IBM

PowerHA SystemMirror for IBM i Cookbook

- Take advantage of PowerHA to configure and manage high availability
- Find guidance on planning and implementing PowerHA
- Benefit from the latest PowerHA solution enhancements

Hernando Bedoya
Abdell Ali-Darwish
Ingo Dimmer
Sabine Jordan
KyoSeok Kim
Akinori Mogi
Nandoo Neerukonda
Tomasz Piela
Marc Rauzier

Redbooks

ibm.com/redbooks

International Technical Support Organization

PowerHA SystemMirror for IBM i Cookbook

January 2012

SG24-7994-00

Note: Before using this information and the product it supports, read the information in “Notices” on page ix.

First Edition (January 2012)

This edition applies to Version 7, Release 1, Modification 0 of IBM i (5770-SS1) and related licensed porgram products.

Contents

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

Active Memory™
AIX®
AS/400®
DB2®
Distributed Relational Database Architecture™
DRDA®
DS6000™
DS8000®
FlashCopy®
Global Technology Services®
i5/OS®
IBM®
iCluster®
iSeries®
Micro-Partitioning®
OS/400®
POWER Hypervisor™
Power Systems™
POWER6+™
POWER6®
POWER7®
PowerHA®
PowerVM®
POWER®
Redbooks®
Redbooks (logo) ®
Storwize®
System i®
System Storage®
System x®
System z®
Tivoli®

The following terms are trademarks of other companies:

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

SAP, and SAP logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

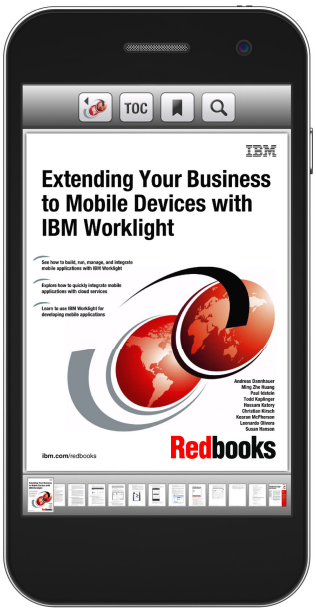
Other company, product, or service names may be trademarks or service marks of others.

Find and read thousands of IBM Redbooks publications

- Search, bookmark, save and organize favorites
- Get up-to-the-minute Redbooks news and announcements
- Link to the latest Redbooks blogs and videos

Get the latest version of the Redbooks Mobile App

iOS **Download Now** Android



Promote your business in an IBM Redbooks publication

Place a Sponsorship Promotion in an IBM® Redbooks® publication, featuring your business or solution with a link to your web site.

Qualified IBM Business Partners may place a full page promotion in the most popular Redbooks publications. Imagine the power of being seen by users who download millions of Redbooks publications each year!

ibm.com/Redbooks
About Redbooks → Business Partner Programs

THIS PAGE INTENTIONALLY LEFT BLANK

Preface

IBM® PowerHA® SystemMirror for i is the IBM high-availability disk-based clustering solution for the IBM i 7.1 operating system. When combined with IBM i clustering technology, PowerHA for i delivers a complete high-availability and disaster-recovery solution for your business applications running in the IBM System i® environment. PowerHA for i enables you to support high-availability capabilities with either native disk storage or IBM DS8000® or DS6000™ storage servers or IBM Storwize® V7000 and SAN Volume Controllers.

The latest release of IBM PowerHA SystemMirror for i delivers a brand new web-based PowerHA graphical user interface that effectively combines the solution-based and task-based activities for your HA environment, all in a single user interface.

This IBM Redbooks® publication gives a broad understanding of PowerHA for i. This book is divided into three major parts:

- Part 1, “Introduction and background” on page 1, provides a general introduction to clustering technology, independent ASPs, PowerHA SystemMirror products, and PowerHA Architecture.
- Part 2, “Concepts and planning” on page 65, describes and explains the various interfaces that PowerHA for i has. It also explains the HA concepts as they pertain to Geographic Mirroring, DS8000 Copy Services, Storwize V7000 and SAN Volume Controller Copy Services, and Advanced Copy Services. It also shows you licensing and ordering information and outlines several considerations to remember when sizing and performance of the HA solution that you are planning to deploy using IBM PowerHA SystemMirror for i.
- Part 3, “Implementation examples and best practices” on page 197, walks you through several scenarios with a step-by-step approach for configuring and managing your IBM PowerHA SystemMirror for i solution. For each scenario, we show you how to perform a planned switchover, and we also discuss the procedures for unplanned failover. In Chapter 15, “Best practices” on page 397, we share our recommendations and best practices to follow in a Highly Available environment that uses various components of the IBM PowerHA SystemMirror for i product.

If you are new to high availability, we recommend that you follow the general structure and flow of this book so that you can start by learning the concepts and progress into the implementation scenarios.

If you are familiar with high availability building blocks or have previous experience with any of the solutions that we discuss in this book, then we recommend that you familiarize yourself with the changes that are new to the latest release of the product.

Since the original writing of this book, IBM iCluster® for IBM Power Systems™, a logical replication software product for high availability and disaster recovery that runs on the IBM i operating system, has been acquired by Rocket Software, Inc. HA Assist, a derivative offering of the iCluster product, was also included in the sale to Rocket Software. HA Assist for IBM i is a specially priced and packaged offering of iCluster, which can be used in conjunction with PowerHA SystemMirror for i for specific situations.

The team who wrote this book

This book was produced by a team of specialists from around the world working at the International Technical Support Organization, Poughkeepsie Center.

Hernando Bedoya is a Senior IT Specialist at STG Lab Services and Training, in Rochester, Minnesota. He writes extensively and teaches IBM classes worldwide in all areas of DB2® for i. Before joining STG Lab Services he worked in the ITSO for nine years writing multiple IBM Redbooks publications. He also worked for IBM Colombia as an IBM AS/400® IT Specialist doing pre-sales support for the Andean countries. He has 25 years of experience in the computing field and has taught database classes in Colombian universities. His areas of expertise are database technology, performance, and data warehousing. He has a master's degree in Computer Science from EAFIT, Colombia.

Abdel Ali-Darwish is a IT Specialist for IBM i and a Technical Solution Architect working in the IBM Global Technology Services® support organization in Lima, Perú. He has eight years of experience with IBM i systems. Over the years he has participated in several pre-sales and post sales support roles including the midrange platform. He currently provides technical pre-sales support to IBM Multicountry for Power Systems with emphasis on IBM i. His current responsibilities include designing solutions and configurations, providing marketing information, and serving as a subject matter expert in technical and delivery assessments.

Ingo Dimmer is an IBM Consulting IT Specialist for IBM i and a PMI Project Management Professional working in the IBM STG Europe storage support organization in Mainz, Germany. He has 12 years of experience in enterprise storage support from working in IBM post-sales and pre-sales support. His areas of expertise include IBM i external disk and tape storage solutions, I/O performance, and high availability. He has been an author of several white paper and IBM Redbooks publications. He holds a degree in electrical engineering from the Gerhard-Mercator University Duisburg.

Sabine Jordan is a Consulting IT Specialist working in IBM Germany. She has worked as a Technical Specialist in the IBM i area for more than 20 years, specializing in high availability since 2004. She has worked on IBM PowerHA System Mirror for i implementations for both SAP and non-SAP environments using geographic mirroring and DS8000 remote copy services. Among these implementations, she has created concepts for the design and implemented the entire project (cluster setup, application changes), in addition to performing customer education and testing. In addition, Sabine presents and delivers workshops (internal and external) on IBM PowerHA System Mirror for i and high availability and disaster recovery.

KyoSeok Kim is a Senior IT Specialist for IBM i working for IBM Global Technology Services in Korea. He has 17 years of experience with AS400, iSeries®, System i, i5/OS®, and IBM i. He is second-level support (Top Gun) for IBM i and Power System for IBM Korea. His areas of expertise include internal, performance, database, and IBM i problem determination. He holds a master's degree in computer engineering and a Bachelor of Science degree in physics from Korea University.

Akinori Mogi is an IBM Consulting IT Specialist for IBM i and the Power Systems platform. He works in the Technical Sales division in IBM Japan. He has over 20 years of experience in AS/400, iSeries, System i, and IBM i pre-sales and post sales support activities as a technical expert. His areas of expertise are high-availability solutions, virtualization, and system performance of the IBM i environment. He has recently joined Power Systems Technical Sales is responsible for ATS and FTSS. Akinori is also an instructor of information technology at a university.

Nandoo Neerukonda is an IBM i Consultant specializing in performance management, query optimization on DB2 for i, high availability, systems programming, and security. He has 15

years of experience with IBM i and its predecessors and has worked for Countrywide Financial (now Bank of America) and Penske Truck Leasing. He currently provides consulting services through his own corporation, Metixis, Inc. He is a speaker at COMMON and a co-author for two other IBM Redbooks, *DB2 Universal Database for iSeries Administration: The Graphical Way on V5R3*, SG24-6092, and *End to End Performance Management on IBM i*, SG24-7808. Nandoo can be contacted via email at nandoo.neerukonda@metixis.com.

Tomasz Piela is an IT Specialist working in the IBM support organization in Katowice, Poland. He has 13 years of experience with IBM i support, consulting and solution implementation. He holds a degree in computer science from Silesian University of Technology in Gliwice. His areas of expertise include IBM i system performance, HA, and DR solutions for IBM i.

Marc Rauzier is an IBM Certified IT Specialist working for IBM France in Global Technology Services - Services Delivery organization in Lyon. He has more than 20 years of experience in information technology, focusing on AS/400, iSeries, System i, i5/OS, and IBM i. He is responsible for the architecture, design, and implementation of IBM Power Systems and the IBM i based-solution for strategic outsourcing of customers in France. His areas of expertise include IBM i, HMC, VIOS, SAN, 35xx series tape libraries, and DS8x00 series external storage related to the IBM i environment.

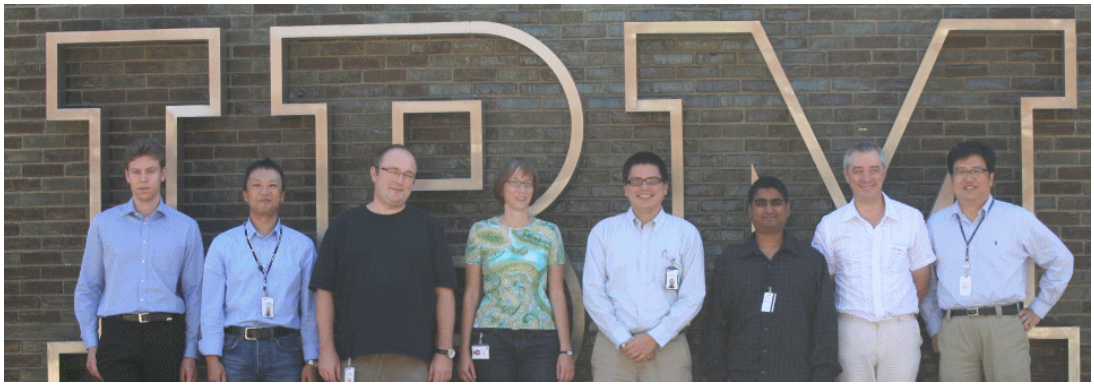


Figure 1 From left to right: Ingo Dimmer, Akinori Mogi, Tomasz Piela, Sabine Jordan, Abdell Ali-Darwish, Nandoo Neerukonda, Marc Rauzier, and KyoSeok Kim

Thanks to the following people for their contributions to this project:

Jenifer Servais
Linda Robinson
International Technical Support Organization

Troy Biesterfeld
Tom Crowley
Jenny Dervin
Steven Finnes
Amanda Fogarty
James Lembke
Curt Schemmel
Christopher Wilk
Ben Rabe
IBM Development Rochester

Laural Bauer
Selwyn Dickey
Jerry Evans

Brian Walker
Tim Klubertanz
John Stroh
Cindy Mestad
Steven Ransom
John Stroh
Charles Farrell
IBM STG Lab Services Rochester

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

Part 1

Introduction and background

For years our clients have been asking when IBM will offer a *hardware* solution for high availability. Over the past decade and with each subsequent release of the operating system, we introduced the building blocks that eventually enabled us to deliver a complete integrated IBM i solution with IBM PowerHA SystemMirror for i. We are pleased to be able to offer our customers a complete set of IBM solution options that address their high-availability and disaster-recovery needs.

IBM recently made significant investments to further enhance PowerHA SystemMirror for i as its strategic high availability product for the IBM i platform.

With the October 2011 announcement, PowerHA SystemMirror for i now also supports IBM System Storage® SAN Volume Controller and IBM Storwize V7000 for storage-based replication solutions. A new PowerHA GUI and further enhancements for IASP-based high-availability solutions complement the new PowerHA functionality.

This book is structured in three parts, with Part 1 covering an introduction and the architecture of IBM PowerHA SystemMirror for i, Part 2, “Concepts and planning” on page 65, providing concepts and information about planning, and Part 3, “Implementation examples and best practices” on page 197, providing implementation examples and scenarios along with best practices.

1

Introduction to PowerHA SystemMirror for i

This chapter provides an overview of the IBM PowerHA SystemMirror for i business continuity solutions, including new enhancements introduced with the October 2011 announcement.

1.1 IBM i Business Continuity Solutions

Increasing business demands for application availability require more and more customers of any size to look after a solution that can help eliminate planned and unplanned downtimes for their IT services.

An unplanned outage that in duration or recovery time exceeds business expectations can have severe implications, including unexpected loss of reputation, customer loyalty, and revenue. Customers who did not effectively plan for the risk of an unplanned outage, never completed their installation of an HA solution, or did not have a tested tape recovery plan in place are especially exposed to negative business impacts.

Figure 1-1 shows an example how a *high-availability* solution can help you to significantly reduce planned and unplanned application downtimes.

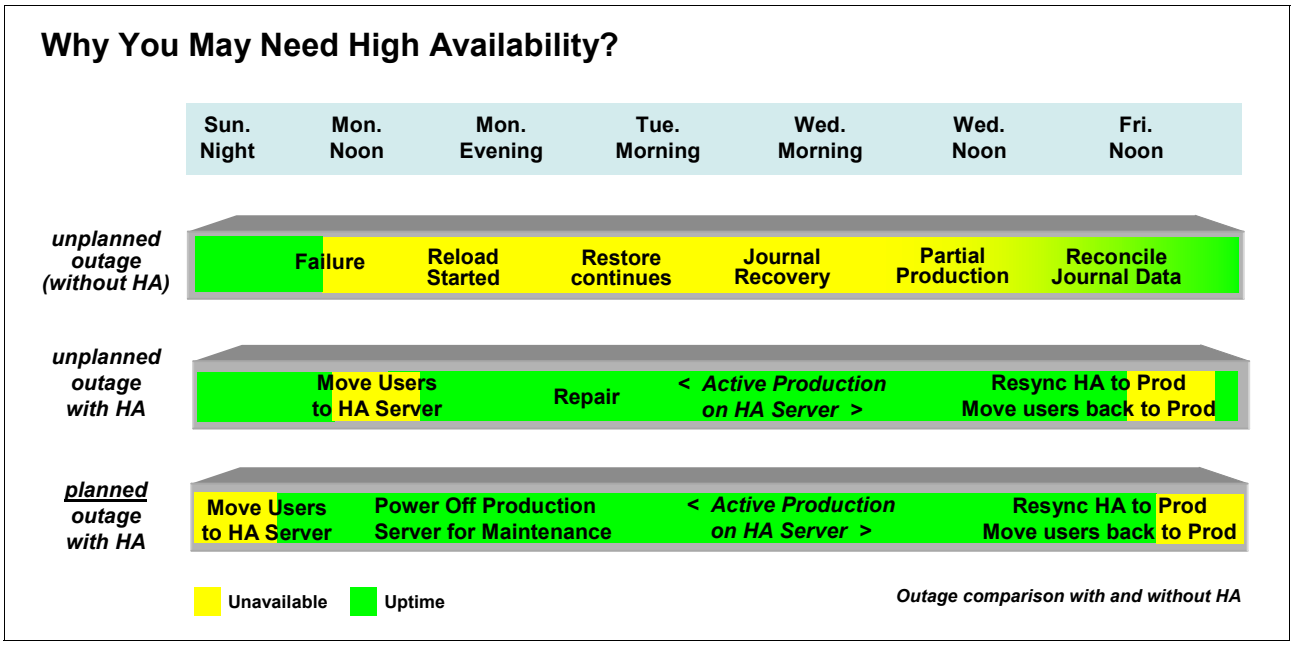


Figure 1-1 An example of why you might need high availability

To address customer needs for high availability for their IBM i environment, IBM announced the PowerHA for i licensed program (LP) with IBM i 6.1, which with IBM i 7.1 is now called IBM PowerHA SystemMirror for i. The IBM PowerHA SystemMirror for i solution offers a complete end-to-end integrated clustering solution for high availability (HA) and disaster recovery (DR). PowerHA provides a data and application resiliency solution that is an integrated extension of IBM i operation system and storage management architecture (Figure 1-2) and has the design objective of providing application high availability through both planned and unplanned outages.

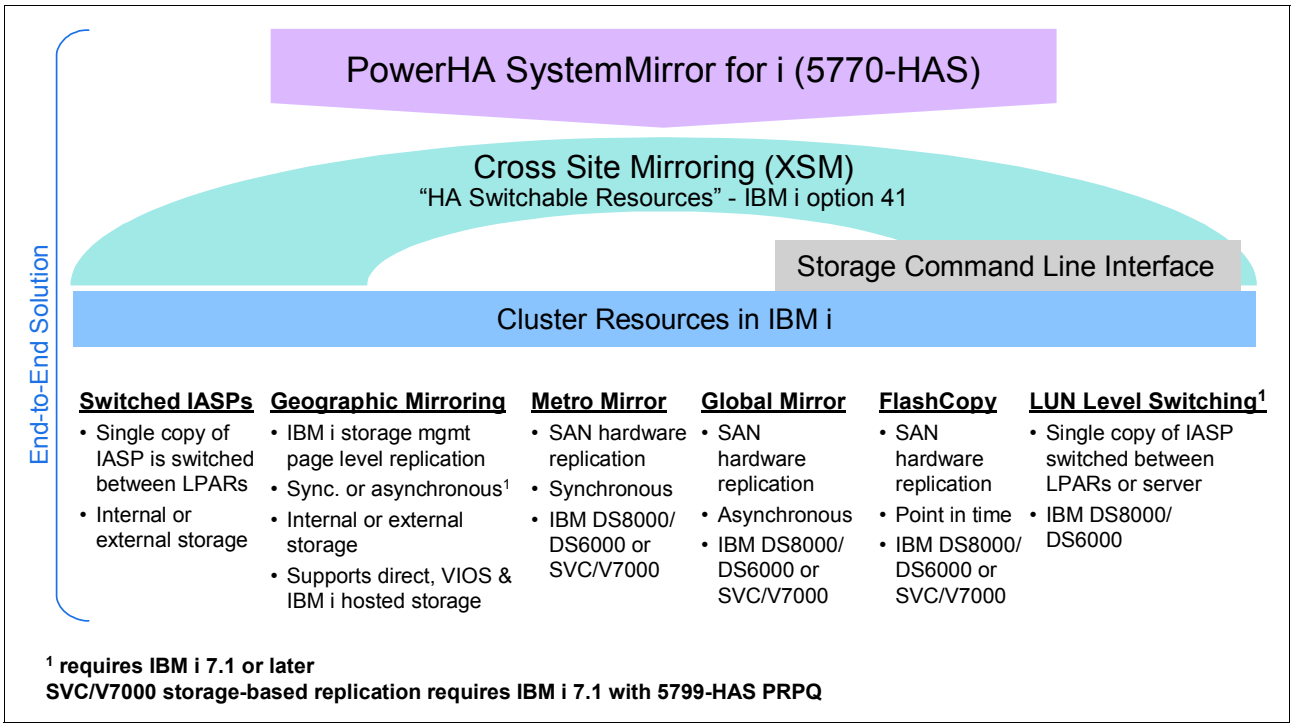


Figure 1-2 PowerHA SystemMirror for i architecture

The key characteristic that our PowerHA customers love is that the solution is automated. Because the data resiliency is completely managed within the IBM i storage management architecture, there is no operator involvement, just as there is no operator involvement with RAID 5 or disk mirroring. Geographic mirroring offers IBM i customers an IBM i-based page-level replication solution for implementing high availability and disaster recovery with any kind of IBM i-supported internal or external storage solution. With IBM System Storage DS8000/DS6000 or SAN Volume Controller (SVC)/Storwize V7000 storage servers our clients are able to exploit storage-based remote replication functions for high availability and disaster recovery, LUN-level switching for local high availability, and FlashCopy® for reducing save window outages by enabling the creation of a copy that is attached to a separate partition for off-line backup to tape.

We offer IBM i customers a full menu of HA/DR solution choices (Figure 1-3).

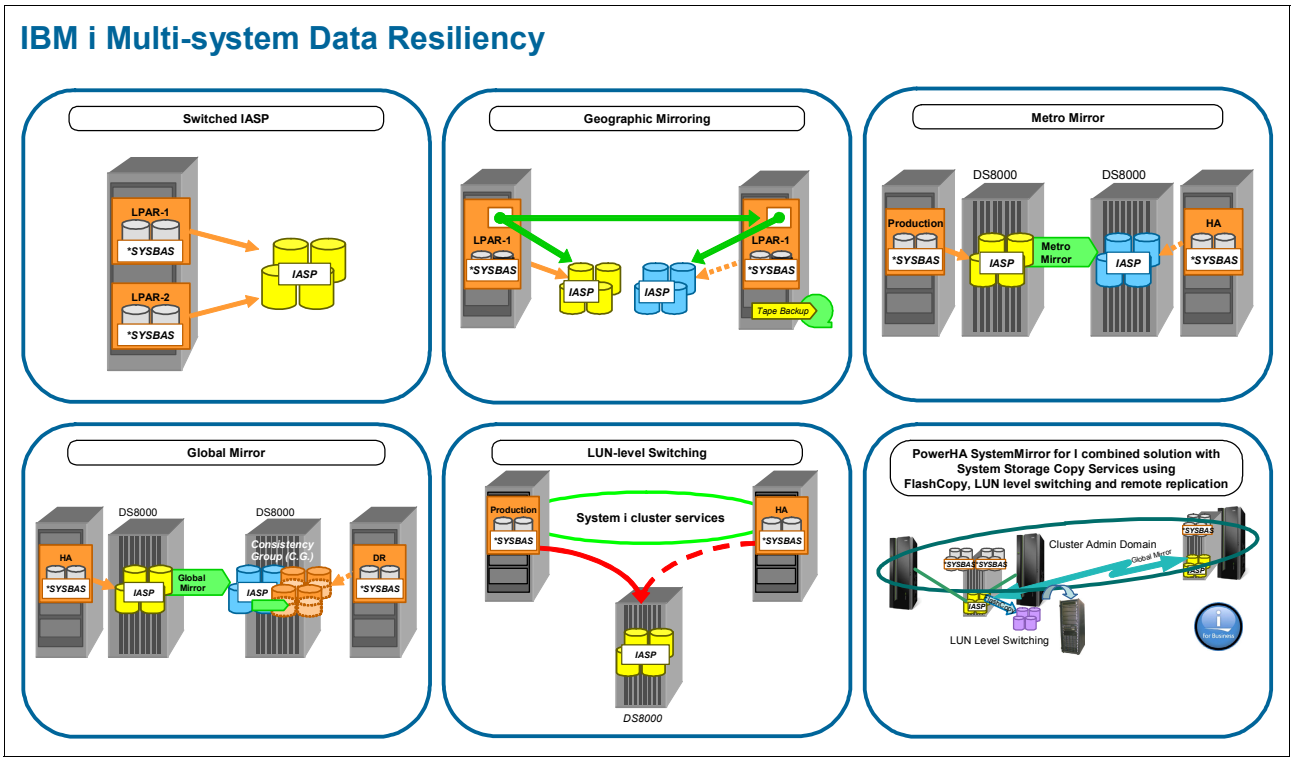


Figure 1-3 PowerHA SystemMirror for i multi-system data resiliency solutions

For more detailed information about IBM PowerHA SystemMirror for i architecture and business resiliency solutions see Chapter 4, “PowerHA architecture” on page 45.

1.1.1 PowerHA SystemMirror for i 7.1 availability capabilities

This section describes these enhancements included with PowerHA SystemMirror for i 7.1:

- PowerHA SystemMirror for i Editions

  IBM PowerHA SystemMirror for i is now offered in two editions for IBM i 7.1:

  - IBM PowerHA SystemMirror for i Standard Edition (5770-HAS *BASE) for local datacenter replication only
  - IBM PowerHA SystemMirror for i Enterprise Edition (5770-HAS option 1) for local or multi-site replication

- PowerHA versioning

  To use any of the new PowerHA SystemMirror for i enhancements all nodes in the cluster need to be upgraded to IBM i 7.1.

- Clustering GUI support change

  The clustering GUI plug-in for System i Navigator from High Availability Switchable Resources licensed program (IBM i option 41) has been removed in IBM i 7.1. Clustering HA environments can continue to be configured and managed using the PowerHA for i licensed product (5770-HAS), CL commands, and the IBM Systems Director Navigator for i web interface.

- N_Port ID virtualization support

  Using NPIV with PowerHA does not require dedicated Fibre Channel IOAs for each SYSBAS and IASP. Instead, virtual adapters can be defined for the partitions.

- Asynchronous Geographic Mirroring

  Asynchronous geographic mirroring is a new function supported by PowerHA SystemMirror for i Enterprise Edition with IBM i 7.1 extending the previously available synchronous geographic mirroring option, which for performance reasons is practically limited to metro area distances up to 40 km.

- LUN-level switching

  One copy of iASP switched between two partitions/systems managed by a cluster resource group device domain and located in IBM System Storage DS8000 or DS6000 series. An ASP session is not required for LUN-level switching, as there is no replication for the IASP involved.

- Space-efficient FlashCopy

  PowerHA for SystemMirror for i with IBM i 7.1 newly supports space-efficient FlashCopy of the IBM System Storage DS8000 series. The IBM System Storage DS8000 series FlashCopy SE licensed feature allows creation of space-efficient FlashCopy target volumes that can help to reduce the required physical storage space for the FlashCopy target volumes. These volumes are typically needed only for a limited time (such as for the duration of a backup to tape).

- Better detection of cluster node outages

  With IBM i 7.1, PowerHA SystemMirror for i now allows advanced node failure detection by cluster nodes. This is done by registering with an HMC or Virtual I/O Server (VIOS) management partition on IVM-managed systems. This way clustering gets notified in case of severe partition or system failures to trigger a cluster failover event instead of causing a cluster partition condition.

- Improved Geographic Mirroring full synchronization performance

  Performance improvements have been implemented in IBM i 7.1 for geographic mirroring full synchronization. The achievable performance improvement varies based on the IASP data. IASPs with a large number of small objects see more benefit than those with a smaller number of large objects.

- Cluster administrative domain enhancements

  PowerHA SystemMirror for i is required to support these new administration domain Monitored Resource Entries (MREs):

  - Authorization lists (*AUTL)
  - Printer device descriptions (*PRTDEV) for LAN or virtual printers

- IBM HA Assist for i

  This is a new licensed product (5733-HAA) for IBM i 6.1 and later that was announced with IBM i 6.1.1 as an extension for PowerHA only. IBM HA Assist for i is based on iCluster code to replicate objects not supported for IASPs or by the cluster administrative domain. It is primarily targeted at customers with existing applications that cannot be fully migrated to an IASP environment.

- IPv6 support

  PowerHA SystemMirror for i on IBM i 7.1 now fully supports IPv6 or a mix of IPv6 and IPv4. All HA-related APIs, commands, and GUIs have been extended for field names holding either a 32-bit IPv4 or a 128-bit IPv6 address.

- New CL commands for programming cluster automation

  With PowerHA SystemMirror for i, these new CL commands are introduced in IBM i 7.1 to better support CL programming for cluster automation management:

  - **RTVCLU** (Retrieve Cluster)
  - **RTVCRG** (Retrieve Cluster Resource Group)
  - **RTVASPCPYD** (Retrieve ASP Copy Description)
  - **RTVASPSSN** (Retrieve ASP Session)
  - **PRTCADMRE** (Print Cluster Administrative Domain Managed Resource Entry)

  For further information about the new PowerHA CL commands see the IBM i 7.1 Information Center at the following web page:

  http://publib.boulder.ibm.com/infocenter/iseries/v7r1m0/topic/rbam6/HAS.htm

1.1.2 PowerHA SystemMirror for i: 2011 enhancements

These new functions are delivered with the October 2011 announcement for PowerHA SystemMirror for i:

- New 5799-HAS PRPQ for IBM PowerHA SystemMirror for i:
  - Support for managing IBM System Storage SAN Volume Controller (SVC) and IBM Storwize V7000 Copy Services functions of FlashCopy, Metro Mirror and Global Mirror.
  - IBM i CL command **CFGDEVASP** for configuring an independent auxiliary storage pool.
  - IBM i CL command **CFGGEOMIR** for configuring geographic mirroring.
  - New PowerHA GUI providing the facility to handle the high-availability solution starting from a single screen. It currently supports these:
    - Geographic mirroring
    - Switched disk (IOA)
    - DS8000/DS6000 FlashCopy
    - Metro Mirror and Global Mirror
  - 5799-HAS PRPQ is English only and requires 5770-HAS PTF SI44148.
- N-2 support for clustering

  This allows you to skip one level of IBM i such as upgrading a V5R4M0 system within a clustered environment directly to i 7.1 by skipping i 6.1.
- Duplicate library error handling

  A message ID CPDB8EB is displayed in the QSYSOPR message queue for a library name conflict between SYSBAS and a varying-on IASP. The vary on can be continued or cancelled after the duplicate library issue is resolved.

1.2 Choosing a solution

Today our customers have many choices and need to evaluate which is their best high-availability and disaster-recovery solution. We suggest that the criteria for choosing the correct solution must be based on business needs like the recovery point objective (RPO), recovery time objective (RTO), geographic dispersion requirements, staffing, skills, and day-to-day administrative efforts. Figure 1-4 shows typical recovery time objectives for various recovery solutions associated with seven tiers of business continuity (BC).

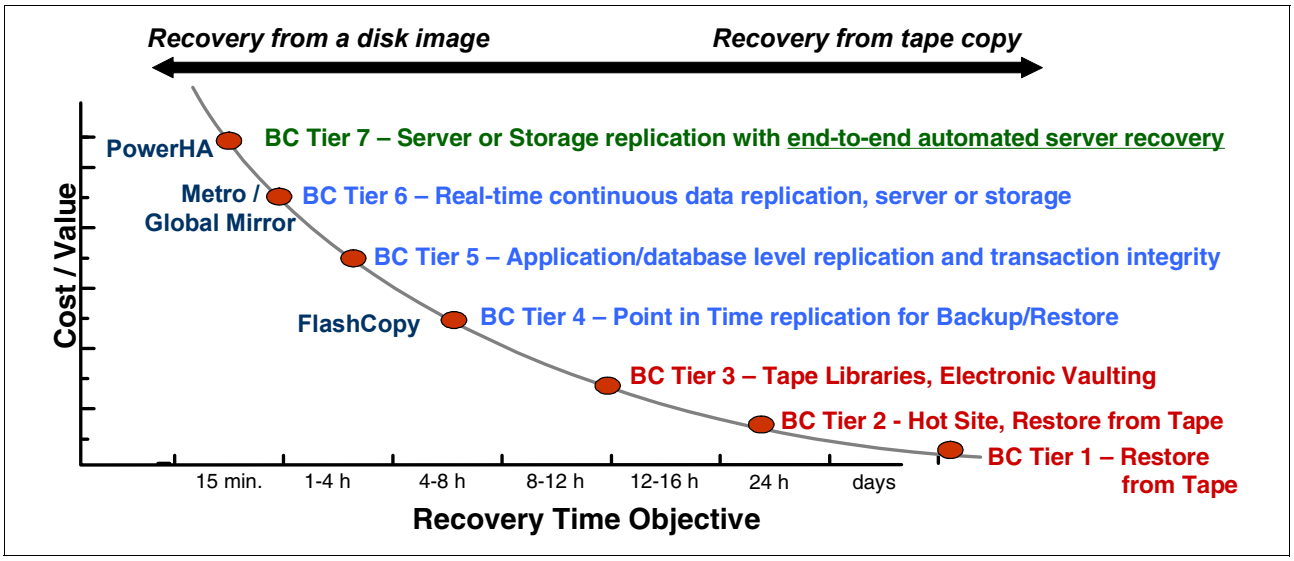


Figure 1-4 Seven tiers of disaster recovery

When you start thinking about implementing HA for your IBM i environment, consider how this criteria apply to your situation before deciding which solution fits your needs best:

- Types of outages to be addressed
  - Unplanned outages (for example, a hardware failure)
  - Planned outages (for example, a software upgrade)
  - Backups (for example, creating a copy of disk for an online save to tape)
  - Disasters (for example, site loss, power grid outage, and so on)
- Recovery objectives
  - Recovery time objective (RTO): The time to recovery from an outage
  - Recovery point objective (RPO): The amount of tolerable data loss (expressed as a time duration)

IBM i data resiliency solutions are either based on logical replication or hardware replication (Figure 1-5). Unlike the previously mentioned PowerHA IASP hardware-based replication solutions, logical replication solutions like IBM iCluster or high availability business partner (HABP) replication solutions send journal entries via TPC/IP from the production system to a backup system where the journal entries are applied to the database. Appendix A, “IBM i data resilience options” on page 451, provides further information, including a comparison of the IBM i data resiliency solutions.

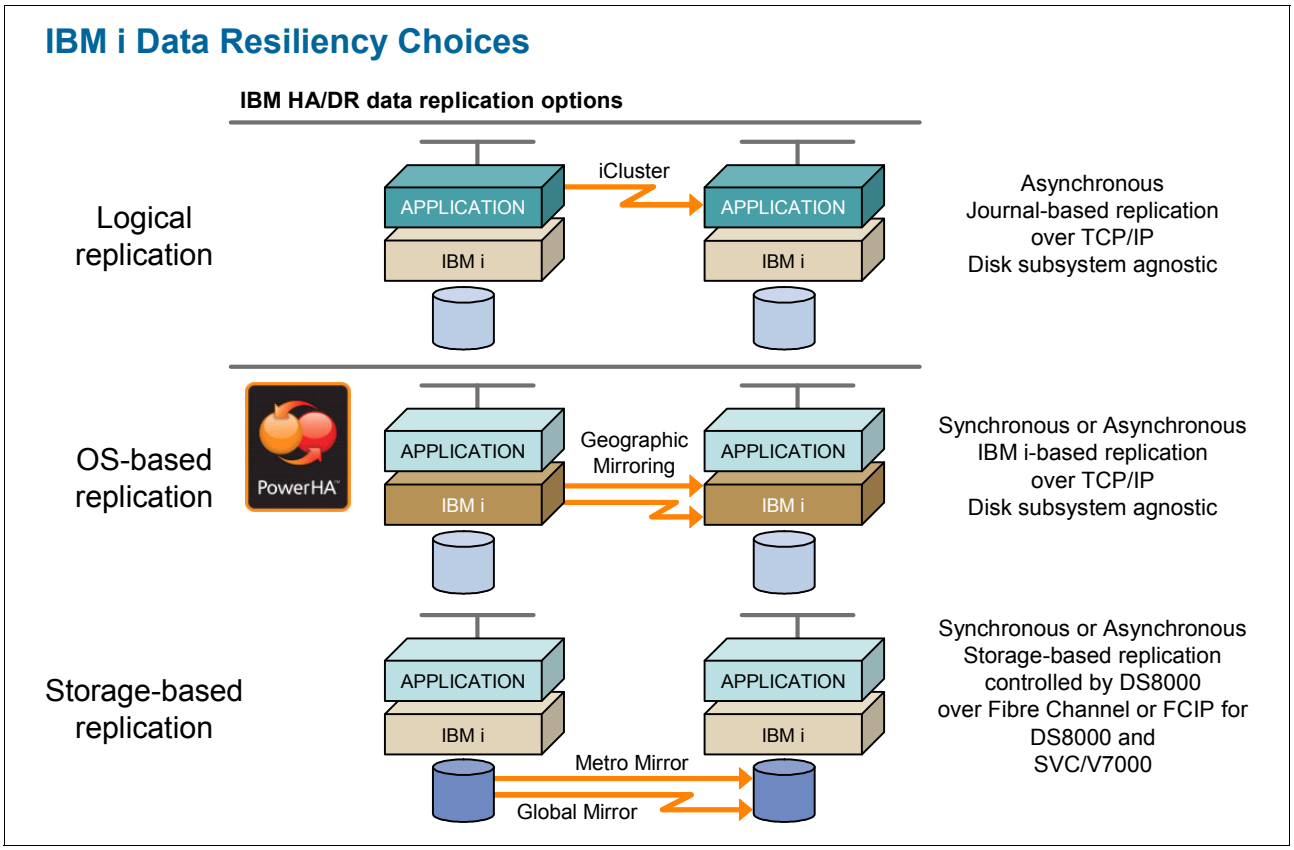


Figure 1-5 IBM i HA/DR data replication options

Solution considerations

In this section we explain concepts that can help you to decide which solution your business requires.

A storage-based *synchronous* replication method is one in which the application state is directly tied to the act of data replication, just as it is when performing a write operation to local disk. You can think of the primary and secondary IASP copies as local disk from the application perspective. This aspect of a synchronous replication approach means that all data written to the production IASP is also written to the backup IASP copy and the application waits just as though it were a write to local disk. The two copies cannot be out of sync, and also the distance between the production and backup copies, in addition to the bandwidth of the communication link, will have an influence on application performance. The farther apart the production and backup copies, the longer the synchronous application steps will need to wait before proceeding to the next application step. For a longer distance exceeding the limits of a metro area network consider using an *asynchronous* hardware replication solution to prevent or minimize performance impacts for critical applications. The huge benefit in comparison to a logical replication approach is that the two copies are

identical minus the data in the “pipe,” and therefore the secondary copy is ready to be varied on for use on a secondary node in the cluster.

The *cluster administrative domain* is the PowerHA function that ensures that the set of objects that are not in an IASP are synchronized across the nodes in the cluster. Thus, the application has the resources that it needs to function on each node in the cluster. Clustering solutions deployed with iASPs and using either storage-based copy services or geographic mirroring replication require little in the way of day-to-day administrative maintenance and were designed from the beginning for role-swap operations. We define an HA environment as one in which the primary and secondary nodes of the cluster switch roles on a regular and sustained basis.

Rule of thumb: If your business does not conduct regular and sustained role swaps, your business does not have a high-availability solution deployment.

The *logical replication* in the IBM i environment is based on IBM i journaling technology, including the option of remote journaling. A key characteristic of logical replication is that only those objects that are journalled by IBM i (that is, database, IFS, data area, data queue) can be replicated in near real time.

Synchronous remote journaling provides synchronous replication for the above-mentioned objects, but all other objects are captured via the audit journal and then replicated to the target system. The practical ramification of this type of replication approach is that there are administrative activities required to ensure that the production and backup copes of data are the same prior to a role-swap operation. Another issue is that there can be a significant out-of-sync condition between the primary and secondary copies of data while the backup server works to apply the data sent from the primary trying to catch up. The benefit of the logical replication approach is that the production and backup systems can be virtually any distance from each other and the backup copy can be used for read operations.

In addition, because one can choose to replicate a subset of objects, the bandwidth requirements are typically not as great in comparison to a hardware-based replication approach.

2

Implementing an independent auxiliary storage pool

Independent auxiliary storage pools (IASPs) are a fundamental building block for implementing Power HA System Mirror for IBM i. In this chapter we provide you with a brief overview of the concept in addition to step-by-step instructions to create them. In addition, we describe the steps necessary to move an existing application environment into an IASP and successfully run it in this new environment.

2.1 IASP technology

IBM i has used the concept of single-level storage since its first release. All space available on disks and in main memory is treated as one continuous address range where users or programs are not actually aware of the location of the information that they want to access.

As the need to segregate groups of programs and data on the same system emerged, the concept of pools developed and was included as part of the operating system. The pools were referred to as *auxiliary storage pools* (ASPs) because they pertained to areas of auxiliary storage (disk space). The new command structures within the operating system used the letters ASP when referring to the auxiliary storage pools.

Enhancements to the concept of pools has led to *independent auxiliary storage pools* introduced with OS/400® V5R1. These are pools that can be brought online, taken offline, and accessed independently of the other pools on the system. They can even be logically or physically switched between systems or logical partitions.

These are the disk pools that are available today:

- System disk pool (disk pool 1)

  The system disk pool contains the load source and all configured disks that are not assigned to any other disk pool.

- Basic disk pools (disk pool 2 to 32)

  Basic disk pools can be used to separate objects from the system disk pool. For example, you can separate your journal receivers from database objects. Basic disk pools and data contained in them are always accessible when the system is up and running.

- Primary disk pool

  This is an independent disk pool that defines a collection of directories and libraries and might have other secondary disk pools associated with it. Primary disk pools and any associated secondary pools can be taken offline or brought online independent of system activity on other disk pools. Data in a primary disk pool can only be accessed by jobs on the system if the disk pool is brought online and the job gets connected to the IASP.

- Secondary disk pool

  This is an independent disk pool that defines a collection of directories and libraries and must be associated with a primary disk pool. It is comparable to basic disk pools because it is again used to separate specific application objects like journal receivers from your main application objects.

- User-defined file system disk pool

  This is an independent disk pool that contains only user-defined file systems (UDFSs).

- Disk pool groups

  Disk pool groups consist of a primary disk pool and zero or more secondary disk pools. Each disk pool is independent in regard to data storage, but in the disk pool group they combine to act as one entity (for example, they are varied on and off together and switchover is done for the entire disk pool group). Making disk pools available to the users is accomplished by using the disk pool group name.

Figure 2-1 illustrates the hierarchy of ASPs on a system.

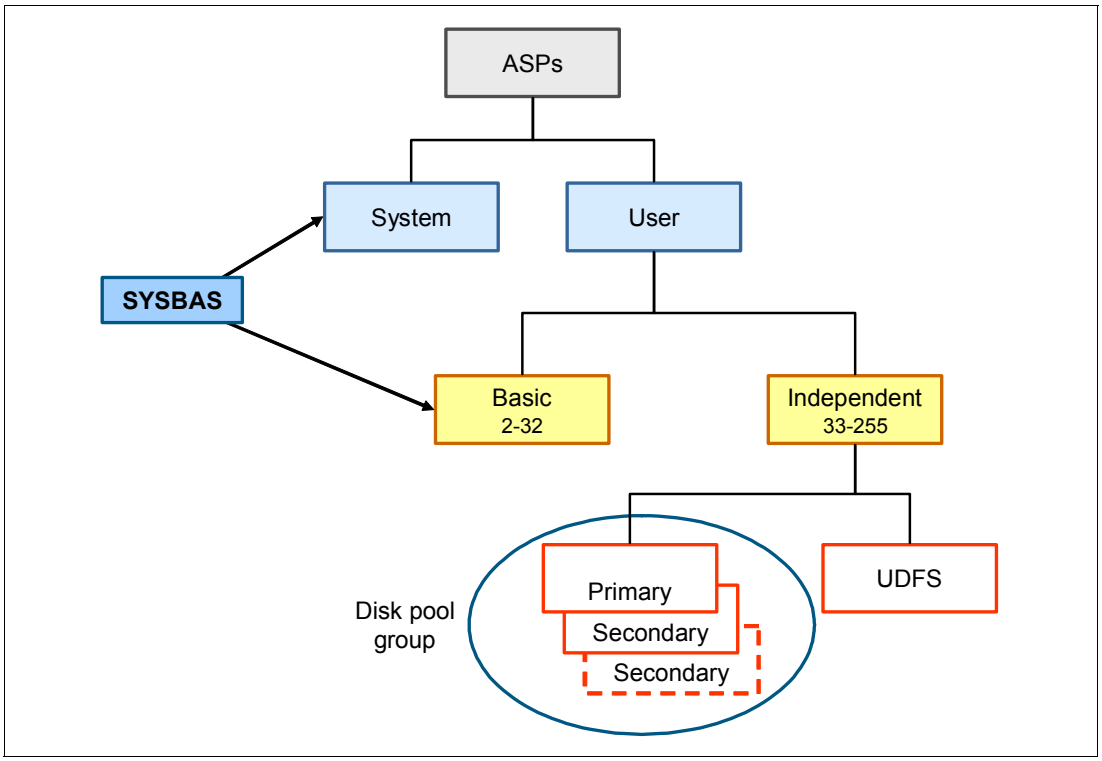


Figure 2-1 ASP hierarchy

There is a difference between basic ASPs and independent ASPs when it comes to data overflow. Basic ASPs overflow and independent ASPs do not. An overflow of a basic user ASP occurs when the ASP fills. The excess data spills into the system ASP. IASPs are designed so that they cannot overflow. Otherwise, they would not be considered independent or switchable. An IASP is allowed to fill up, and the application that is responsible for filling it up simply halts. There is no automatic cancellation of the responsible job. If this job is running from a single-threaded JOBQ, in a single-threaded subsystem all further processing is stopped until user action is initiated.

When an IASP fills up, the job that generates the data that filled up the disk pool might not be complete. The system generates an MCH2814 message indicating this condition. This might have serious ramifications. Jobs that only read data are still able to work, but any job trying to add data to the IASP is on hold. The system does not automatically cancel the offending job. If the job is from a single-threaded JOBQ or a single-threaded subsystem, other jobs behind it are held up until the offending job is handled. Possible scheduling impacts might occur.

2.1.1 Name space

Prior to the introduction of library-capable IASPs, any thread, including the primary or only thread for a job, can reference the following libraries by name:

- The QTEMP library for the thread's job, but not the QTEMP library of any other job
- All libraries within the system ASP
- All libraries within all existing basic user ASPs

This set of libraries formed the library *name space* for the thread and was the only possible component of that name space. Although there was not a formal term for this name space component, it is now referred to as the *SYSBAS component of the name space. It is a required component of every name space.

With library-capable IASPs, a thread can reference, by name, all of the libraries in the IASPs of one ASP group. This adds a second, but optional, component to the name space and is referred to as the *ASP group component* of the name space. A thread that does not have an ASP group component in its name space has its library references limited to the *SYSBAS component. A thread with an ASP group component to its library name space can reference libraries in both the *SYSBAS and the ASP group components of its name space.

Library names no longer must be unique on a system. However, to avoid ambiguity in name references, library names must be unique within every possible name space. Because *SYSBAS is a component of every name space, presence of a library name in *SYSBAS precludes its use within any IASP. Because all libraries in all IASPs of an ASP group are part of a name space, for which the ASP group is a component, existence of a library name within one IASP of an ASP group precludes its use within any other IASP of the same ASP group. Because a name space can have only one ASP group component, a library name that is not used in *SYSBAS can be used in any or all ASP groups.

IBM i has a file interface and an SQL interface to its databases. The file interface uses the name space to locate database objects. For compatibility, SQL maintains a catalog for each ASP group. This catalog resides in the primary IASP of the ASP group. The catalog is built from the objects that are in a name space that has the ASP group and *SYSBAS as its two components. The names database and the name space are somewhat interchangeable because they refer to the same set of database objects.

Each name space is treated as a separate relational database by SQL. It is required that all RDBs whose data is accessible by SQL are defined in the RDB directory on the system.

Note that the name space is a thread attribute and can be specified when a job is started. When it is referenced as a *job attribute*, it technically means the “thread attribute for the initial thread of a single-threaded job.”

2.1.2 Relational Database directory

The Relational Database (RDB) directory allows an application requester (AR) to accept an RDB name from the application and translate this name into the appropriate IP address or host name and port. In addition, the RDB directory can also specify the user's preferred outbound connection security mechanism. The relational database directory can also associate an Application Requester Driver (ARD) program with an RDB name.

Each IBM i system in the distributed relational database network must have a relational database directory configured. There is only one relational database directory on a system. Each AR in the distributed relational database network must have an entry in its relational database directory for its local RDB and one for each remote and local user RDB that the AR accesses. Any system in the distributed RDB network that acts only as an application server does not need to include the RDB names of other remote RDBs in its directory.

The RDB name assigned to the local RDB must be unique from any other RDB in the network. Names assigned to other RDBs in the directory identify remote RDBs or local user databases. The names of remote RDBs must match the name that an ASP uses to identify its local system database or one of its user databases, if configured. If the local system RDB name entry for an application server does not exist when it is needed, one is created automatically in the directory. The name used is the current system name displayed by the Display Network Attributes (**DSPNETA**) command.

Figure 2-2 gives an example of the RDB directory on a system with an IASP configured. Notice that there is one entry present with a remote location of local. This is the RDB entry representing the database in SYSBASE. In addition, an RDB entry gets created by the operating system when you vary on an IASP. In our case this is the entry IASP1 with a remote location of loopback. By default, the relational database name of an IASP is identical to the IASP device name, but you can also choose another name here. When migrating an application environment with a large number of accesses through the RDB name, you might want to change the SYSBASE RBD name to a different value and use the “old” SYSBASE RDB name as the database name for the IASP. This way, you do not have to change RDB access to your environment.

```
               Work with Relational Database Directory Entries

Position to  . . . . . .

Type options, press Enter.
  1=Add   2=Change   4=Remove   5=Display details   6=Print details

                            Remote
Option  Entry               Location                  Text

        IASP1               LOOPBACK                  Entry added by system
        S10C78FP            *LOCAL                    Entry added by system




                                                                        Bottom
F3=Exit   F5=Refresh   F6=Print list   F12=Cancel   F22=Display entire field
(C) COPYRIGHT IBM CORP. 1980, 2009.
```

Figure 2-2 Work with Relational Database Directory Entries

Although the objects in the system RDB are logically included in a user RDB, certain dependencies between database objects have to exist within the same RDB. These include:

- A view into a schema must exist in the same RDB as its referenced tables, views, or functions.
- An index into a schema must exist in the same RDB as its referenced table.
- A trigger or constraint into a schema must exist in the same RDB as its base table.
- Parent table and dependent table in a referential constraint both have to exist in the same RDB.
- A table into a schema has to exist in the same RDB as any referenced distinct types.

Other dependencies between the objects in the system RDB and the user RDB are allowed. For example, a procedure in a schema in a user RDB might reference objects in the system RDB. However, operations on such an object might fail if the other RDB is not available, such as when the underlaying IASP is varied off and then varied on to another system. A user RDB is local to IBM i while the IASP is varied on. But as an IASP can be varied off on one server and then varied on to another server, a user RDB might be local to a given server at one point in time and remote at a different point in time.

2.1.3 Connections

In an SQL environment, SQL CONNECT is used to specify the correct database. To achieve the best performance, make sure that the database being connected to corresponds with your current library name space. You can use **SETASPGRP** or the INLASPGRP parameter in your job description to achieve this. If the SQL CONNECT function is not operating within the same library name space, the application uses Distributed Relational Database Architecture™ (DRDA®) support, which can affect performance.

2.1.4 Object creation

While it is possible to create files, tables, and so on, into QSYS2, the corresponding library in the independent disk pool prevents this from occurring. Most applications that create data in QSYS2 do not realize it and fail when running in an independent disk pool.

Consider the example in Example 2-1 with library demo10 residing in an IASP and the job running the SQL being attached to an IASP. In this example, the view ICTABLES is not built in the current library (DEMO10) as you would expect. It is built in the library of the first table that is mentioned, which is QSYS2 (where SYSTABLES is located). It fails when accessing the independent disk pool because creation of objects in QSYS2XXXXX is prevented. In the example mentioned, you must explicitly specify that you want to create the view either in QSYS2 or in a user library in the IASP.

Example 2-1 Create view on SYSTABLES

```
CHGCURLIB DEMO10
create view ICTABLES(Owner, tabname, type) as select table_schema, TABLE_NAME, TABLE_TYPE
from SYSTABLES where table_name like'IC%'
```

2.1.5 System-wide statement cache (SWSC)

A separate SWSC is created and maintained on each IASP. Multiple sets of system cross-reference and SQL catalog tables are defined and maintained on each IASP.

The IASP version of QSYS and QSYS2 contains cross-reference and SQL catalog tables with merged views of all the SQL and database objects that are accessible when connected to the IASP.

2.2 Creating an IASP

To create an IASP, you need to have at least one unconfigured disk available on your system. The IASP can either be created using Systems Director Navigator for IBM i or using the new CL command **CFGDEVASP**, available with the 5799-HAS PRPQ.

If you want to use Systems Director Navigator for IBM i to create an IASP, the following tasks have to be performed before doing so:

1. Ensure that the IBM i user profile that you are using to access disk units has these authorities:
   - *ALLOBJ: All object authority
   - *SERVICE
2. Start DST via the service panel function 21.
3. Sign on to DST using your service tools user ID and password.
4. When the Use Dedicated Service Tools (DST) display is shown, select option 5 (Work with DST environment) and press Enter. The Work with DST Environment display is shown.
5. At the Work with DST Environment menu, select option 6 (Service tools security data).
6. At the Work with Service Tools Security Data menu, select option 6 (Change password level). Make sure that the password level is set to Secure Hash Algorithm (SHA) encryption or password level 2, and press F12.
7. At the Work with DST Environment display, select option 3 (Service tools user IDs) to work with service tools user IDs.
8. Create a service tools user ID that matches the IBM i user profile and that also has the same password in uppercase. The service tools user ID and password must match the IBM i user profile and password of the user using IBM Systems Director Navigator for IBM i. For example, if the user profile and password combination is BOB and my1pass, then the DST user ID and password combination must be BOB and MY1PASS. If the service tool user ID that you intend to use existed before changing the password level, then you have to change its password before using IBM i Systems Director Navigator.
9. Give this service tools user ID at least these authorities:
   - Disk units: operation
   - Disk units: administration
10. Press Enter to enable these changes and exit DST.

When starting IBM i Systems Director Navigator, you can find the Configuration and Service tasks in the main task menu (Figure 2-3).

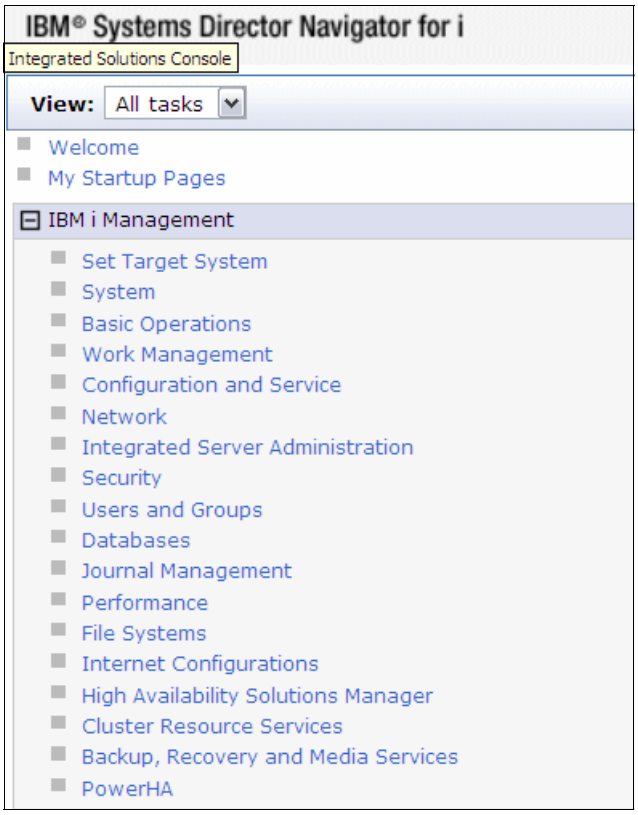


Figure 2-3 System Director Navigator main menu

After choosing **Configuration and Service**, a a list of options displays (Figure 2-4):

1. To create a new independent ASP, choose **Disk Pools**.

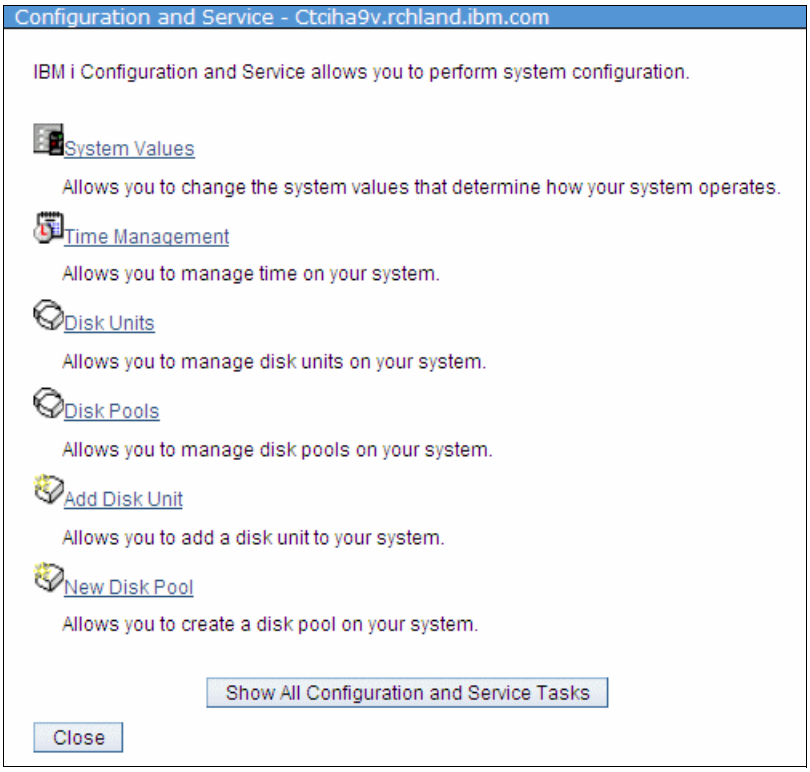


Figure 2-4 System Director Navigator: Configuration and service tasks

You are presented with the current disk pool configuration (Figure 2-5).

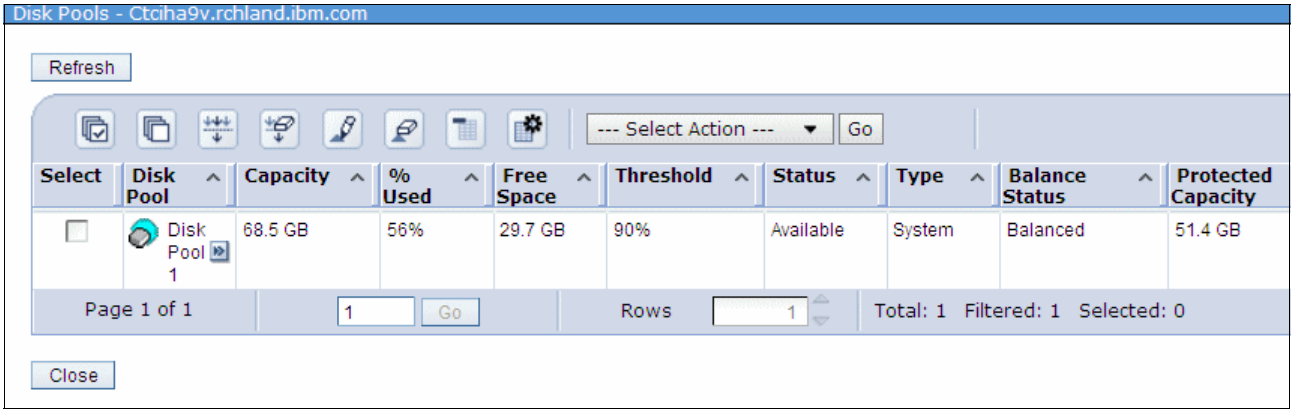


Figure 2-5 System Director Navigator: Current disk pool configuration

2. From the Select Action pull-down menu, choose **New Disk Pool** (Figure 2-6). Click **Go** after making your selection.

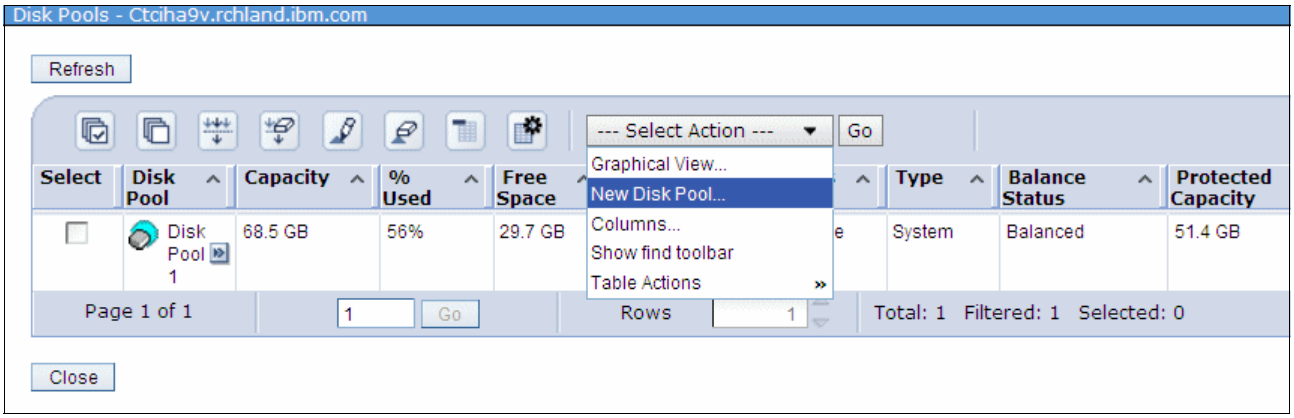


Figure 2-6 System Director Navigator: New disk pool

3. In the next window, choose **Primary** as the type of disk pool (Figure 2-7). You can then enter the name of your IASP. The database name defaults to the name of the IASP, but you can also enter a different name for the database. Be aware that the IASP database name cannot be identical to the system ASP database name.

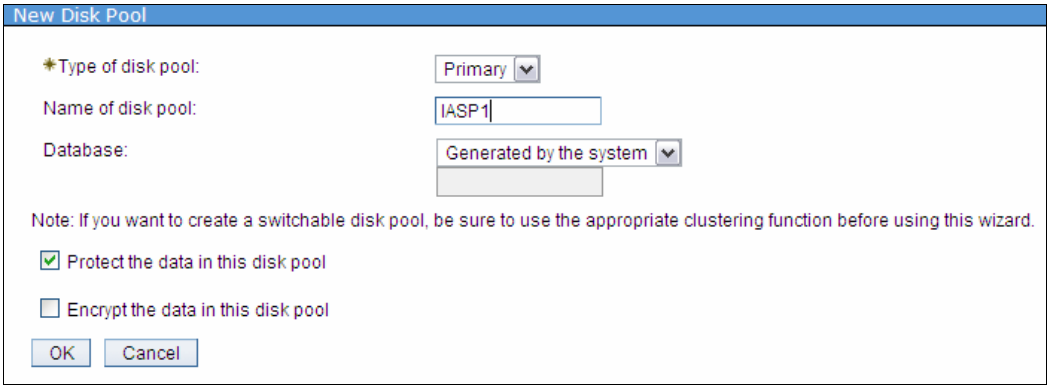


Figure 2-7 System Director Navigator: Disk pool details

4. If the **Protect data in this disk pool** check box is marked, the GUI will give you the ability to select disks for the IASP that are either parity protected or that can be mirrored. If the check box is not marked, you can only add unprotected disks to your IASP. Because we marked the check box in our example, we can choose whether we want to add parity protected disk or disk that can be mirrored (Figure 2-8).

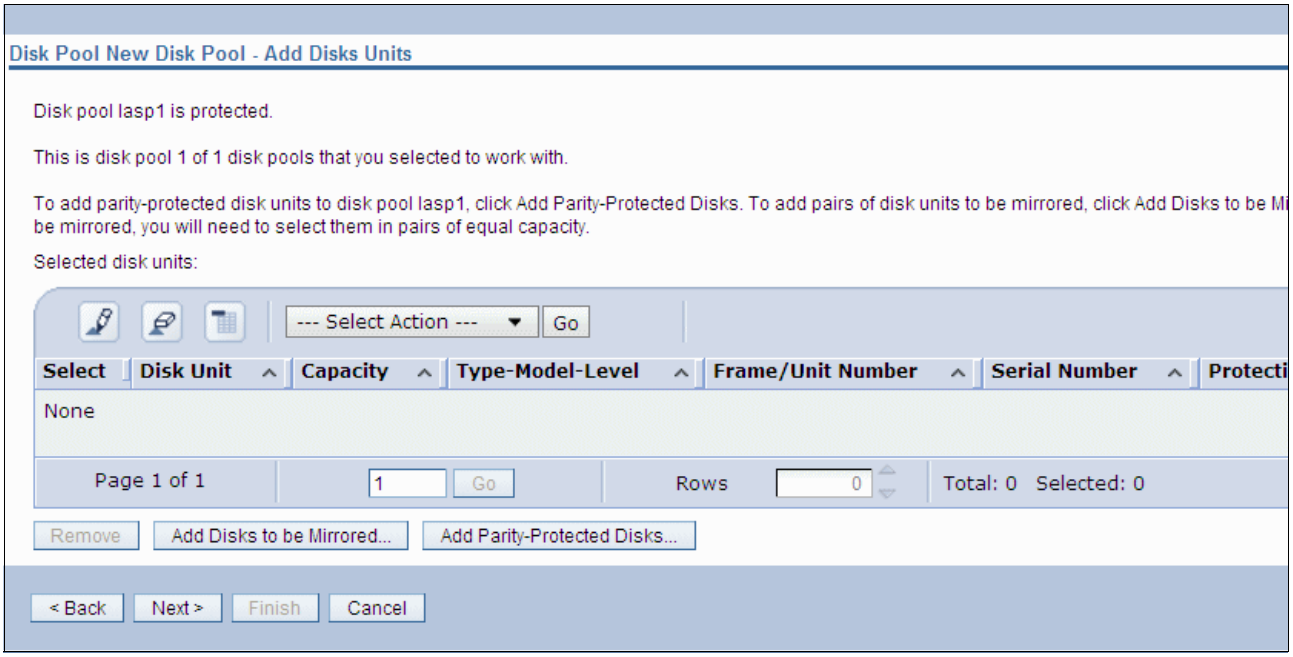


Figure 2-8 System Director Navigator: Add Disk Units

5. Choosing **Add Disks to be Mirrored** then gives us a list of disks that can be put into the new IASP (Figure 2-9). Make sure to select all the disks that you want to have in your IASP and click **Add**.

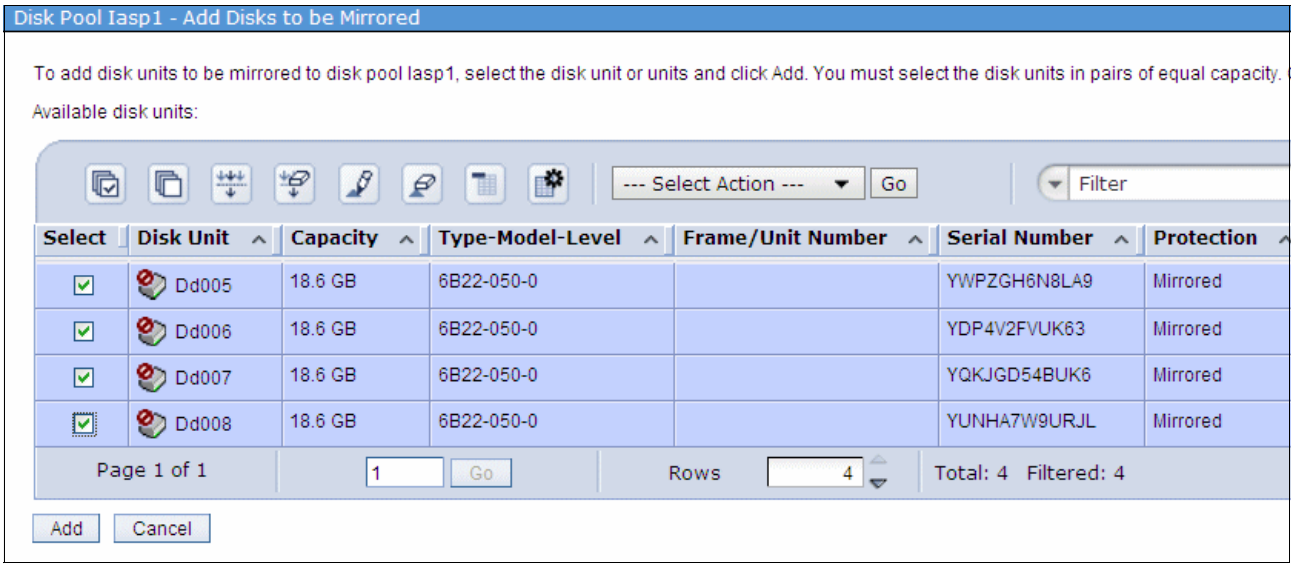


Figure 2-9 System Director Navigator: Choose disks

6. The summary window gives you an overview of the new disk pool configuration that you are about to create (Figure 2-10). Verify that is what you wanted to achieve and click **Finish**.

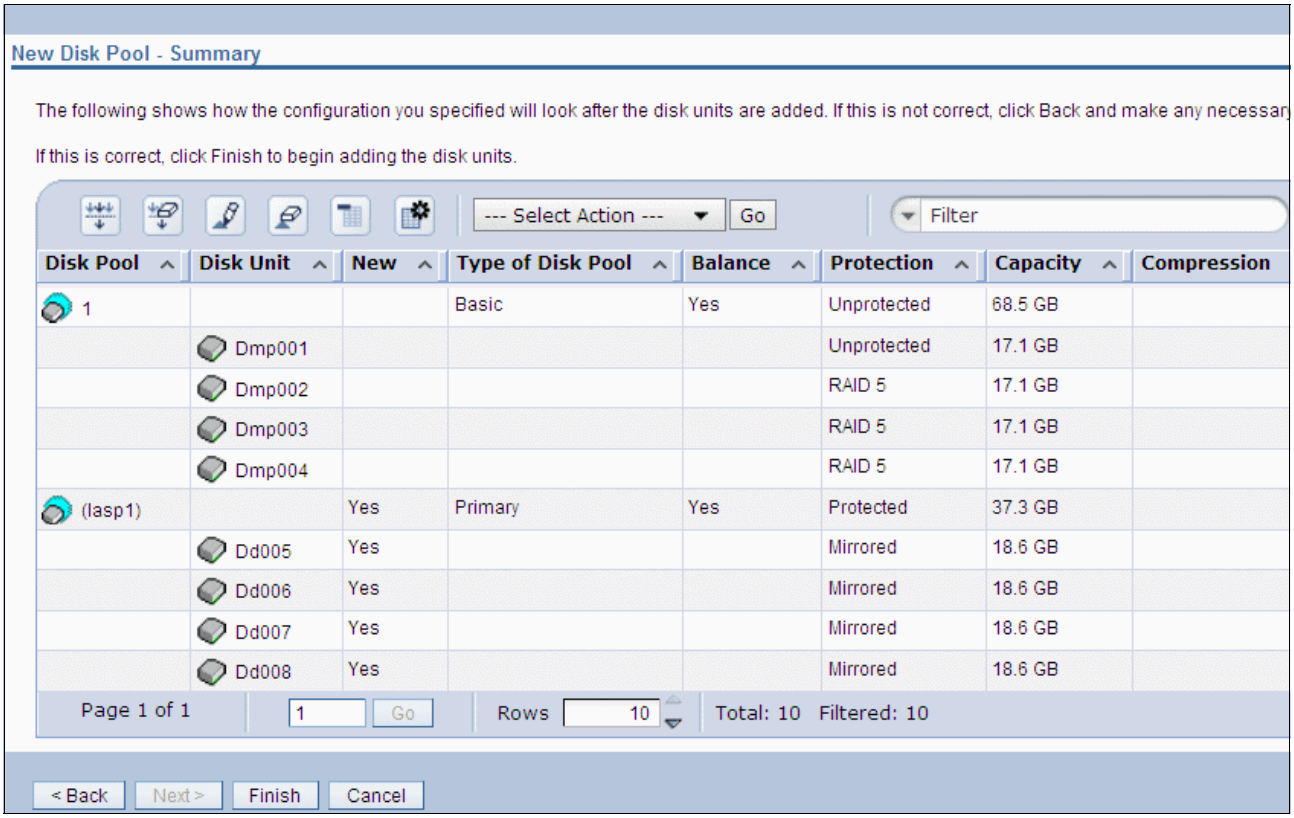


Figure 2-10 System Director Navigator: New Disk Pool Summary

7. The IASP creation then starts (Figure 2-11). Be aware that this screen does not automatically refresh. Refresh must be done manually.

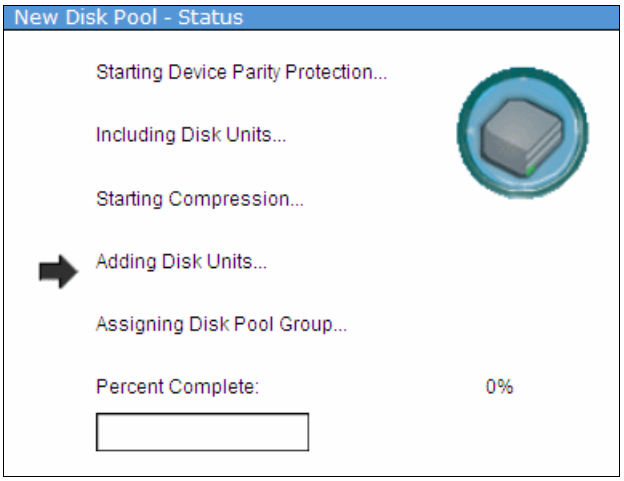


Figure 2-11 System Director Navigator: Disk pool creation

In 11.3, “Creating an IASP” on page 215, we show how to create an IASP using the new PowerHA GUI.

Alternatively, you can use a CL command to create your iASP. Figure 2-12 shows the parameter required for **CFGDEVASP**.

```
                      Configure Device ASP (CFGDEVASP)

Type choices, press Enter.

ASP device . . . . . > IASP1         Name
Action . . . . . > *CREATE       *CREATE, *DELETE
ASP type . . . . . .   *PRIMARY      *PRIMARY, *SECONDARY, *UDFS
Protection . . . . . .   *NO           *NO, *YES
Encryption . . . . . .   *NO           *NO, *YES
Disk units . . . . . .   *SELECT       Name, *SELECT
                                                  + for more values




                                                                  Bottom
F1=Help   F9=Calculate Selection   F11=View 2   F12=Cancel
```

Figure 2-12 CFGDEVASP command to create IASP

Specifying *SELECT for the disk unit parameter shows the screen shown in Figure 2-13. It provides you with a list of unconfigured disks on your system. Choose which ones you want to add to your IASP and press Enter to create the IASP.

```
                      Select Non-Configured Disk Units

ASP device . . . . . :   IASP1
Selected capacity  . . . . . :   0
Selected disk units  . . . . . :   0

Type options, press Enter.
  1=Select

     Resource
Opt  Name        Serial Number    Type Model  Capacity  Rank  Eligible
 1   DD007       YQKJGD54BUK6     6B22 0050      19088  002      Yes
 1   DD006       YDP4V2FVUK63     6B22 0050      19088  002      Yes
 1   DD008       YUNHA7W9URJL     6B22 0050      19088  002      Yes
 1   DD005       YWPZGH6N8LA9     6B22 0050      19088  002      Yes



                                                                  Bottom
F1=Help   F9=Calculate Selection   F11=View 2   F12=Cancel
Configuration of ASP device IASP1 is 8% complete.
```

Figure 2-13 CFGDEVASP command: Select disks to put into IASP

A message on the bottom of the screen shows the progress of the IASP creation. Your screen is locked as long as the creation of the IASP is running. You can also follow the status of the IASP creation using **DSPASPSTS**.

2.3 Moving applications to an IASP

Having successfully created in IASP, the next step is to look at the considerations for migrating your application into it. We discuss what object types can be moved into an IASP and which cannot, how you get access to objects in an IASP, and various aspects of your application environment with an IASP.

2.3.1 Object considerations

To understand the steps necessary to move an application into an IASP we first need to have an understanding of which objects can be located in an IASP and which objects cannot. Table 2-1 shows which object types can be put into an IASP.

Table 2-1 Object types that can be put into IASP

*ALRTBL	*FIFO	*MGTCOL	*QRYDFN
*BLKSF	*FILE	*MODULE	*SBSD
*BNDDIR	*FNTRSC	*MSGF	*SCHIDX
*CHTFMT	*FNTTBL	*MSGQ	*SPADCT
*CHRSF	*FORMDF	*NODGRP	*SQLPKG
*CLD	*FTR	*NODL	*SQLUDT
*CLS	*GSS	*OVL	*SVRPGM
*CMD	*IGCDCT	*OUTQ	*STFM
*CRQD	*JOBD	*PAGDFN	*SVRSTG
*CSI	*JOBQ	*PAGSEG	*SYMLNK
*DIR	*JRN	*PDG	*TBL
*DSTFMT	*JRNRCV	*PGM	*USRIDX
*DTAARA	*LIB	*PNLGRP	*USRQ
*DTADCT	*LOCALE	*PSFCFG	*USRSPC
*DTAQ	*MEDDFN	*QMFORM	*VLDL
*FCT	*MENU	*QMQRY	*WSCST

For some of these object types, special considerations have to be taken into account:

- If network attributes reference the *alert table*, then this object must reside in the system ASP.
- If an active subsystem references a *class object*, then this class object must reside in the system ASP.
- *Database files* that are either multiple-system database files or that have DataLink fields that are created as link control cannot be located in an independent disk pool. If an active subsystem references the file object, *FILE must exist in the system disk pool, for example, the sign-on display file.
- *Subsystem descriptions* can be placed into in IASP. However, if you want to start a subsystem, its subsystem description has to be in the system ASP at that point in time.

We therefore recommended putting subsystem descriptions into an extra library located in the system ASP.

- The same is true for *job descriptions*. There are a number of cases in which job descriptions can only be used when they are located in the system ASP, for example, when they are used for autostart jobs.
- *Job queues* in an IASP are operationally identical to job queues in the system ASP. Users can manipulate the jobs (submit, hold, release, and so on) or the job queues themselves (clear, hold, release, delete, and so on). However, the internal structures holding the real content of the job queue still reside in the system ASP. Job entries in a job queue in an IASP do not survive the vary off and vary on of an IASP and therefore are lost when switching over from production to backup system.
- *Journals* and *journal receivers* must be located in the same IASP group as objects being journaled. Journal receivers can be moved to a secondary IASP within that IASP group.
- A *library* that is specified by CRTSBSD SYSLIBLE() must exist in the system ASP. In addition, libraries referenced in the system values QSYSLIBL or QUSRLIBL cannot be located in an IASP.
- If network attributes reference a *message queue*, then that message queue must be located in the system ASP.
- *Programs* referenced in subsystem descriptions (for example, in routing entries or prestarted jobs) must be found in the system ASP when that subsystem is activated. The same is true if a program is associated with the attention key.

Table 2-2 shows which object types cannot be put into an IASP.

Table 2-2 Object types that cannot be put into an IASP

*AUTHLR	*CTLD	*IGCTBL	*NWSD
*AUTL	*DDIR	*IPXD	*PRDAVL
*CFGL	*DEVD	*JOBSCD	*RCT
*CNNL	*DOC	*LIND	*SOCKET
*COSD	*EDTD	*MODD	*SSND
*CRG	*EXITRG	*M36	*S36
*CSPMAP	*FLR	*M36CFG	*USRPRF
*CSPTBL	*IGCSRT	*NTBD	IBM libraries

If you look at the list of object types you will notice that most of them are either legacy objects (such as folders or documents), configuration objects (such as device descriptions or line descriptions), or objects closely related to system security (such as user profiles or authority lists).

If you want to keep objects in the system ASP in sync between your production and your backup system (such as user IDs and passwords) you can use the administrative domain to help you with this task. See 3.5, “Advanced node failure detection” on page 43.

2.3.2 Accessing objects in an IASP

By default, each job on the system can only access objects that are stored in the system ASP. To get access to objects in an IASP, that IASP has to be varied on. In addition, the IASP has to be added to the namespace of the job. This can be done in these ways:

- By using the CL command **SETASPGRP**
- By changing the job description that the job is using

The recommended way is to use job descriptions, as that is also applicable to prestarted jobs.

Important: *Never* change the QDFTJOBD job description. Doing so leaves you with an unusable system after an IPL.

2.3.3 Considerations for specific environments

Now that we know how you generally access data in an IASP, let us look at specific environments and the impact using an IASP has on them.

Considerations for system values

Before you implement independent disk pools, examine how you use the following system values. System values have no access to **SETASPGRP**. In most cases, the programs that they reference as their values must exist in *SYSBAS. The system values that are affected by an implementation of independent disk pools are as follows:

- QALWUSRDMN: Allows user domain objects in libraries

  This value specifies which libraries can contain user domain user (*USRxxx) objects. You can specify up to 50 individual libraries or all libraries on the system. Specifying the name of a library makes all libraries with that name (which might exist in separate independent auxiliary storage pools) eligible to contain user domain user objects.

- QATNPGM: Attention program

  This value specifies the name and library of the attention program. This program must exist in the system ASP or in a basic user ASP.

- QCFGMSGQ: Configuration message queue

  This system value allows you to specify the default message queue that the system uses when sending messages for lines, controllers, and devices. The message queue must exist in the system ASP or in a basic user ASP.

- QCTLSBSD: Controlling subsystem

  The controlling subsystem is the first subsystem to start after an IPL. At least one subsystem must be active while the system is running. This is the controlling subsystem. Other subsystems can be started and stopped. If this subsystem description cannot be used (for example, it is damaged), the backup subsystem description QSYSSBSD in the library QSYS can be used. A subsystem description specified as the controlling subsystem cannot be deleted or renamed after the system is fully operational. The subsystem description specified here must be located in the system ASP.

- QIGCCDEFNT: Double-byte code font

  This value is used when transforming an SNA character string (SCS) into an Advanced Function Printing Data Stream (AFPDS). It is also used when creating an AFPDS spooled file with shift in/shift out (SI/SO) characters present in the data. The IGC coded font must exist in the system ASP or in a basic user ASP. The shipped value is different for different countries or regions.

- QINACTMSGQ: Inactive job message queue

  This value specifies the action that the system takes when an interactive job has been inactive for an interval of time (the time interval is specified by the system value QINACTITV). The interactive job can be ended, disconnected, or message CPI1126 can be sent to the message queue that you specify. The message queue must exist in the system ASP or in a basic user ASP.

  If the specified message queue does not exist or is damaged when the inactive timeout interval is reached, the messages are sent to the QSYSOPR message queue. All of the messages in the specified message queue are cleared during an IPL. If you assign a user's message queue as QINACTMSGQ, the user loses all messages that are in the user's message queue during each IPL.

- QPRBFTR: Problem log filter

  This value specifies the name of the filter object used by the Service Activity Manager when processing problems. The filter must exist in the system ASP or in a basic user ASP.

- QPWDVLDPGM: Password validation program

  This value provides the ability for a user-written program to perform additional validation on passwords. The program must exist in the system ASP or in a basic user ASP.

- QRMTSIGN: Remote sign-on control

  This system value specifies how the system handles remote sign-on requests. The program option allows you to specify the name of a program and library to decide which remote sessions to allow and which user profiles to automatically sign on from which locations. The program must exist in the system ASP or in a basic user ASP.

- QSRTSEQ: Sort sequence

  This system value specifies the default sort sequence algorithm to be used by the system. The sort sequence table must exist in the system ASP or in a basic user ASP.

- QSTRUPPGM: Startup program

  This value specifies the name of the program called from an autostart job when the controlling subsystem is started. This program performs setup functions, such as starting subsystems and printers. The program must exist in the system ASP or in a basic user ASP.

- QSYSLIBL: System part of the library list

  When searching for an object in the library list, the libraries in the system part are searched before any libraries in the user part are searched. The list can contain as many as 15 library names. The libraries must exist in the system ASP or in a basic user ASP.

- QUPSMSGQ: Uninterruptible power supply (UPS) message queue

  This value specifies the name and library of the message queue that will receive UPS messages. It allows you to monitor the message queue and control the power down. If the message queue is not the system operator message queue (QSYS/QSYSOPR), all UPS messages are also sent to the system operator message queue.

- QUSRLIBL: User part of the library list

  When searching for an object in the library list, the libraries in this part are searched after the libraries in the system part and after the product library and current library entries. The list might contain as many as 25 library names. The libraries must exist in the system ASP or in a basic user ASP.

Considerations for network attributes

When you set up independent disk pools for the first time or move applications to independent disk pools, consider the keywords and parameters for the system network attributes. If the keywords and parameters highlighted in the following sections are in use, review them for the impact that independent disk pools might have on their use. These parameters are on the Change Network Attributes (**CHGNETA**) command. Some of them are on the Retrieve Network Attributes (**RTVNETA**) command.

For more information about these commands, see the CL Command Finder function in the iSeries Information Center:

http://publib.boulder.ibm.com/eserver/ibmi.html

To access this function, type `CL Command Finder` in the Search field:

- Alert Filters (ALRFTR)

  This parameter specifies the qualified name of the alert filter used by the alert manager when processing alerts. The alert filter must exist in the system ASP or in a basic user ASP.

- Message Queue (MSGQ)

  This parameter specifies the qualified name of the message queue where messages received through the SNADS network are sent for users with no message queue specified in their user profile or whose message queue is not available. The message queue must exist in the system ASP or in a basic user ASP.

- Distributed Data Management Access (DDMACC)

  This parameter specifies how the system processes distributed data management (DDM) and DRDA requests from remote systems for access to the data resources of the system. The DDM and DRDA connections refer to APPC conversations or active TCP/IP or OptiConnect connections. Changes to this parameter are immediate and apply to DRDA, DDM, or DB2 Multisystem applications. However, jobs that are currently running on the system do not use the new value. The DDMACC value is accessed only when a job is first started. You must specify a special value or program name that dictates how the requests are to be handled. If a program name is specified, the program must exist in the system ASP or in a basic user ASP.

- PC Support Access (PCSACC)

  This parameter specifies how Client Access/400 requests are handled. You must specify a special value or program name that dictates how the requests must be handled. This permits greater control over Client Access/400 applications. Changes to this parameter are immediate. However, jobs currently running on the system do not use the new value. The PCSACC value is used only when a job is first started. If a program name is specified, the program must exist in the system ASP or in a basic user ASP.

Considerations for ODBC/JDBC

ODBC and JDBC work with prestarted jobs that run under user profile QSYS. After a request comes into the system, that request gets connected to one of the prestarted jobs. That request has to use a user profile and password to authenticate. The prestarted job then gets changed to run with this user profile and the environment defined in the user profile settings. Provided that the user profile uses a job description associated with an IASP, the ODBC or JDBC connection can therefore access objects located in the IASP.

If for any reason using a specific job description is not possible with some ODBC or JDBC connections, then they both also provide parameters in the connection setup to explicitly include an IASP in their namespace (Example 2-2) for ODBC.

Example 2-2 Setting access to IASP1 with ODBC

```
SQLAllocHandle(...);
SQLSetEnvAttr(...);
SQLAllocHandle(...);
SQLDriverConnect(hdbc, NULL,
   “DSN=myDSN;DATABASE=IASP1;UID=myUID;PWD=myPWD;”,
   SQL_NTS,NULL,0,NULL,SQL_DRIVER_NOPROMPT);
```

This gives the ODBC connection access to IASP1.

For JDBC, connecting to IASP1 is shown in Example 2-3.

Example 2-3 Setting access to IASP1 in JDBC

```
DriverManager.registerDriver(new AS400JDBCDriver());
AS400JDBCDataSource ds = new AS400JDBCDataSource(“SYS1”);
   ds.setUser(“xxxxxxxxx”);
   ds.setPassword(“yyyyyyyyy”);
   ds.setNaming(“sql”);
   ds.setDatabaseName(“IASP1”);
```

Considerations for FTP

FTP also uses prestarted jobs that run under user profile QTCP. After a request comes into the system, that request gets connected to one of the prestarted jobs. Normally, this request has to use a user profile and password to authenticate. The prestarted job then gets changed to run with this user profile and, provided that the user profile uses a job description associated with an IASP, can access objects located in the IASP.

If this does not work in some cases in your environment (for example, because you provide anonymous FTP access to your system), then you can use the following command to get access to IASP1 within the FTP job:

```
quote rcmd SETASPGRP IASP1
```

Considerations for the Integrated File System (IFS)

IFS objects are stored in a directory structure. Access to the objects is by a path that navigates the directory structure to reach the object. An available IASP has a directory in the root directory that has the same name as the IASP. When the IASP is available, the contents of the IASP are mounted to the IASP directory.

When an IASP is not available (or before the IASP is created), it is possible to create a directory with the name of the IASP. If there is a directory with the same name as an IASP, when the IASP is varied on:

- The MOUNT operation will be successful if the existing directory is empty.
- The MOUNT operation will fail if there are any objects in the existing directory. The vary-on will not fail. The first indication of failure is likely to occur when users try to access objects in the directory. The objects will be missing or incorrect. The only indication that the MOUNT operation failed is message CPDB414 - file system failure with a reason code 1 (The directory to be mounted over is not empty) in the joblog of the thread that performed the vary-on operation. If the IASP environment uses IFS, each vary-on operation should be checked to ensure that the IFS mounted properly.

Access to IFS objects is not affected by **SETASPGRP** or by a job description pointing to an IASP but has to be done using the hierarchical path structure of the IFS. Therefore, if you do not want to change hard-coded paths to IFS objects in your applications, you can create symbolic links from the original location to the IASP location.

Tip: Check that you do not have any IFS directories with the same name as the primary IASP that you want to create.

Considerations for DRDA

There are certain DRDA-related objects that cannot be contained in an IASP. DDM user exit programs must reside in libraries in the system database, as must any ARD programs.

Be aware that the process of varying on an IASP causes the RDB directory to be unavailable for a short period of time. This can cause attempts by a DRDA application requester or application server to use the directory to be delayed or to time out.

Local user database entries in the RDB directory are added automatically the first time that the associated databases are varied on. They are created using the *IP protocol type and with the remote location designated as LOOPBACK. LOOPBACK indicates that the database is on the same server as the directory.

Considerations for database access using SQL

SQL connects to the database that is set in a job's environment. Therefore, if your job description connects you to an IASP, any SQL statement within that job uses the IASP as its database.

When a static SQL program is run, an access plan is created and stored with the program. If the SQL program is located in the system ASP, a job or thread with IASP1 in its namespace will create an access plan for data in IASP1. When a job or thread with IASP2 in its namespace runs the same SQL program, the existing access plan is invalidated, so a new access plan is created and stored in the program. We therefore recommend creating separate static SQL applications in each IASP for best performance if you use more than one IASP in your environment.

For extended dynamic SQL, create a separate SQL package in each IASP for best performance.

Query Management Query and Query Management Procedures

You can resolve the SQL objects (tables, functions, views, types) that are referenced in a Query Management Query (*QMQRY) object. To do this, you use the RDB specified on the RDB parameter or the RDB specified on the **CONNECT**/**SET CONNECTION** commands. This RDB

might be an IASP. The query management objects referenced must be in the current RDB (name space).

When output from **STRQMQRY** is directed to an output file, Query Management ensures that the output file is created on the RDB (name space) that was current at the time that **STRQMQRY** is executed.

Considerations for DB2 Web Query

Web query only references objects in the current RDB (namespace). A *QRYDFN object created in *SYSBAS might reference files in an IASP and vice versa. If a *QRYDFN object created to reference objects in an IASP runs when a different IASP is set as the current RDB (namespace), the *QRYDFN runs successfully if the new IASP contains objects with the same name and the file formats are compatible.

Considerations for spool files

Outqueues can be put into an IASP so that spoolfiles are available on the backup system after a switchover or failover situation. You can only do this for outqueues that are not attached to a physical printer device, as those outqueues have to be placed in QUSRSYS. In addition, for outqueues in an IASP, the connection between a job and its spoolfile ends when the job itself ends. Users can therefore no longer access their old spoolfiles using **WRKSPLF**, but instead have to use **WRKOUTQ**.

2.3.4 Steps for application migration

With the IASP created and understanding the behavior of an IASP, you can start to move your applications to an IASP environment. The general steps to achieve this are:

1. Restore your application libraries into the IASP. Be aware that you cannot have identical library names in the system ASP and in the IASP. If your application was installed in the system ASP of the system that you are using to do the migration, then you have to delete the original libraries before you can restore them into the IASP. Objects not supported in an IASP have to be copied to a different library in the system ASP.
2. Copy IFS data that is part of your application into the new directory inside the IASP. Delete the original IFS data and create symbolic links that redirect access from the old directories to the new ones in the IASP.
3. If you have application objects stored in QGPL or QUSRSYS, move them to a library inside the IASP. QGPL and QUSRSYS cannot be moved to the IASP.
4. Change your application job description to connect to the IASP using the INLASPGRP parameter. If you are using the QDFTJOBD job description then you have to first copy it and change the copy. *Never* change QDFTJOBD to access an IASP (as it is used for the startup program during an IPL, the IPL would fail because at this point in time the IASP is not yet available). Make sure that you set your library environment correctly in the job description because libraries in an IASP cannot be referenced by system values QSYSLIBL or QUSRLIBL.
5. If you created new job description, change user profiles to use them. Make sure that you do not change your IBM i administration user profiles to use a job description with an IASP included. If the IASP is not varied on for any reason, you cannot sign on to the system if your profile points to a job description with an IASP.
6. Test your application.
7. Change your application environment to work with the IASP. Think of save and restore procedures or changes in the startup program to vary on the IASP.

8. Decide on procedures to synchronize objects in the system ASP from the primary to the backup system.
9. Switch your application over to the backup system and test it there.

3

IBM i clustering

In this chapter, we discuss key components of IBM i cluster technology. Before exploring the implementation of IBM PowerHA SystemMirror for i, it is important to first understand IBM i clustering technology and capabilities.

3.1 Cluster

A cluster is a collection of complete systems that work together to provide a single, unified computing resource. The cluster is managed as a single system or operating entity (Figure 3-1). It is designed specifically to tolerate component failures and to support the addition or subtraction of components in a way that is transparent to users. Clusters can be simple, consisting of only two nodes, or very complex with a large number of nodes.

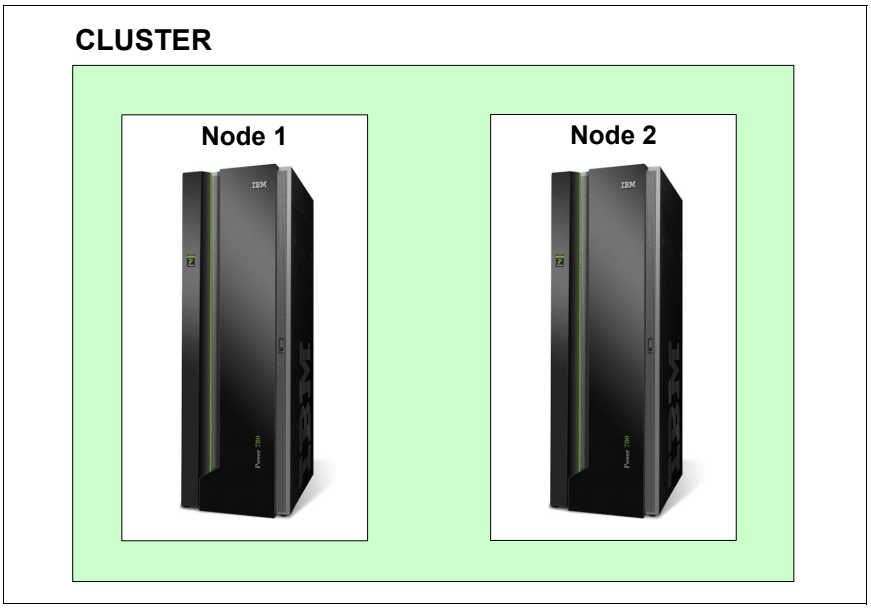


Figure 3-1 A cluster consisting of two nodes

These are the major benefits that clustering offers a business:

- Simplified administration of servers by allowing a customer to manage a group of systems as a single system or single database
- Continuous or high availability of systems, data, and applications
- Increased scalability and flexibility by allowing a customer to seamlessly add new components as business growth develops

Attributes normally associated with the concept of clustering include these:

- Simplified single system management
- High availability and continuous availability
- High-speed interconnect communication
- Scalability and flexibility
- Workload balancing
- Single system image
- Shared resources

Note: Small outages, tolerated just a few years ago, can now mean a significant loss of revenue and of future opportunities for a business. The most important aspect of clustering is high availability (that is, the ability to provide businesses with resilient resources).

A cluster, device domain, device CRG, and device description are configuration objects used to implement independent ASPs or clusters. Figure 3-2 illustrates the inter-relationship of each IASP and cluster configuration object.

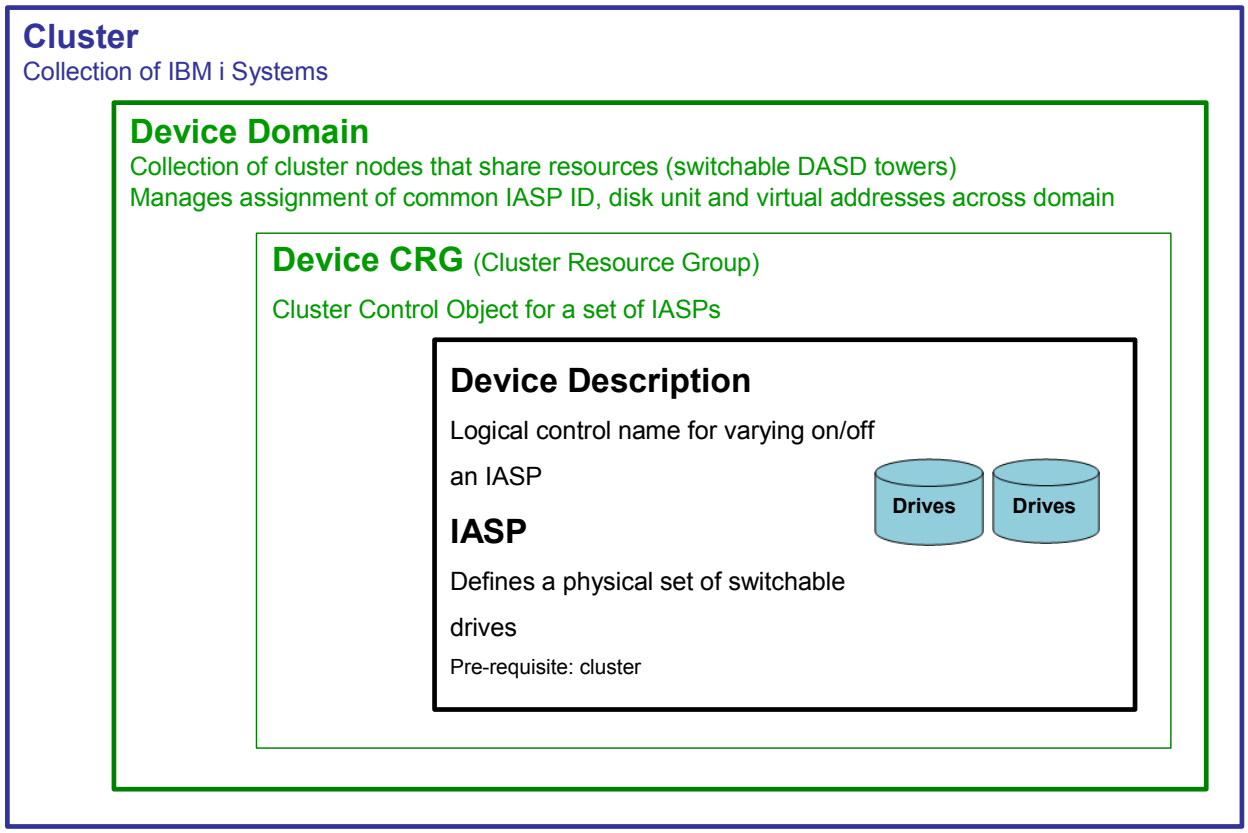


Figure 3-2 Switchable IASP object relationship

Base cluster functions

Several basic IBM i cluster functions monitor the systems within the cluster to detect and respond to potential outages in the high-availability environment. Cluster resource services provide a set of integrated services that maintain cluster topology, perform heartbeat monitoring, and allow creation and administration of cluster configuration and cluster resource groups. Cluster resource services also provides reliable messaging functions that keep track of each node in the cluster and ensure that all nodes have consistent information about the state of cluster resources:

- Heartbeat monitoring

  Heartbeat monitoring is an IBM i cluster base function that ensures that each node is active by sending a signal from every node in the cluster to every other node in the cluster to convey that they are still active.

- Reliable message function

  The reliable message function of cluster resource services keeps track of each node in an IBM i cluster and ensures that all nodes have consistent information about the state of cluster resources.

Why you want clustering

The concept of high availability in the sense of disaster recovery is an important consideration. However, disasters are not the only reason why high availability is important.

Disasters or unplanned outages account for only 20% of all outages. The majority of outages consist of planned ones, such as a shutdown to perform an upgrade or complete a total system backup. A relatively straightforward action, like the backup of databases and other objects, accounts for 50% of all planned outages.

Clusters are a very effective solution for continuous availability requirements on an IBM i system or logical partition, providing fast recovery for the widest range of outages possible, with minimal cost and overhead.

Some people might think that a backup of the server is not an outage. But IBM i users are not interested in such technicalities. If access to their data on the system is not possible, the user is most concerned about when the system is available again so that work can continue.

IBM i *clustering technology* offers you state-of-the-art and easy-to-deploy mechanisms to put your business on the path to continuous availability (Figure 3-3).

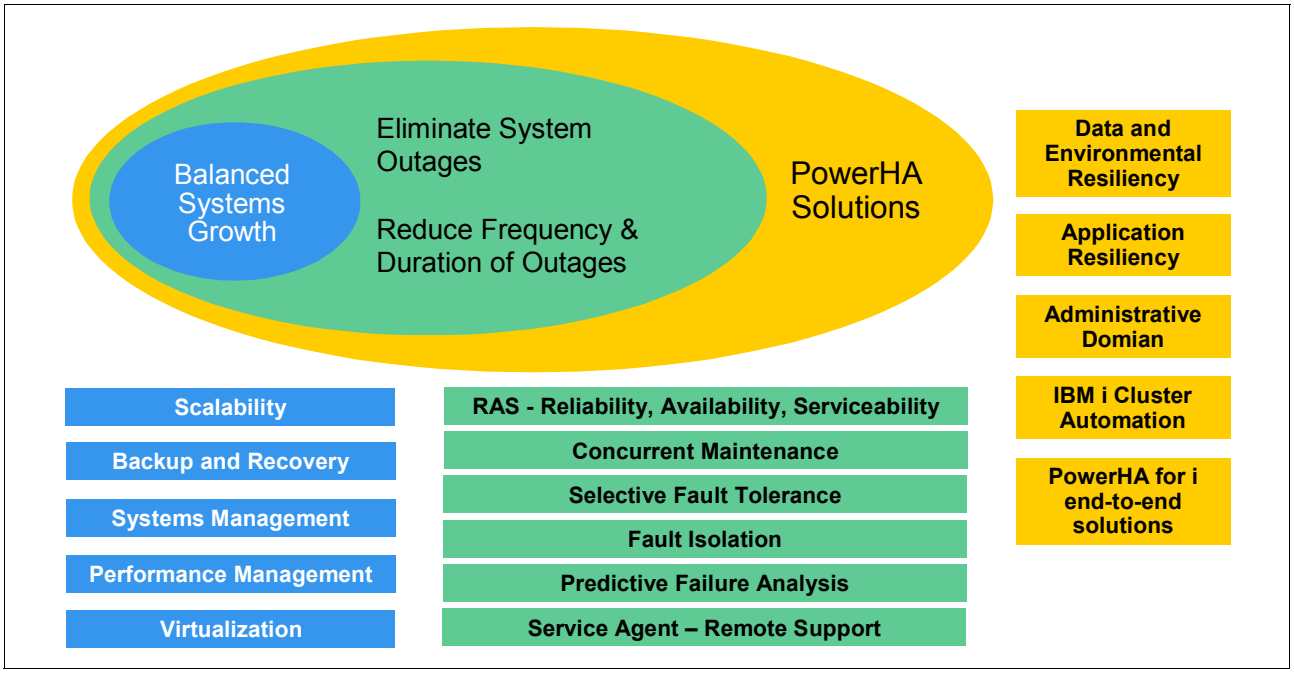


Figure 3-3 IBM i clustering technology state-of-the-art

3.2 Cluster nodes

A cluster node is any IBM i system or partition that is a member of a cluster. Cluster nodes must be interconnected on an IP network. A cluster node name is an eight-character cluster node identifier. Each node identifier is associated with one or two IP addresses that represent the system.

Any name can be given to a node. However, we recommend that you make the node name the same as the system name. Cluster communications that run over IP connections provide the communications path between cluster services on each node in the cluster. The set of cluster nodes that is configured as part of the cluster is referred to as the cluster membership list.

A cluster consists of a minimum of two nodes. The environment can be extended to a cluster with a maximum of 128 nodes.

A node of a cluster can fill one of three possible roles within a recovery domain. These are the roles and associated functions:

- Primary node
  - The point of access for a resilient device.
  - Contains the principal copy of any replicated resource.
  - The current owner of any device resource.
  - All CRG objects can fail over to a backup node.
- Backup node
  - Can take over the role of primary access at failure of the current primary node.
  - Contains a copy of the cluster resource.
  - Copies of data are kept current via replication.
- Replicate node
  - Has copies of cluster resources.
  - Unable to assume the role of primary or backup (typically used for functions such as data warehousing).

3.3 Device domain

A device domain is the first of the cluster constructs to be defined when creating a switchable IASP. It is a logical construct within Cluster Resource Services that is used to ensure that there are no configuration conflicts that prevent a switchover or failover.

The device domain is a subset of cluster nodes.

The set of configuration resources associated with a collection of resilient devices can be switched across the nodes in the device domain. Resource assignments are negotiated to ensure that no conflicts exist. The configuration resources assigned to the device domain must be unique within the entire device domain. Therefore, even though only one node can use a resilient device at any given time, that device can be switched to another node and brought online (Figure 3-4).

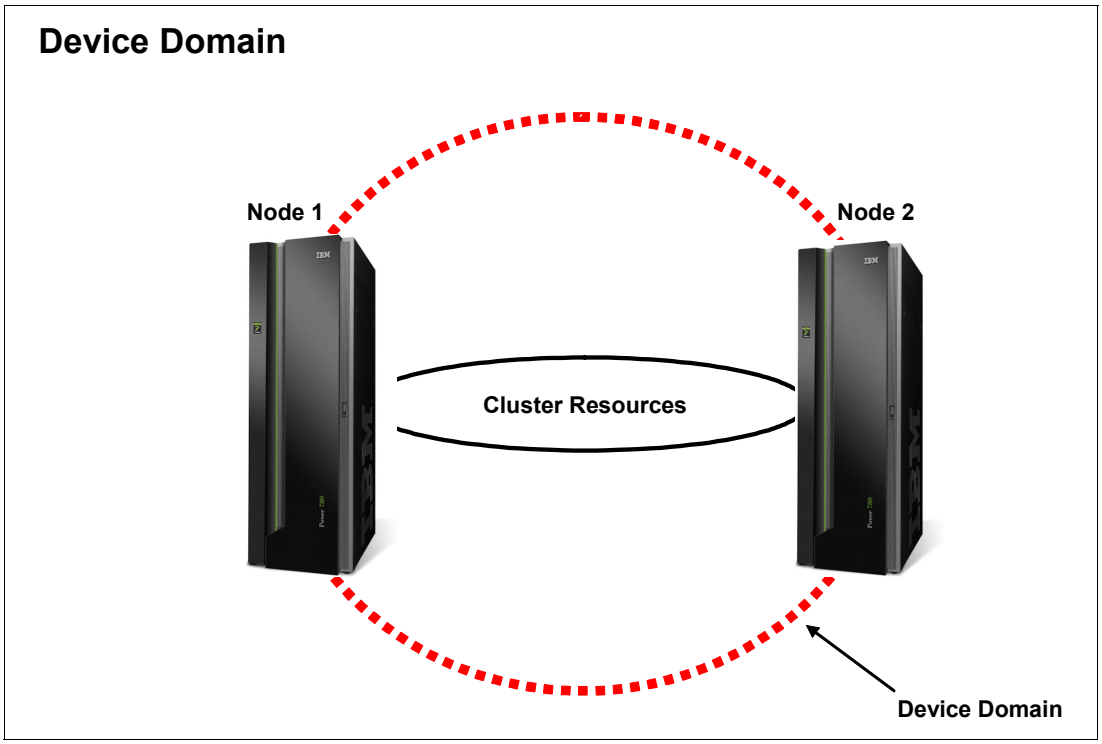


Figure 3-4 Device domain

These cluster resources are negotiated across a device domain to ensure that there are no conflicts:

- IASP number assignments

  IASPs are automatically assigned a number to correlate the name of the IASP. The user chooses the resource name. The system manages the assigned IASP numbers, which might not be in numerical order. The order depends on a number of factors, including the creation date and the creation of IASPs on other nodes in the device domain.

- DASD unit number assignments

  To keep from conflicting with the permanently attached disk units of each node, all IASP unit numbers begin with a four.

- Virtual address assignments

  The cluster configuration determines the virtual address space required for the IASP. Virtual address assignments (the cluster configuration) are ensured not to conflict across all nodes in the device domain.

Note: The collection of switched disks, the independent disk pool identification, disk unit assignments, and virtual address assignments must be unique across the entire device domain.

3.4 Cluster resource group

A cluster resource group is an IBM i system object that is a set or grouping of cluster resources. The cluster resource group (and replication software) is a foundation for all types of resilience.

Resources that are available or known across multiple nodes within the cluster are called cluster resources. A cluster resource can conceptually be any physical or logical entity (that is, database, file, application, device). Examples of cluster resources include IBM i objects, IP addresses, applications, and physical resources. When a cluster resource persists across an outage, that is any single point of failure within the cluster, it is known to be a resilient resource. As such, the resource is resilient to outages and accessible within the cluster even if an outage occurs to the node currently hosting the resource.

Cluster nodes that are grouped together to provide availability for one or more cluster resources are called the recovery domain for that group of cluster resources. A recovery domain can be a subset of the nodes in a cluster, and each cluster node might participate in multiple recovery domains. Resources that are grouped together for the purposes of recovery action or accessibility across a recovery domain are known as a cluster resource group (Figure 3-5).

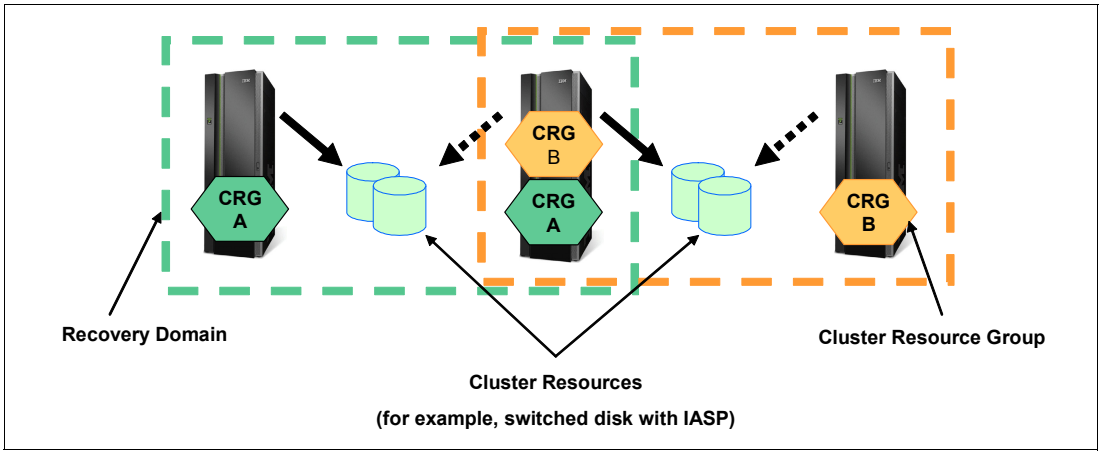


Figure 3-5 Cluster resource group

There are four cluster resource group (CRG) object types that are used with Cluster Services at V7R1:

- Application CRG

  An application CRG enables an application (program) to be restarted on either the same node or a different node in the cluster.

- Data CRG

  A data CRG enables data resiliency so that multiple copies of data can be maintained on more than one node in a cluster.

- Device CRG

  A device CRG enables a hardware resource to be switched between systems. The device CRG is represented by a (device) configuration object as a device type of independent ASP (IASP).

- Peer CRG

  A peer CRG is a non-switchable cluster resource group in which each IBM i node in the recovery domain plays an equal role in the recovery of the node. The peer cluster resource group provides peer resiliency for groups of objects or services. It is used to represent the cluster administrative domain. It contains monitored resource entries, for example, user profiles, network attributes, or system values that can be synchronized between the nodes in the CRG.

The cluster resource group defines the recovery or accessibility characteristics and behavior for that group of resources. A CRG describes a recovery domain and supplies the name of the cluster resource group exit program that manages cluster-related events for that group. One such event is moving the users from one node to another node in case of a failure.

Recovery domain

A recovery domain is a subset of nodes in the cluster that are grouped together in a cluster resource group for purposes such as performing a recovery action. Each cluster resource group has a recovery domain that is a subset of the nodes in the cluster. Here are facts about recovery domains:

- The nodes within a recovery domain participate in any recovery actions for the resources of the domain.
- Different CRGs might have different recovery domains.
- As a cluster goes through operational changes (for example, nodes end, nodes start, nodes fail), the current role of a node might change. Each node has a preferred role that is set when the CRG is created.
- A recovery domain can be a subset of the nodes in a cluster, and each cluster node might participate in multiple recovery domains.

CRG exit programs

In IBM i high availability environments, cluster resource group exit programs are called after a cluster-related event for a CRG occurs and responds to the event.

An exit program is called when a CRG detects certain events, such as a new node being added to the recovery domain, or the current primary node failing. The exit program is called with an action code indicates what the event is. Furthermore, the exit program has the capability to indicate whether to process the event. *User-defined* simply means that the IBM i cluster technology does not provide the exit program. Typically the exit program is provided by the application or data replication provider. The exit program is the way that a CRG communicates cluster events to the exit program provider. The exit program can perform the appropriate action based on the event, such as allowing a resource access point to move to another node. The exit program is optional for a resilient device CRG but is required for the other CRG types. When a cluster resource group exit program is used, it is called on the occurrence of cluster-wide events.

For detailed information about the cluster resource group exit programs, including what information is passed to them for each action code, see:

http://publib.boulder.ibm.com/infocenter/iseries/v7r1m0/topic/apis/clrgexit.htm

3.5 Advanced node failure detection

There have been sudden cluster node outages, such as a main storage dump, HMC immediate partition power-off, or a system hardware failure, which so far resulted in a partitioned cluster. In this case, the user is alerted with a failed cluster communication message CPFBB22 sent to QHST and an automatic failover not started message CPFBB4F sent to the QSYSOPR message queue on the first backup node of the CRG.

With IBM i 7.1, PowerHA SystemMirror for i now allows advanced node failure detection by cluster nodes. This is done by registering with an HMC or Virtual I/O Server (VIOS) management partition on IVM-managed systems. This way clustering gets notified in case of a severe partition or system failure to trigger a cluster failover event instead of causing a cluster partition condition.

If a node detects a failure, it notifies the other nodes in the cluster via a distress message. This triggers a failover if the failing node is the primary node of a CRG. If instead we get a heartbeating failure, then a partition occurs. Partitions are usually the result of network failures, and the nodes automatically merge after the problem has been fixed. However, there are cases in which the node fails too quickly for a distress message to go out, which can result in a “false” partition. The user can use **CHGCLUNODE** to change a partitioned node to a failed node. Through the use of advanced node failure detection, the occurrences of false partitions are reduced.

For LPAR failure conditions it is the POWER® Hypervisor™ (PHYP) that notifies the HMC that a LPAR failed. For system failure conditions other than a sudden system power loss, it is the flexible service processor (FSP) that notifies the HMC of the failure. The CIM server on the HMC or VIOS can then generate an IBM Power state change CIM event for any registered CIM clients.

Important: For systems that can be managed with a Hardware Management Console (HMC), the Common Information Model (CIM) server runs on the HMC. The HMC affords the most complete node failure detection because it is not part of the system and thus can continue to operate when a system has completely failed. When using VIOS, the CIM server must be started in the VIOS management partition by running **startnetsvc cimserver** in the VIOS partition.

Whenever a cluster node is started, for each configured cluster monitor IBM i CIM client APIs are used to subscribe for the particular power state change CIM event. The HMC CIM server generates such a CIM event and actively sends it to any registered CIM clients (that is, there is no heartbeat polling involved with CIM). On the IBM i cluster nodes the CIM event listener compares the events with available information about the nodes constituting the cluster to determine if it is relevant for the cluster to act upon. For relevant power state change CIM events, the cluster heartbeat timer expiration is ignored (that is, IBM i clustering immediately triggers a failover condition in this case).

Using advanced node failure detection requires SSH and CIMOM TCP/IP communication to be set up between the IBM i cluster nodes and the HMC or VIOS. Also, a cluster monitor needs to be added to the IBM i cluster nodes, for example, through the new **ADDCLUMON** command (Figure 3-6), which enables communication with the CIM server on the HMC or VIOS.

```
                        Add Cluster Monitor (ADDCLUMON)

Type choices, press Enter.

Cluster  . . . . . . . . . . . . > PWRHA_CLU     Name
Node identifier  . . . . . . . . > CTCIHA9V      Name
Monitor type . . . . . . . . . .   *CIMSVR       *CIMSVR
CIM server:
  CIM server host name . . . . . > HMC1
_____________________________________________________________________________
_____________________________________________________________________________
_____________________________________
  CIM server user id . . . . . . > hmcuser
  CIM server user password . . . > password




                                                                        Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 3-6 IBM i ADDCLUMON command

Note: Notification of failures by an HMC or VIOS depends on a TCP/IP application server running on the cluster node that is to receive the notification. If the application server is not running, the advanced node failure detection is not aware of node failures. The application server must be started and left running anytime that the cluster node is active. Use **STRTCPSVR *CIMOM CL** to start the application server.

4

PowerHA architecture

This chapter provides information on the basic concepts and mechanisms used in clusters built with IBM PowerHA SystemMirror for i.

4.1 PowerHA technologies

In this chapter we discuss the technologies of exchanging application data between PowerHA cluster nodes.

Clustering solutions build on PowerHA, based on the technologies of the physical exchange of the data included on the independent ASP (iASP). These are the mechanisms used by PowerHA described here:

- Switched disks
- Host-based replication
- Storage-based replication
- Administrative domain

Figure 4-1 and Figure 4-2 on page 47 provide information about PowerHA SystemMirror feature availability in relation to the hardware being used.

	Internal SAS/SSD	DS5000	DS6000	DS8000
	POWER5/6/7	**POWER6/7**	**POWER5/6/7**	**POWER5/6/7**
PowerHA SystemMirror 6.1 or 7.1				
FlashCopy	No	No	**Yes**	**Yes**
Metro Mirror	No	No	**Yes**	**Yes**
Global Mirror	No	No	**Yes**	**Yes**
Switched IASP	**Yes**	**Yes**	**Yes**	**Yes**
LUN Level Switching	No	No	**Yes** (7.1)	**Yes** (7.1)
Geographic Mirroring	**Yes**	**Yes**	**Yes**	**Yes**
PowerHA SystemMirror 6.1 or 7.1 *plus* Advanced Copy Services (ACS)				
FlashCopy	No	**Yes**	**Yes**	**Yes**
Metro Mirror	No	**Yes**	**Yes**	**Yes**
Global Mirror	No	**Yes**	**Yes**	**Yes**
LUN Level Switching	No	**Yes**	**Yes** (6.1)	**Yes** (6.1)
Metro/Global Mirror	No	No	**Yes**	**Yes**
External Storage Full System Copy (*crash consistent copy and cloning*)				
FlashCopy	No	**Yes**	**Yes**	**Yes**
Global Mirror	No	**Yes**	**Yes**	**Yes**
Metro Mirror	No	**Yes**	**Yes**	**Yes**
Logical Replication Add-on Software				
iCluster and others	**Yes**	**Yes**	**Yes**	**Yes**

Figure 4-1 IBM i native attach storage and resiliency

	DS3000	DS4000	DS5000	DS6000	DS8000	XIV	SVC / V7000
	BladeCenter	**POWER6/7 BladeCenter**	**POWER6/7 BladeCenter**	**POWER6/7 BladeCenter**	**POWER6/7 BladeCenter**	**POWER6/7 BladeCenter**	**POWER6/7 BladeCenter**
PowerHA SystemMirror 6.1 or 7.1							
FlashCopy	No	No	No	No	**Yes**[1]	No	**Yes**[2]
Metro Mirror	No	No	No	No	**Yes**[1]	No	**Yes**[2]
Global Mirror	No	No	No	No	**Yes**[1]	No	**Yes**[2]
Switched IASP	No	No	No	No	No	No	No
LUN Level Switch	No	No	No	No	**Yes**[1]	No	No
Geo'mirroring	**Yes**	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**
PowerHA SystemMirror 6.1 or 7.1 *plus* Advanced Copy Services (ACS)							
FlashCopy	No	No	**Yes**[1]	No	**Yes**[1]	No	No
Metro Mirror	No	No	**Yes**[1]	No	**Yes**[1]	No	No
Global Mirror	No	No	**Yes**[1]	No	**Yes**[1]	No	No
LUN Level Switch	No	No	No	No	**Yes**[1]	No	No
External Storage Full System Copy *(crash consistent copy and cloning)*[3]							
FlashCopy	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Metro Mirror	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Global Mirror	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Logical Replication Add-on Software							
iCluster, et. al.	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

1 Requires NPIV capable fiber channel adapter. DS5000 NPIV support requires IBM i 7.1 TR2
2 Requires IBM i 7.1 TR3 and PRPQ 5799-HAS
3 See Redbook "IBM i and Midrange External Storage SG247668" and "IBM i Virtualization and DS4000 Read-me First"

Figure 4-2 IBM i PowerVM® VIOS storage and resiliency

4.1.1 Switched disks

Switched disk is a solution based on the concept of sharing the same pool of disks (iASP) between different systems.

To illustrate this idea we modified the generic cluster model to the one shown in Figure 4-3.

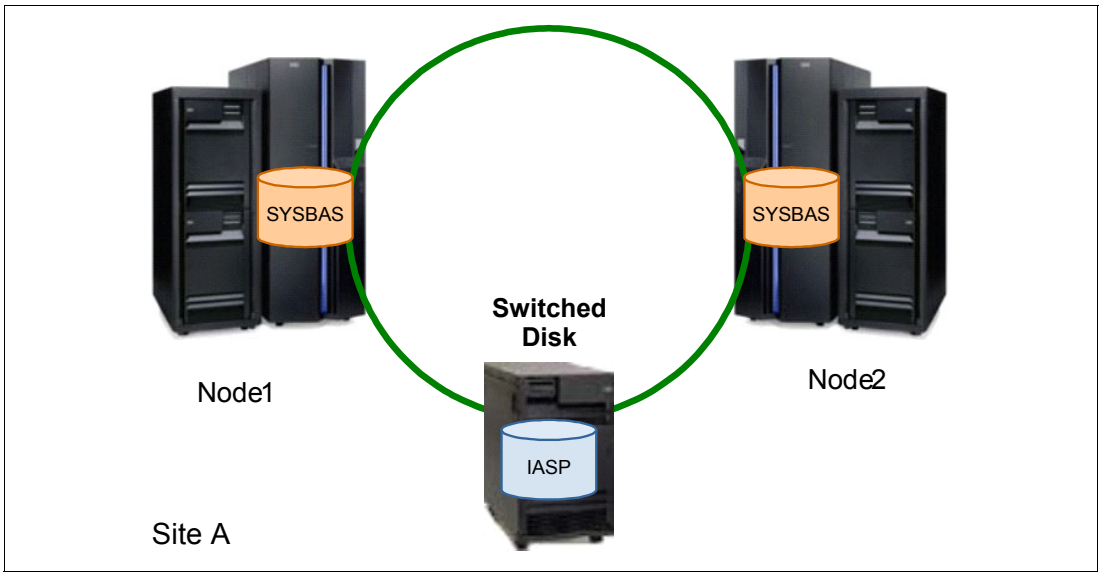


Figure 4-3 Model of switched disk solution

These are the types of switched disk solutions:

- Switching disks between two separate systems using the commonHSL loop
- Switching disks between two LPARs in the same system

Note: POWER6® server is the last one that uses the HSL loop and the last server where switching disks build on this technology is supported.

All solutions are the same from the logical standpoint.

Using the switched disk solution we have a single IASP that can be owned by one or the other node in the cluster at a given time. During the switch process the switchable IASP changes its owner and all the data on it is accessible by the other node.

These are the advantages of this mechanism:

- Simplicity
- Low cost

These are the disadvantages of this mechanism:

- HA within single POWER system: When using the HSL loop or switching between LPARs in the same system. When using switched LUNs it is possible to switch the disks between different systems.
- Low resiliency to data loss: You need to implement disk redundancy at the storage level.

Switched disks are not advised to be a single HA solution for a production system. However they can be a complementary element of more sophisticated solutions (for example, a 3-node solution with storage-based replication).

Note: The data on the switched disks is in one copy, so it should be protected on the disk level (for example, by one of the RAID mechanisms).

4.1.2 Host-based replication (geographic mirroring)

Host-based replication is used to keep consistent copy of an iASP assigned to one of the cluster nodes onto another iASP assigned to another node. The processes that keep the iASP in synchronization run on the nodes that own the iASPs. In most cases the TCP/IP network is being used for sending the changes to backup iASP.

In the case of the PowerHA System Mirror for i, the host-based replication feature is called geographic mirroring.

Figure 4-4 shows the idea of host-based replication.

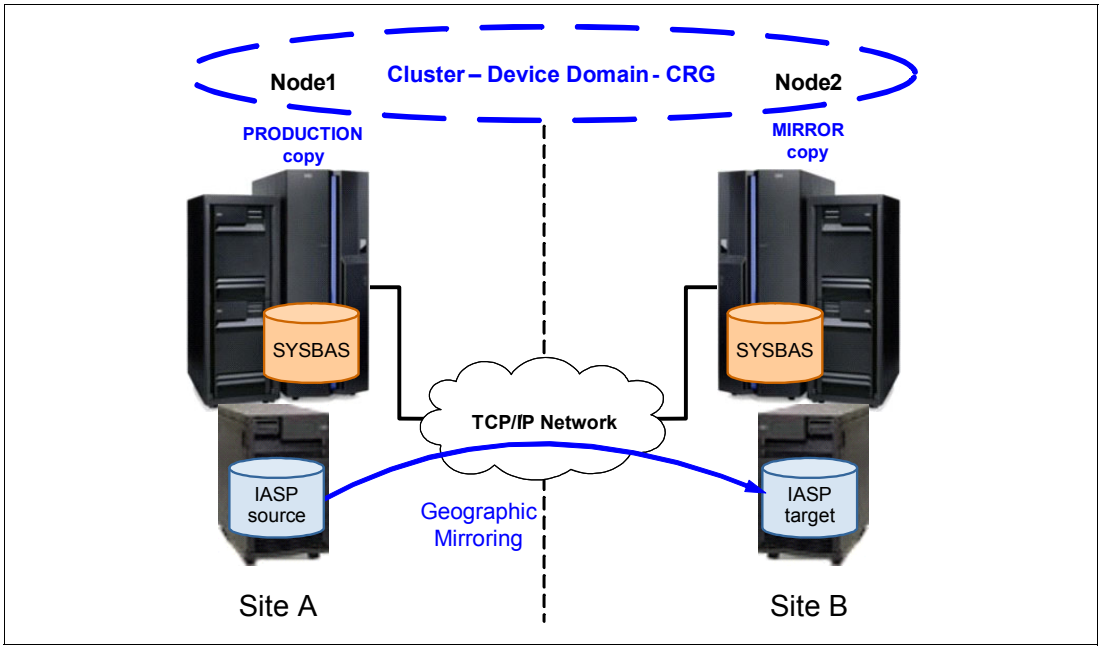


Figure 4-4 Host-based replication in PowerHA: Geographic mirroring

Note: Notice that iASPs can be built on internal or external disk drives.

Geographic mirroring guarantees that the changes to the replicated iASP will be applied to the target copy in the original order, because in case of failure the target copy will be usable and accessible by the node that owns it.

Note: It must be stressed that the commitment control must be in place to guarantee the database consistency in case of unexpected failure.

Geographic mirroring introduces high-availability protection in case of the server failure, and in case the servers are in a different location, it allows protection in case of a site outage.

The copy owned by the primary node is the production copy and the copy owned by the backup system at the other site is the mirror copy. Users and applications can only access the independent disk pool on the primary node, the node that owns the production copy. Changes that are made to the production copy (source system) are guaranteed by the geographic mirroring functionality to be made in the same order on the mirror copy (target system).

Geographic mirroring allows for the production and mirrored copies to be in the same site for high-availability protection in the event of server failure. It is also possible to separate the two systems geographically for disaster recovery protection in the event of a site-wide outage, provided that the communication link between the two sites is fast enough. In the case of using synchronous delivery mode, communication speed and throughput have an impact on application response time on the production system. This is due to the fact that the production system waits until a write operation has at least reached main memory on the backup system and the backup system has sent a confirmation back to the production system before a local write to the iASP of the production system is considered finished. PowerHA for i Enterprise additionally offers the asynchronous mode of delivery of the changes to the remote IASP. In this case the network delivery time does not affect the application performance. However, the changes are not guaranteed to be delivered to the backup system in case of the unexpected failure of the primary node.

Synchronous and asynchronous modes of operation

When using geographic mirroring replication for iASP, there are two parameters that are related to synchronism in the data replication:

- Transmission delivery

  This parameter is available when using PowerHA for System i Enterprise. It allows you to decide whether the data will be delivered to the adjacent node in synchronous or asynchronous mode. Synchronous mode of delivery requires that when the data is received by the backup node it sends the confirmation to the primary, and after the confirmation is received the operation is confirmed. In asynchronous mode of delivery the data is sent to the backup node and this completes the operation.

- Mirroring mode

  Mirroring mode of operation decides whether the data needs to be written to the disk on the backup node in order to have a completed status on the primary node. This is synchronous delivery mode. If the mirroring mode is asynchronous, the operation is completed on the primary node when it is received by the backup node. It does not have to be put on disk before it is completed on the primary.

The relationship between the operations in synchronous and asynchronous modes of operations are shown in Figure 4-5, and the order of the operations for each type of delivery is shown in Table 4-1.

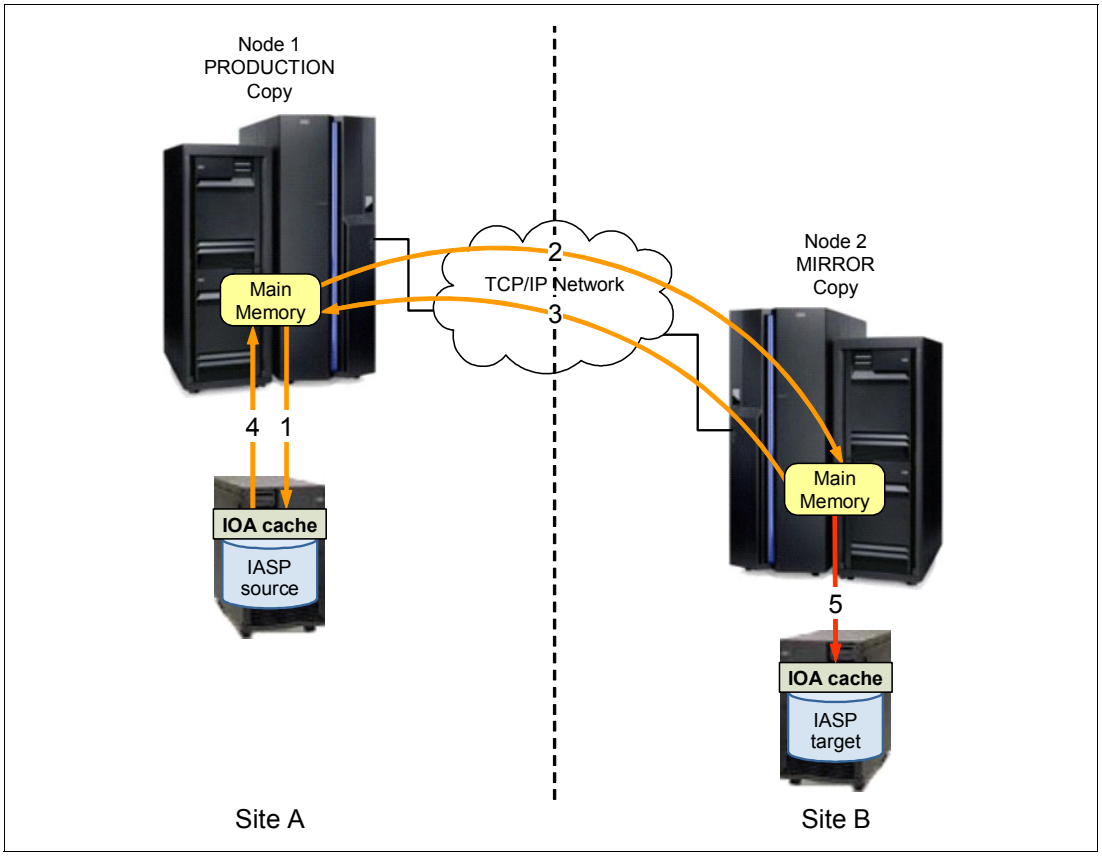


Figure 4-5 Geographic mirroring operations

Table 4-1 Geographic mirroring operations relationship

Transmission delivery	Mirroring mode	Local operations	Remote operations
*ASYNC	*ASYNC	1,4	2
*SYNC	*ASYNC	1,4	2,3
*SYNC	*SYNC	1,4	2,5,3

More details about the geographic mirroring can be found in Chapter 5, “Geographic Mirroring” on page 67.

Note: When using host-based replication in synchronous mode of operation, the network latency caused by the distance between the hosts can cause performance degradation of the applications using the replicated iASP. The mirroring mode can severely affect performance when the target system’s disk subsystem has lower performance than the source system.

Advantages of host-based replication

Geographic mirroring is the only physical replication that can be used for the IASPs that are built on the internal disks subsystems. It has a relatively low cost of implementation due to the

fact that it is a software-based solution and does not require external storage devices to be in place for replication.

Disadvantages

The disadvantage of this solution can be the fact that it uses the host resources, such as CPU and memory. That might affect performance of other processes running on the system.

4.1.3 Storage-based replication

In contrast to host-based replication (geographic mirror), storage-based replication is done at the storage subsystem devices level. The types of the storage level replication supported by IBM PowerHA SystemMirror for i are as follows:

- Metro Mirror
- Global Mirror
- FlashCopy
- LUN-level switching

Metro Mirror (synchronous)

Metro Mirror is a replication solution based on the synchronous data replication between external IBM Storage Systems connected to IBM i. This means that all the operations are acknowledged to the source system when they are completed on both storage susbsystems. Data between storage systems is replicated over the Storage Area Network (SAN), either using FC or FC over IP. Synchronous replication in case of Metro Mirror means that the source and a copy of the IASP is in a consistent state when an unexpected failure occurs on any of the cluster nodes.

When using a Metro Mirror replication with PowerHA System Mirror, each node has a locally attached external IBM Storage System, and the replicated IASP must be located in it. Figure 4-6 shows the most common configuration. The System ASP can reside on internal or external disks.

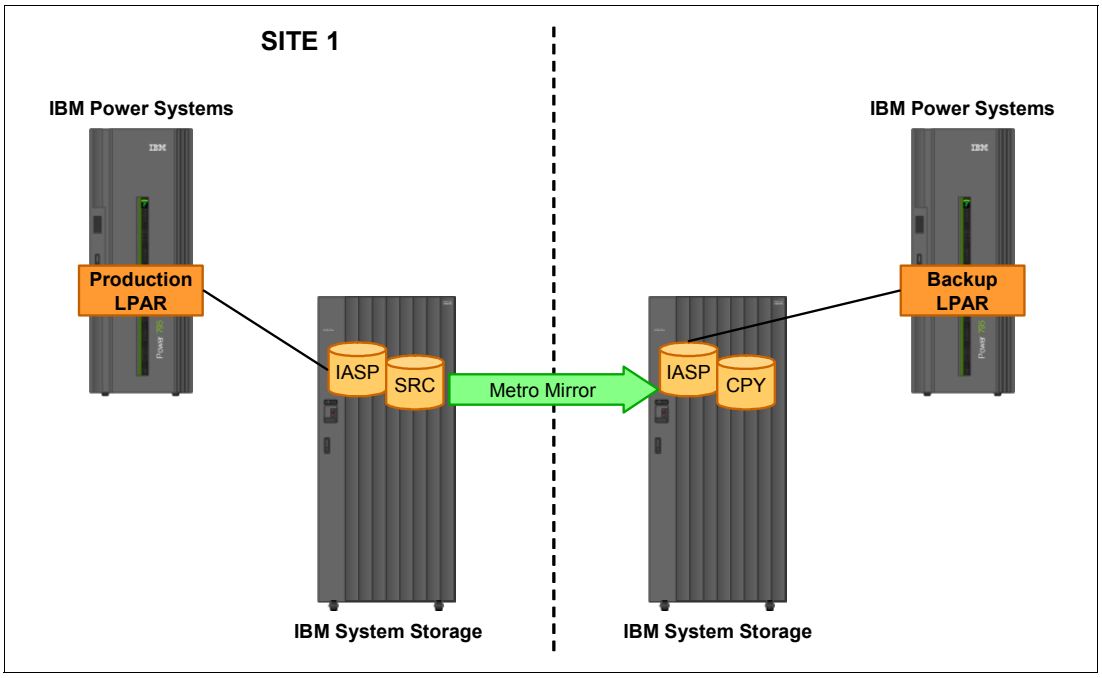


Figure 4-6 Metro Mirror architecture

Because of the synchronous mode of operation, the distance between the replication sites is limited to metro distances, which is about 30 kilometers. For more information about Metro Mirror see 6.2, “Metro Mirror” on page 93.

In addition to IBM TotalStorage devices you can build a Metro Mirror solution using IBM Storwize V7000 and SAN Volume Controller. For more details about this type of implementation see 7.2, “Metro Mirror” on page 119.

Global Mirror (asynchronous)

Global Mirror is an asynchronous replication between local and remote IBM Storage Systems. Similar to Metro Mirror, this replication is also controlled by storage systems and data is being replicated on the disk level. Due to asynchronous replication, this solution can be used for replication between storage systems located a long distance from each other because the delay introduced by the remote write is not affecting the local operations.

Global Mirror is based on these copy services functions of a Storage system:

- Global Copy
- FlashCopy
- FlashCopy consistency group

The use of FlashCopy and the FlashCopy consistency group with Global Copy allows you to maintain a consistent copy of the IASP in the backup site. Figure 4-7 shows the general architecture for the Global Mirror replication.

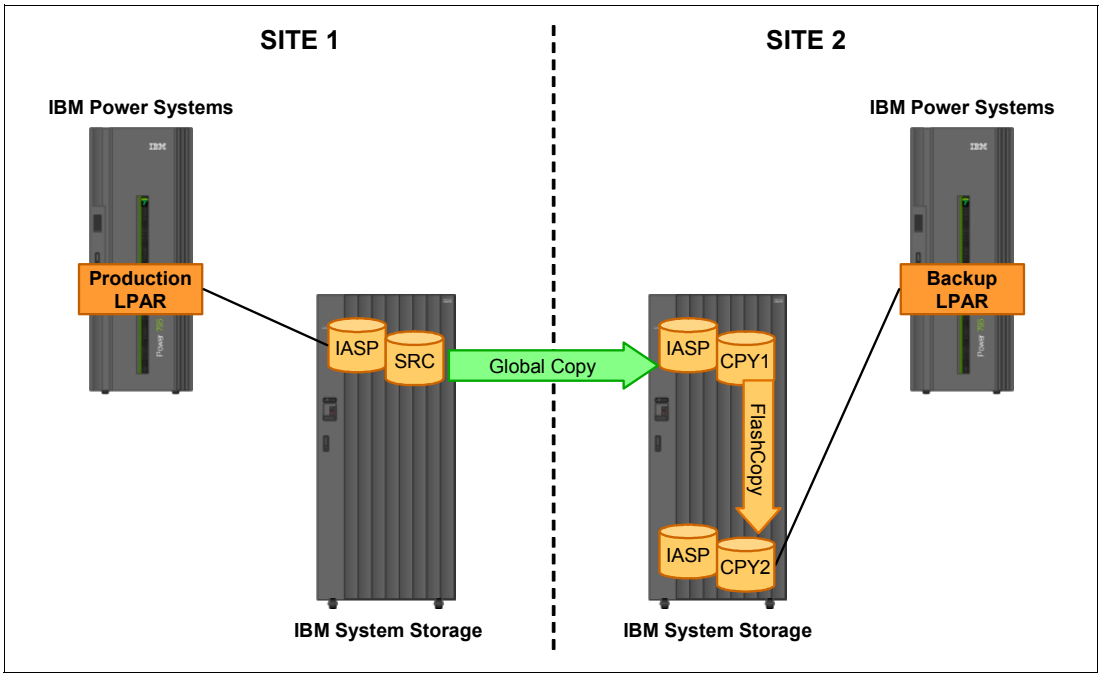


Figure 4-7 Global Mirror architecture

How it works

Global Mirror, as a long-distance remote copy solution, is based on an efficient combination of Global Copy and FlashCopy functions. It is the Storage system microcode that provides, from the user perspective, a transparent and autonomic mechanism to intelligently utilize Global Copy in conjunction with certain FlashCopy operations to attain consistent data at the remote site.

For more details about the Global Mirror solution see 6.3, “Global Mirror” on page 98, which describes the Global Mirror for DS8000 Copy Services.

Global Mirror can be used in solutions based on IBM Storwize V7000 and SAN Volume Controller. See 7.3, “Global Mirror” on page 123.

LUN-level switching

LUN-level switching is a new function provided by PowerHA SystemMirror for i. It allows you to switch a set of LUNs (a volume group) between systems. The idea of this solution is similar to the switched disks described in 4.1.1, “Switched disks” on page 47. For more details about this solution, see 6.4, “LUN-level switching” on page 105.

FlashCopy

FlashCopy allows you to create a point-in-time copy of the logical volume. By doing a FlashCopy, a relationship is *established* between a source volume and a target volume. Both are considered to form a FlashCopy *pair*. As a result of the FlashCopy, either all physical blocks from the source volume are copied (when using the copy option) to the target volume, or when using the nocopy option. Only those parts are copied that are changing in the source data since the FlashCopy was established. The target volume needs to be the same size or bigger than the source volume whenever FlashCopy is used to flash an entire volume. Figure 4-8 shows an example of the architecture for the FlashCopy. In this example, when the production IASP FlashCopy is established, the copy can be accessed by another LPAR. The copy of the IASP is read and write capable, so you can use it for backup or testing purposes.

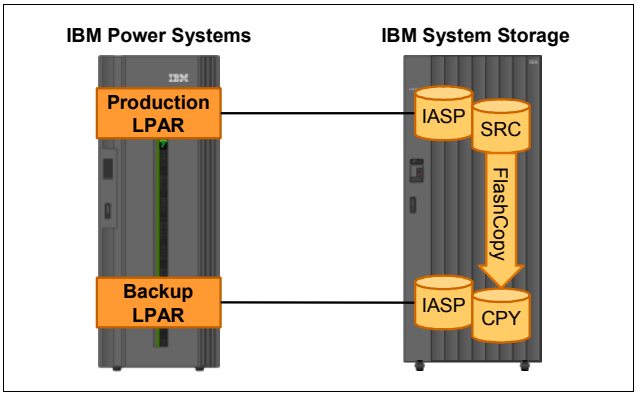


Figure 4-8 FlashCopy architecture example

Within PowerHA for i, the classic FlashCopy and FlashCopy SE are supported. When using classic FlashCopy, the target volume size that is reported to the system must be allocated in the TotalStorage device.

When using FlashCopy SE (Space Efficient) you can create virtual volumes that are space efficient and use them as a target for the FlashCopy operation. The space efficient volumes are volumes that have a defined virtual size, but the physical storage is not physically allocated to them.

Typically, large applications such as databases have their data spread across several volumes, and their volumes should all be FlashCopied at exactly the same point-in-time. FlashCopy offers consistency groups, which allows multiple volumes to be FlashCopied at exactly the same instance.

4.1.4 Administrative domain

A *cluster administrative domain* provides a mechanism for maintaining a consistent operational environment across cluster nodes within an IBM i high availability environment. A cluster administrative domain ensures that highly available applications and data behave as expected when switched to or failed over to backup nodes.

There are often configuration parameters or data associated with applications and application data, which are known collectively as the operational environment for the application. Examples of this type of data include user profiles used for accessing the application or its data, or system environment variables that control the behavior of the application. With a high-availability environment, the operational environment needs to be the same on every system where the application can run, or where the application data resides. When a change is made to one or more configuration parameters or data on one system, the same change needs to be made on all systems. A cluster administrative domain lets you identify resources that need to be maintained consistently across the systems in an IBM i high availability environment. The cluster administrative domain then monitors for changes to these resources and synchronizes any changes across the active domain.

When a cluster administrative domain is created, the system creates a peer CRG with the same name. The nodes that make up the cluster administrative domain are defined by the CRGs recovery domain. Node membership of the cluster administrative domain can be modified by adding and removing nodes from the recovery domain using these commands:

- Add Admin Domain Node Entry (**ADDCADNODE**).
- Remove Admin Domain Node Entry (**RMVCADNODE**).
- Work with Cluster (**WRKCLU**).

Each cluster node can be defined in only one cluster administrative domain within the cluster.

Note: To work with cluster CL commands or the Cluster Resource Services graphical interface, you must have IBM PowerHA SystemMirror for i licensed program installed.

After the cluster administrative domain is created, it can be managed with CL commands or the Cluster Resource Services graphical interface in IBM Systems Director Navigator for i.

The type of objects that can be managed in a cluster administrative domain, also known as monitored resources, have been enhanced in IBM i 7.1. See Table 4-1 on page 51:

Table 4-2 Monitored resource entry type support

Object or attribute description	Type
Authorization lists (*)	*AUTL
Classes	*CLS
Ethernet line descriptions	*ETHLIN
Independent disk pools device descriptions	*ASPDEV
Job descriptions	*JOBD
Network attributes	*NETA
Network server configuration for connection security	*NWSCFG
Network server configuration for remote systems	*NWSCFG
Network server configurations for service processors	*NWSCFG

Object or attribute description	Type
Network server descriptions for iSCSI connections	*NWSD
Network server descriptions for integrated network servers	*NWSD
Network server storage spaces	*NWSSTG
Network server host adapter device descriptions	*NWSHDEV
Optical device descriptions	*OPTDEV
Printer device descriptions for LAN connections*	*PRTDEV
Printer device descriptions for virtual connections*	*PRTDEV
Subsystem descriptions	*SBSD
System environment variables	*ENVVAR
System values	*SYSVAL
Tape device descriptions	*TAPDEV
Token-ring line descriptions	*TRNLIN
TCP/IP attributes	*TCPA
User profiles	*USRPRF

* Available from IBM i 7.1. PowerHA SystemMirror for i is required to support these new administration domain monitored resource entries.

Monitored resources

A monitored resource is a system resource that is managed by a cluster administrative domain. Changes made to a monitored resource are synchronized across nodes in the cluster administrative domain and applied to the resource on each active node. Monitored resources can be system objects such as user profiles or job descriptions. A monitored resource can also be a system resource not represented by a system object, such as a single system value or a system environment variable. These monitored resources are represented in the cluster administrative domain as monitored resource entries (MREs).

A cluster administrative domain supports monitored resources with simple attributes and compound attributes. A compound attribute differs from a simple attribute in that it contains zero or more values, while a simple attribute contains a single value. Subsystem Descriptions (*SBSD) and Network Server Descriptions (*NWSD) are examples of monitored resources that contain compound attributes.

For MREs to be added, the resource must exist on the node from which the MREs are added. If the resource does not exist on every node in the administrative domain, the monitored resource is created. If a node is later added to the cluster administrative domain, the monitored resource is created. MREs can only be added to the cluster administrative domain if all nodes in the domain are active and participating in the group. MREs cannot be added in the cluster administrative domain if the domain has a status of Start of changePartitionedEnd of change.

You can add the MRE using **ADDCADMRE** or with the PowerHA GUI. The PowerHA GUI allows you to select the MRE's monitored parameters from a list, while in the command you need to specify them.

To remove MRE from the administrative domain, you can use **RMVCADMRE** or the PowerHA GUI.

Determine the status of the cluster administrative domain and the status of the nodes in the domain by using Cluster Resource Services graphical interfaces using these options:

- IBM Systems Director Navigator for i.
- Display CRG Information (DSPCRGINF).
- Work with Cluster (**WRKCLU**) commands.

Note: To use the Cluster Resource Services graphical interface or the Display CRG Information (**DSPCRGINF**) command, you must have the IBM PowerHA SystemMirror for i licensed program installed.

4.2 ASP copy descriptions

ASP copy descriptions are used by PowerHA to manage geographic mirroring, Metro Mirror, Global Mirror, and FlashCopy copies. The copy description defines all the parameters that are needed to access the disk units assigned to given iASP and execute Copy Services operations by PowerHA. Figure 4-9 shows the relations between management objects described in this section. The IASPs that are in the device domain for the cluster have the same ASP copy descriptions on all nodes of the cluster.

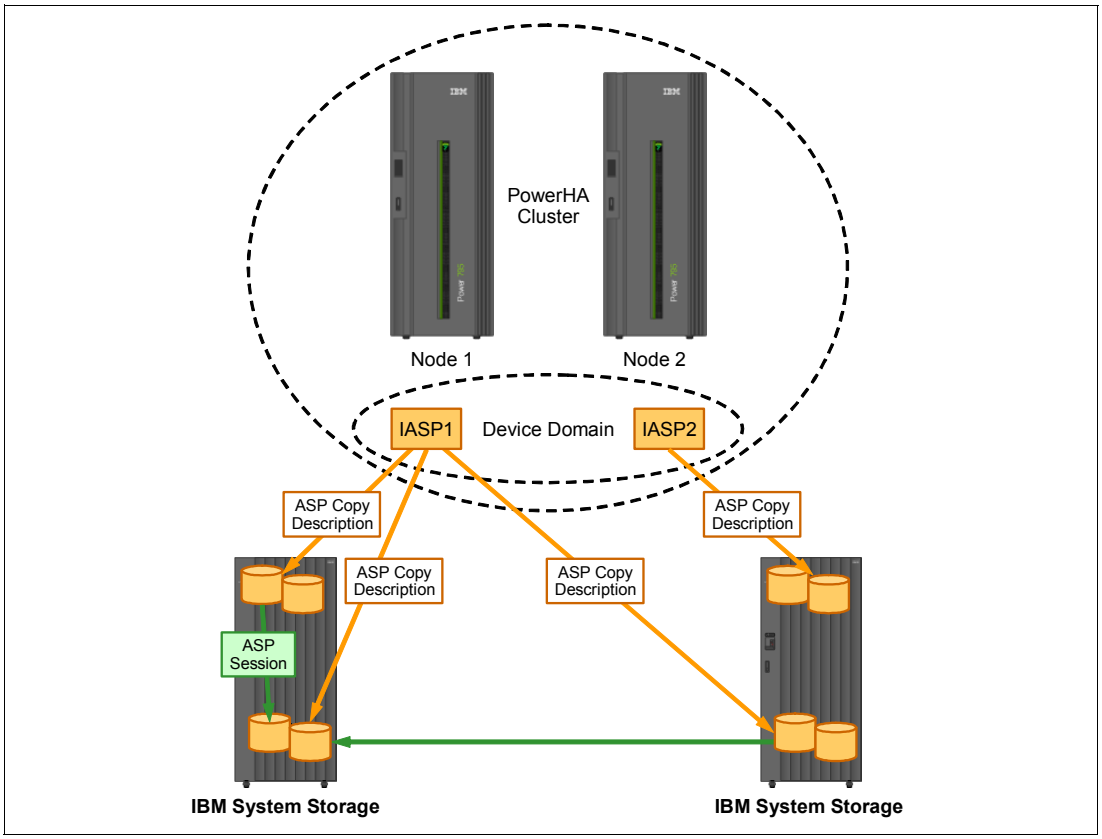


Figure 4-9 General view of the ASP copy descriptions and ASP sessions

Figure 4-10 shows an example of the ASP copy description being added to the system.

```
 Add ASP Copy Description (ADDASPCPYD)

 Type choices, press Enter.

 ASP copy . . . . . . . . . . . . . ASPCPY       > ASPCPYD1
 ASP device . . . . . . . . . . . . ASPDEV       > IASP1
 Cluster resource group . . . . . . CRG          > PWRHA_CRG1
 Cluster resource group site  . . . SITE         > SITE1
 Storage host:                      STGHOST
   User name  . . . . . . . . . .                > DS_USER
   Password . . . . . . . . . . .                > DS_PASSWORD
   Internet address . . . . . . .                > '10.10.10.10'



 Location . . . . . . . . . . . . . LOCATION     > NODE1

Logical unit name:                 LUN
   TotalStorage device  . . . . .                > 'IBM.2107-75AY032'
   Logical unit range . . . . . .                > 'A010-A013'
                            + for more values
   Consistency group range  . . .
                            + for more values
 Recovery domain:                   RCYDMN
   Cluster node . . . . . . . . .                  *NONE
   Host identifier  . . . . . . .
                            + for more values
   Volume group . . . . . . . . .
                            + for more values
                            + for more values
```

Figure 4-10 ADDASPCPYD example

Table 4-3 describes the parameters of this command.

Table 4-3 Parameters of the ASP copy description.

Parameter name	Description
ASPCPY	Defines the name of the ASP copy description. Needs to be unique in the cluster.
ASPDEV	This is the name of the IASP for which the ASP copy description is being created. One IASP can have many ASP copy descriptions.
CRG	Cluster resource group in which this ASP is being used. This parameter is valid for replication solutions (geographic mirroring, Metro Mirror, and Global Mirror) and for switched disks solution. This parameter is not used for FlashCopy purposes.

Parameter name	Description
SITE	Same as in the case of CRG, this parameter is for replication solutions and is needed for switching the IASP between the sites. This ASP copy description will be used when the cluster resource group will be switched to the site specified in this parameter.
STGHOST	This group of parameters defines the address, user ID, and password for the HMC controlling the IBM Data Storage system.
LOCATION	This specifies the node that is used for this ASP copy description. This is used for FlashCopy to define which node will be using this copy of IASP. In other cases it should be set to *DEFAULT.
LUN	This group defines which storage system to use and what LUNs to be used for this ASP copy description. This parameter is used to bind the IASP with storage on the IBM Data Storage system.
RCYDMN	Recovery domain used for switchable LUNs solution.

4.3 ASP sessions

An ASP session is used to link two ASP copy descriptions and start the Copy Services functions between them. The status of the ASP session describes the current status of the replication. Figure 4-11 shows an example of the status of the ASP session.

```
Display ASP Session                        NODE1
                                                          09/15/11  15:42:49
Session . . . . . . . . . . . . . . . . . . :   IASP1SSN
  Type  . . . . . . . . . . . . . . . . . . :     *METROMIR









                              Copy Descriptions

ASP
device         Name                 Role          State     Node
IASP1          IASP1CPYD1          SOURCE        UNKNOWN     NODE1
IASP1          IASP1CPYD2          TARGET         ACTIVE     NODE2



                                                                        Bottom
Press Enter to continue

F3=Exit    F5=Refresh    F12=Cancel    F19=Automatic refresh
```

Figure 4-11 ASP session example

4.3.1 Start/end

To start the session you can issue **straspssn** from the 5250 session (Figure 4-12 on page 61). Table 4-4 describes the parameters for this command.

Table 4-4 STRASPSSN command parameters

Parameter	Description
SSN	Session name
TYPE	Type of the session. The possible values are: ► *GEOMIR for the geographic mirroring session ► *METROMIR for the Metro Mirror session ► *GLOBALMIR for the Global Mirror session ► *FLASHCOPY for the FlashCopy session
ASPCPY	The names of the ASP copy descriptions that the session is started between.
FLASHTYPE	When you use type of the session *FLASHCOPY you can choose here whether is the FlashCopy will be COPY or NOCOPY type.

Parameter	Description
PERSISTENT	In case you use FlashCopy, you choose whether the relation should be persistent.

```
Start ASP Session (STRASPSSN)

Type choices, press Enter.

Session  . . . . . . . . . . . . SSN          > ASPSSN1
Session type . . . . . . . . . . TYPE         > *METROMIR
ASP copy:                        ASPCPY
  Preferred source . . . . . .                > ASPCPYD1
  Preferred target . . . . . .                > ASPCPYD2
                         + for more values
FlashCopy type . . . . . . . . . FLASHTYPE      *NOCOPY
Persistent relationship  . . . . PERSISTENT     *NO

                                                                      Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 4-12 STRASPSSN

The ASP session is also started as a result of making IASP highly available in the PowerHA cluster. To stop the session use this command:

```
ENDASPSSN SSN(IASP1SSN)
```

Or use the **wrkaspcpyd** option 24 next to the related ASP copy description. When the ASP session is ended, the relation between the ASP copy descriptions is removed.

4.3.2 Changing attributes

In addition to the previously mentioned starting and stopping of the ASP sessions, you can change some of the attributes of the session, including its status. Most of the changes described later in this section can be done using **CHGASPSSN**.

The Change Auxiliary Storage Pool Session (**CHGASPSSN**) command can be used to change an existing geographically mirrored, Metro Mirrored, Global Mirrored, or FlashCopy session.

Figure 4-13 is an example. Note that all values are not possible for all session types.

```
                      Change ASP Session (CHGASPSSN)

Type choices, press Enter.

Session  . . . . . . . . . . . . >              Name
Option . . . . . . . . . . . . .                *CHGATTR, *SUSPEND...
ASP copy:
  Preferred source . . . . . . .   *SAME        Name, *SAME
  Preferred target . . . . . . .   *SAME        Name, *SAME
  Consistency source . . . . . .   *SAME        Name, *SAME, *NONE
  Consistency target . . . . . .   *SAME        Name, *SAME, *NONE
              + for more values
Suspend timeout  . . . . . . . .   *SAME        60-3600, *SAME
Mirroring mode . . . . . . . . .   *SAME        *SAME, *SYNC, *ASYNC
Synchronization priority . . . .   *SAME        *SAME, *LOW, *MEDIUM, *HIGH
Tracking space . . . . . . . . .   *SAME        0-100, *SAME
FlashCopy type . . . . . . . . .   *SAME        *SAME, *COPY
Persistent relationship  . . . .   *SAME        *SAME, *YES, *NO
ASP device . . . . . . . . . . .   *ALL         Name, *ALL
              + for more values
                                                                    More...
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 4-13 Example of CHGASPSSN panel

To suspend or resume geographic mirroring, Metro Mirror, or Global Mirror through this command, you must have already configured the mirror copy disks and cluster resource group with a recovery domain that has two sites.

To change session attributes (the *CHGATTR parameter on the **CHGASPSSN** command) when using geographic mirroring, the production copy of the iASP must be varied off.

CHGASPSSN with the *Detach, *Reattach, *Suspend, and *Resume options can be confusing. It is hard to know which node you have to run which option from, in addition to what states the iASP copies must be in first. Table 4-5 provides a quick reference list. To simplify the chart we are letting *source* also mean *production copy* and *target* also mean *mirror copy*.

Table 4-5 What options can be run from where and what status the iASP copies must be in

CHGASPSSN option	Environment	Can run from source?	Can run from Target?	Source IASP must be varied off	Target IASP must be varied off
*Detach	Metro Mirror	Yes	No	Yes	Yes
*Reattach	Metro Mirror	No	Yes	No	Yes
*Suspend	Metro Mirror	Yes	Yes	No	Yes
*Resume	Metro Mirror	Yes	Yes	No	Yes
*Detach	Geographic Mirroring	Yes	No	No	Yes
*Reattach	Geographic Mirroring	Yes	No	No	Yes

CHGASPSSN option	Environment	Can run from source?	Can run from Target?	Source IASP must be varied off	Target IASP must be varied off
*Suspend	Geographic Mirroring	Yes	No	No	Yes
*Resume	Geographic Mirroring	Yes	No	No	Yes

CHGASPSSN *SUSPEND or *DETACH will offer the tracking *YES or *NO (Figure 4-14). However, the parameter is ignored if this is not a geographic mirroring solution. Both Metro Mirror and Global Mirror will track regardless of this option.

```
                      Change ASP Session (CHGASPSSN)

Type choices, press Enter.

Session  . . . . . . . . . . . . > ASPSSN        Name
Option . . . . . . . . . . . . . > *DETACH       *CHGATTR, *SUSPEND...
Track  . . . . . . . . . . . . .   *YES          *YES, *NO

                                                                      Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

*Figure 4-14 CHGASPSSN *DETACH for geographic mirroring*

CHGASPSSN with geographic mirror *will* allow a *DETACH if the iASP is varied on. Because the iASP is in use, this means that the mirror copy will have to go through an abnormal vary-on, and there are risks to your data.

CHGASPSSN with Metro Mirroring will *not* allow a *DETACH if the iASP is varied on. If you try you will get a CPD26B9 (Figure 4-15).

```
                      Additional Message Information

Message ID . . . . . . :   CPD26B9       Severity . . . . . . . :   40
Message type . . . . . :   Diagnostic
Date sent  . . . . . . :   05/12/08      Time sent  . . . . . . :   09:44:35

Message . . . . :   Device METRO must be varied off for this change.
Cause . . . . . :   The requested changes cannot be made on device METRO while
  it is varied on.
Recovery  . . . :   Vary off the device (VRYCFG command).  Then try the
  request again.
```

Figure 4-15 msgCPD26B9 example

CHGASPSSN using the *DETACH parameter removes a FlashCopy relation, but it keeps the ASP session.

CHGASPSSN using the *REATTACH parameter recreates a FlashCopy using the parameters in an already created ASP session.

When you are going to reattach a Metro Mirror session, a message (CPF9898) will go to the QSYSOPR message queue that you will have to use to confirm the reattach.

Error messages will not appear at the bottom of the panel if you run this message through WRKASPCPYD. You will need to check your joblog specifically.

Part 2

Concepts and planning

In this part we provide concepts and planning information for implementing the IBM High Availability solution with PowerHA SystemMirror for i. We also introduce you to the various interfaces that are available in IBM PowerHA for i.

This part includes these chapters:

5

Geographic Mirroring

In this chapter, we introduce geographic mirroring and how it works from a synchronous/asynchronous point of view. We illustrate several scenarios that serve as disaster recovery solutions by using switched disk for local high availability and geographic mirroring between remote sites.

These are the topics that we discuss:

- Concept of geographic mirroring
- Synchronous geographic mirroring
- Asynchronous geographic mirroring
- Switched disk for local HA + Geographic Mirroring for DR scenario

5.1 Concept of geographic mirroring

Geographic mirroring has been available in i5/OS V5R3M0. Geographic mirroring specifically refers to the IBM i host-based management storage mirroring solution. Now geographic mirroring is provided as a function of PowerHA SystemMirror for i products. PowerHA SystemMirror for i also includes replication solutions such as Metro Mirror and Global Mirror by using IBM Storage products.

The basis of geographic mirroring is extended IBM i disk mirroring technology to multiple systems environment. Geographic mirroring is managed by IBM i storage management capability so that replication is performed by each memory page segment (4096 bytes).

Geographic mirroring is intended for use by clustered system environments and uses data port services. Data port services are System Licensed Internal Code that supports the transfer of large volumes of data between a source system and one of any specified target systems. This is a transport mechanism that communicates over TCP/IP.

Prior to IBM i 7.1, it provides only synchronous delivery mode. Be aware that a local write waits for the data to reach the main storage of the backup node before the write operation is considered to be finished.

Asynchronous geographic mirroring (asynchronous delivery mode) is supported by PowerHA SystemMirror for i Enterprise Edition (5770-HAS option 1) with IBM i 7.1 extending the previously available synchronous geographic mirroring option, which for performance reasons is practically limited to metro area distances up to 30 km.

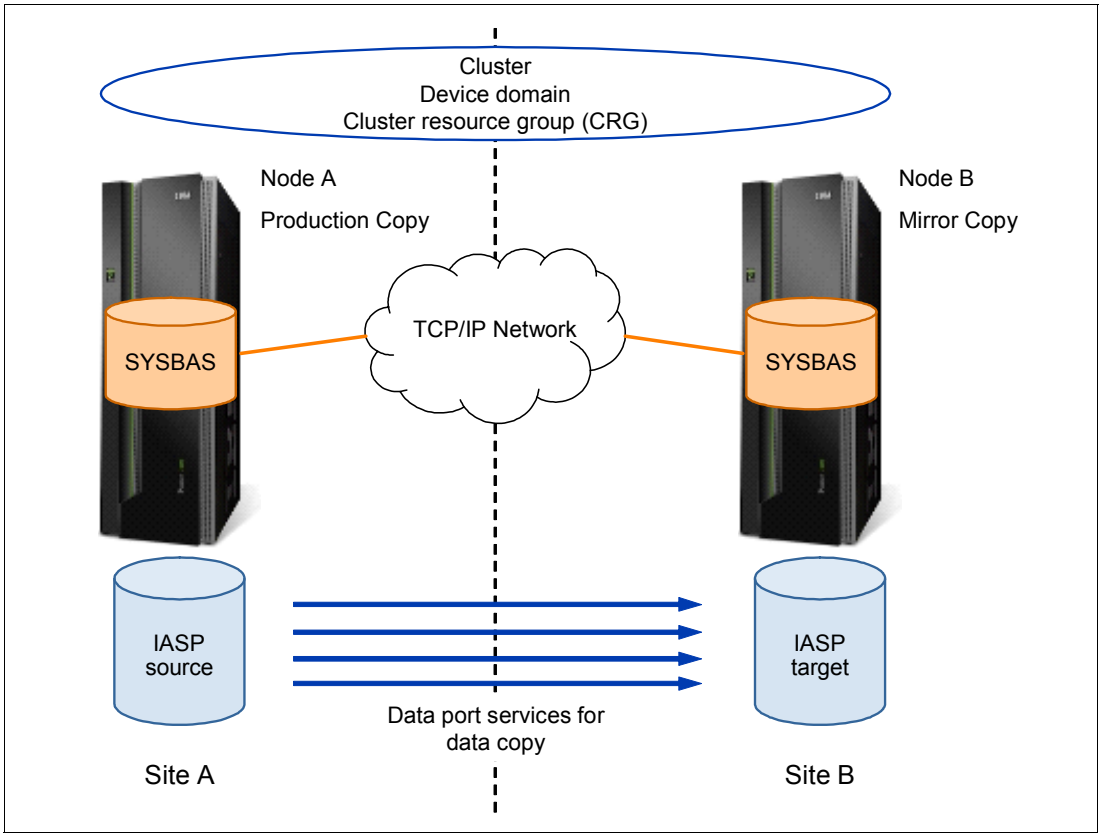


Figure 5-1 Geographic mirroring between two sites

Note: The data replication process between the two sets of disks (production copy and mirror copy), described by the blue arrow in Figure 5-1, is performed based on TCP/IP communication between the two systems.

The copy owned by the primary node is the production copy, and the copy owned by the backup system at the other site is the mirror copy. Users and applications can only access the independent disk pool on the primary node, the node that owns the production copy. Changes that are made to the production copy (source system) are guaranteed by the geographic mirroring functionality to be made in the same order on the mirror copy (target system or partition).

Geographic mirroring allows for the production and mirrored copies to be at the same site for high-availability protection in the event of server failure, and it is also possible to separate the two systems geographically for disaster recovery protection in the event of a site-wide outage, provided that the communication link between the two sites is enough for your synchronization timing policy.

Geographic mirroring functionality involves the use of these cluster components:

- Cluster
- Device domain
- Cluster resource group
- ASP copy description
- ASP session
- Cluster administrative domain
- Independent auxiliary storage pools (IASPs)

Configuration basis for geographic mirroring

The nodes participating in geographic mirroring must be part of the same cluster and of the same device domain, and their role is defined in the recovery domain of the cluster resource group (CRG). Before configuring geographic mirroring, you must specify a site name and the TCP/IP address to be used for the mirroring process (up to four) for each node in the recovery domain within the device CRG of your IASP. When you configure the cluster for geographic mirroring, you have many options for defining the availability and the protection of the independent disk pool.

Geographic mirroring happens on the page level of storage management. Therefore, the size of individual disks and the disk protection used on the production IASP can differ from what is used on the backup IASP. The overall size of the two IASPs should be about the same on both systems though.

Important: If both disk pools (source and target) do not have the same disk capacity available, when the mirrored copy reaches 100%, geographic mirroring is suspended. You can still add data to your production IASP, but you lose your high availability. If, however, the production copy reaches 100%, you are no longer able to add data to that IASP, and applications trying to do so come to a stop. An IASP can never overflow to the system ASP, as that would compromise your high-availability environment.

Data port services

Geographic mirroring provides logical page-level mirroring between independent disk pools through the use of data port services. Data port services manages connections for multiple IP addresses (up to four), which provides redundancy and greater bandwidth in geographic mirroring environments. We strongly suggest that different IP interfaces, connected to different networks, be used for data port services and cluster heartbeating. You can see more

detailed considerations of the communication environment in “Communications lines” on page 147.

Mirroring option

You can specify how to replicate from your production data in IASP to back up IASP. Mirroring options of geographic mirroring can be specified as ASP session attributes, which are a combination of two parameters:

- Transmission delivery
- Mirroring mode

Table 5-1 shows you the combination of these parameters that you can specify.

Table 5-1 Mirroring option

		Mirroring mode	
		*SYNC	*ASYNC
Transmission delivery	*SYNC	Synchronous geographic mirroring (synchronous mirroring mode)	Synchronous geographic mirroring (asynchronous mirroring mode)
	*ASYNC	N/a	Asynchronous geographic mirroring

As shown in Table 5-1, synchronous geographic mirroring has two modes:

- Synchronous mirroring mode
- Synchronous mirroring mode (also called semi-asynchronous mode)

Both mirroring modes use synchronous communication between sites. Synchronous geographic mirroring mode is configured to specify transmission delivery as synchronous in IBM i 7.1.

Figure 5-2 shows the **DSPASPSSN** command screen as configuring synchronous geographic mirroring with synchronous mirroring mode.

```
                          Display ASP Session                    DEMOGEO1
                                                          09/30/11  11:20:48
 Session . . . . . . . . . . . . . . . . . . :   GEOMIRROR
   Type  . . . . . . . . . . . . . . . . . . :     *GEOMIR
   Transmission Delivery . . . . . . . . . . :     *SYNC
   Mirroring Mode  . . . . . . . . . . . . . :     *SYNC
   Suspend timeout . . . . . . . . . . . . . :     120
   Synchronization priority  . . . . . . . . :     *MEDIUM
   Tracking space allocated  . . . . . . . . :     100%
...
```

Figure 5-2 Display ASP session command: Synchronous mirroring mode

Figure 5-3 shows the **DSPASPSSN** command screen as configuring synchronous geographic mirroring with asynchronous mirroring mode.

```
                          Display ASP Session                    DEMOGEO1
                                                        09/30/11  11:15:49
 Session . . . . . . . . . . . . . . . . . . :   GEOMIRROR
   Type  . . . . . . . . . . . . . . . . . . :     *GEOMIR
   Transmission Delivery . . . . . . . . . . :     *SYNC
   Mirroring Mode  . . . . . . . . . . . . . :     *ASYNC
   Suspend timeout . . . . . . . . . . . . . :     120
   Synchronization priority  . . . . . . . . :     *MEDIUM
   Tracking space allocated  . . . . . . . . :     100%
...
```

Figure 5-3 Display ASP session command: Asynchronous mirroring mode

Figure 5-4 shows the **DSPASPSSN** command screen as configuring asynchronous geographic mode.

```
                          Display ASP Session                    DEMOGEO1
                                                        09/30/11  11:17:20
 Session . . . . . . . . . . . . . . . . . . :   GEOMIRROR
   Type  . . . . . . . . . . . . . . . . . . :     *GEOMIR
   Transmission Delivery . . . . . . . . . . :     *ASYNC
   Mirroring Mode  . . . . . . . . . . . . . :     *ASYNC
   Suspend timeout . . . . . . . . . . . . . :     120
   Synchronization priority  . . . . . . . . :     *MEDIUM
   Tracking space allocated  . . . . . . . . :     100%

...
```

Figure 5-4 Display ASP session command: Asynchronous geographic mode

Synchronization

When geographic mirroring is resumed after suspend or detach, the mirror copy will be resynchronized with the production copy. The production copy can function normally during synchronization, but performance might be negatively affected. During synchronization, the contents of the mirror copy are unusable, and it cannot become the production copy. If the independent disk pool is made unavailable during the synchronization process, synchronization resumes where it left off when the independent disk pool is made available again. Message CPI095D is sent to the QSYSOPR message queue every 15 minutes to indicate progression of the synchronization.

These are the two types of synchronization:

- Full synchronization

  Indicates that a complete synchronization takes place. Changes to the production copy were not tracked to apply to the synchronization. A full synchronization first deletes all data in the backup IASP and then copies the current data from the production IASP to the backup IASP.

- Partial synchronization

  Indicates that changes to the production copy were tracked while geographic mirroring was suspended or detached. This might shorten the synchronization time considerably

because a complete synchronization is unnecessary. In this case when the mirror copy is reattached and geographic mirroring is resumed, only tracked changes will need to be synchronized. Changes made on the production copy (since the detach has been done) are sent to the mirror copy, and any change made on the mirror copy will be overlayed with the original production data coming from the production copy of the IASP. Logically, any changes made on the detached mirror copy are undone, and any tracked changes from the production copy are applied.

Message CPI095D indicates what types of synchronization happened. Figure 5-5 shows the message.

```
…
The synchronization is of type 2. The synchronization types and their meanings
are as follows:
1 - The synchronization being performed is a synchronization of tracked
changes.
2 - The synchronization being performed is a synchronization of all data.
…
```

Figure 5-5 Message CPI095D

Two parameters can be used to better manage IASP copies synchronization and application performances when geographic mirroring is used:

- Synchronization priority

  When you set the attributes for geographic mirroring, you can set the synchronization priority. You can select synchronization priority as high, medium, or low. If synchronization priority is set high, the system uses more resources for synchronization, which results in a sooner completion time. The mirror copy is eligible to become a production copy faster, so you are protected sooner. However, high priority can cause degradation to your application. We recommend that you try high priority first, so you are protected as soon as possible. If the degradation to your application performance is not tolerable, then lower the priority. Be aware that you need to vary off the IASP to perform this change.

- Suspend timeout

  In addition to synchronization priority, you can also set the suspend timeout. The suspend timeout specifies how long your application can wait when geographic mirroring cannot be performed. When an error, such as a failure of the communication link, prevents geographic mirroring from occurring, the source system waits and retries for the specified suspend timeout before suspending geographic mirroring, which allows your application to continue.

Figure 5-6 shows you as example of display geographic mirroring attribute via **DSPASPSSN**.

```
                          Display ASP Session                    DEMOGEO1
                                                        09/30/11  11:13:03
Session . . . . . . . . . . . . . . . . . . . . :   GEOMIRROR
  Type  . . . . . . . . . . . . . . . . . . . . :     *GEOMIR
  Transmission Delivery . . . . . . . . . . . . :     *SYNC
  Mirroring Mode  . . . . . . . . . . . . . . . :     *SYNC
  Suspend timeout . . . . . . . . . . . . . . . :     120
  Synchronization priority  . . . . . . . . . . :     *MEDIUM
  Tracking space allocated  . . . . . . . . . . :     100%


                             Copy Descriptions

ASP           ASP                                          Data
Device        Copy               Role         State       State    Node
IASP1         S1CPYD        PRODUCTION     AVAILABLE       USABLE   DEMOGEO1
IASP1         S2CPYD            MIRROR        ACTIVE       USABLE   DEMOGEO2

                                                                     Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel   F19=Automatic refresh
```

Figure 5-6 Display geographic mirroring attribute via DSPASPSSN command

Tracking space

The tracking space was introduced with IBM i 5.4. It enables geographic mirroring to track changed pages while in suspended status. With tracked changes we can avoid full resynchronization after a resume in many cases, therefore minimizing exposed times where you do not have a valid mirror copy. Tracking space gets configured when you configure geographic mirroring or change geographic mirroring attributes via **CHGASPSSN**.

Tracking space is allocated inside of the independent ASPs. The more tracking space that you specify, the more changes the system can track. The amount of space for tracking can be defined by the user up to 1% of the total independent ASP capacity. When we specify tracking space size, we can specify a percentage of total usable tracking space size. If you specify 100%, you have 1% of your total independent ASP capacity as tracking space size. For example, if you have an independent ASP with 100 GB, you can have a maximum of 1 GB of storage space as the tracking space, and if you specify the tracking space parameter as 50%, you have 500 MB of storage space as tracking space. Be aware that this tracking space does not contain any changed data. It just holds information about what pages in the IASP have been changed.

5.2 Synchronous geographic mirroring

In this section we discuss synchronous geographic mirroring.

5.2.1 Synchronous geographic mirroring with synchronous mirroring mode

Synchronous geographic mirroring with synchronous mirroring mode needs to specify that both transmission delivery and mirroring mode are synchronous.

When geographic mirroring is active in synchronous mode (Figure 5-7), the write on disk operation waits until the operation is complete to the disk (actually the data write to the IOA cache) on both the source (acknowledgement operation #4) and target systems (acknowledgement operation #3) before sending the acknowledgment to the storage management function of the operating system of the production copy. See the operations numbered 1 - 4 with green arrows in Figure 5-7.

The mirror copy is always eligible to become the production copy, because the order of writes is preserved on the mirror copy. If you are planning to use geographic mirroring as helocal HA solution, we recommend trying synchronous mode first. If your performance remains acceptable, continue to use synchronous geographic mirroring.

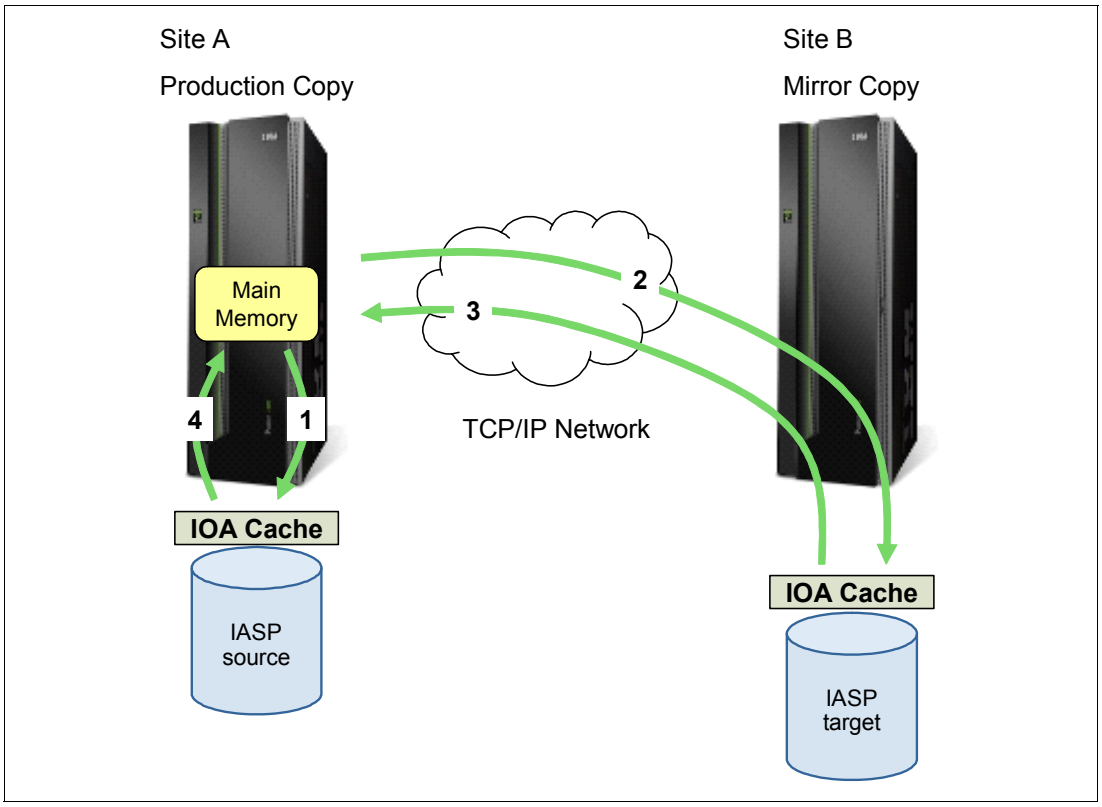


Figure 5-7 Synchronous Mirroring mode

5.2.2 Synchronous geographic mirroring with asynchronous mirroring mode

When geographic mirroring is using asynchronous mirroring mode, the write on disk operation must wait to get an acknowledgement from the production copy for the write operation when it is completed to the disk (actually to the IOA cache - operation #4) on the

source system and is received for processing on the target system (actually in main memory - operation #2 and acknowledgement operation #3) only. See the operations numbered 1 - 4 with green arrows shown in Figure 5-8. The physical write operation, #5 in orange in the figure, is performed later (asynchronously) to the disk on the mirror copy (target system).

In IBM i 7.1, asynchronous mirroring mode can activate to specify that transmission delivery is synchronous and mirroring mode is asynchronous.

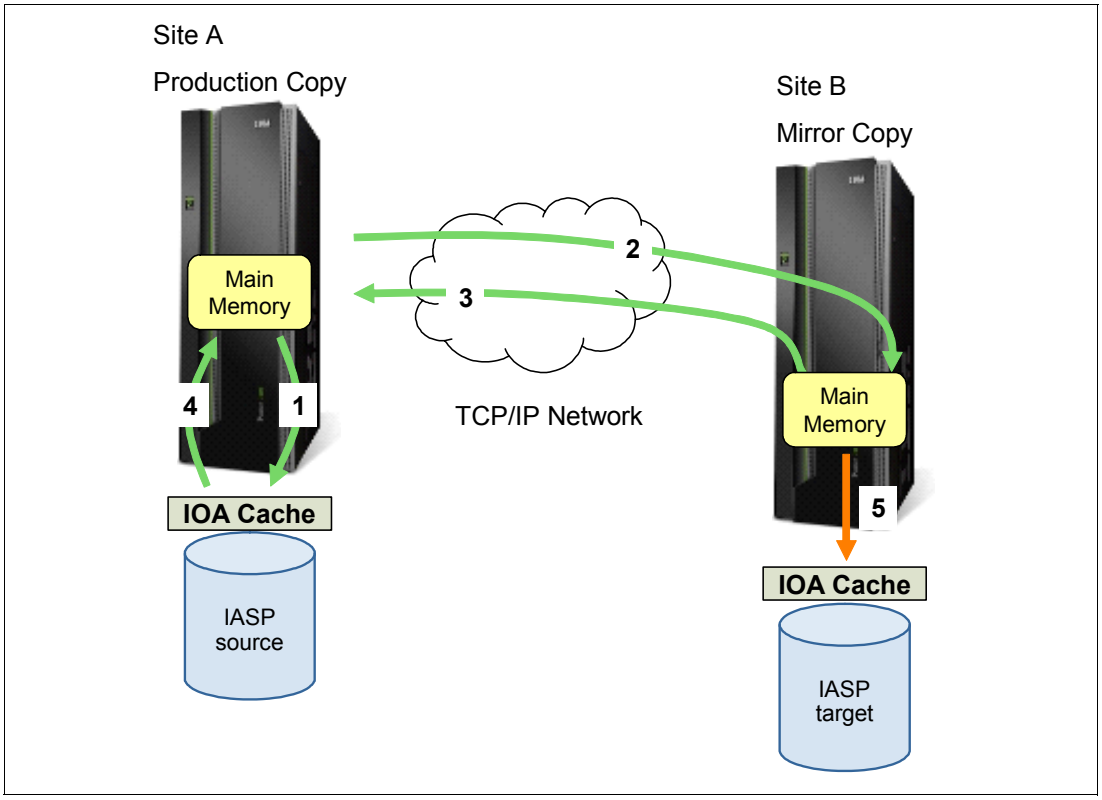


Figure 5-8 Asynchronous mirroring mode

In this mode, the pending updates must be completed before the mirror copy can become the production copy. This means that while you might see a slightly better performance during normal operation, your switchover or failover times might be slightly longer because changes to the backup IASP might still reside in the main memory of your backup system. They must be written to disk before the IASP can be varied on.

Important: Because this mode still waits for the write to cross to the target system, it is not truly asynchronous. We recommend this for situations in which the source and target systems are less than 30 km apart.

5.3 Asynchronous geographic mirroring

Asynchronous geographic mirroring is a new capability supported by PowerHA SystemMirror for i Enterprise Edition (5770-HAS option 1) with IBM i 7.1, extending the previously available synchronous geographic mirroring option, which for performance reasons is practically limited to metro area distances up to 30 km.

The asynchronous delivery of geographic mirroring (not to be confused with the asynchronous mirroring mode of synchronous geographic mirroring) allows IP-based hardware replication beyond synchronous geographic mirroring limits (Figure 5-7 on page 74 and Figure 5-8 on page 75). The write on disk operation does not wait until the operation is delivered to target system (operation #1 and #2).

Asynchronous transmission delivery, which also requires the asynchronous mirroring mode, works by duplicating any changed IASP disk pages in the *BASE memory pool on the source system and sending them asynchronously while preserving the write-order to the target system in operation #3 in Figure 5-9. Therefore, at any given time, the data on the target system (though not up-to-date) still represents a so-called crash-consistent copy of the source system.

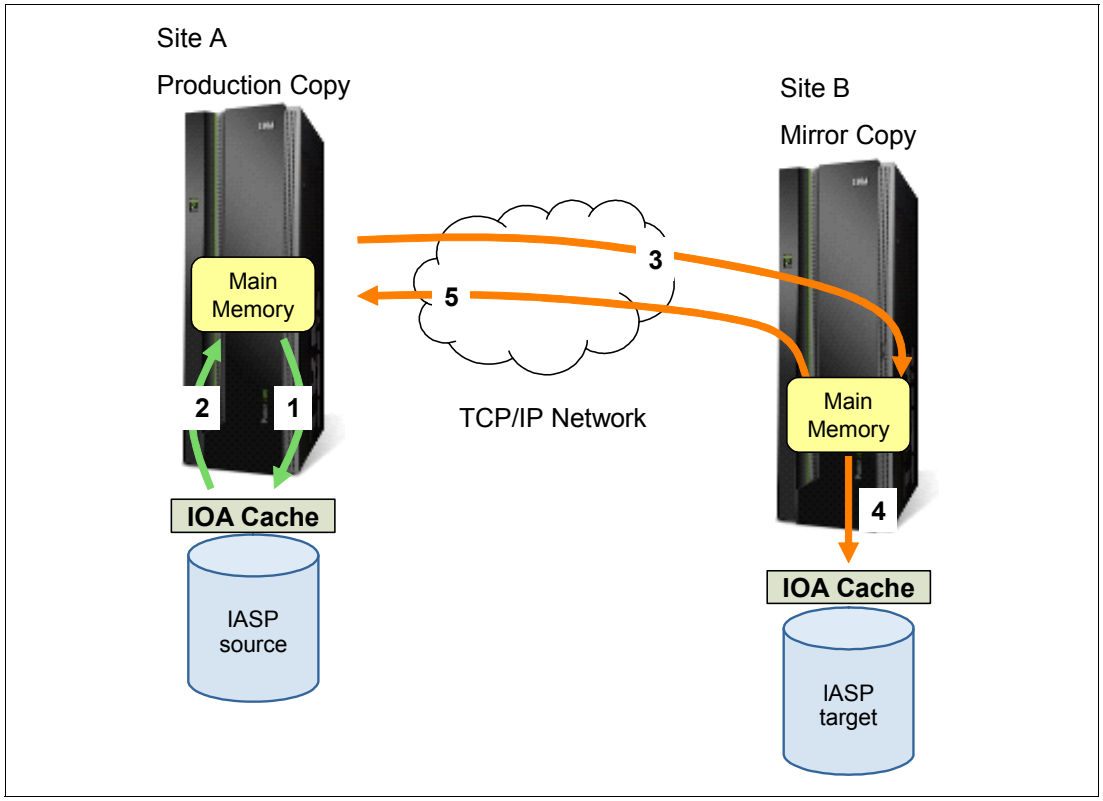


Figure 5-9 Asynchronous geographic mirroring

The asynchronous geographic mirroring option has potential performance impacts to system resources, such as processor and memory because communication lines with longer latency times might tie up additional memory resources for keeping their changed data.

Also consider the distance and amount of latency of your communication lines by using mirror activities. If you have very write-intensive applications, they can flood the communication line and suspend your session with he auto resume process. Then you might tune the timeout value by using **CHGASPSSN** (Figure 5-10). The default value of the SSPTIMO (Suspend timeout) parameter is 120 seconds.

```
                      Change ASP Session (CHGASPSSN)

Type choices, press Enter.

Session  . . . . . . . . . . . . > GEOMIRROR     Name
Option . . . . . . . . . . . . . > *CHGATTR      *CHGATTR, *SUSPEND...
ASP copy:
  Preferred source . . . . . . . > S1CPYD        Name, *SAME
  Preferred target . . . . . . . > S2CPYD        Name, *SAME
              + for more values
Suspend timeout  . . . . . . . . > 120           60-3600, *SAME
Transmission delivery  . . . . . > *SYNC         *SAME, *SYNC, *ASYNC
Mirroring mode . . . . . . . . . > *SYNC         *SAME, *SYNC, *ASYNC
Synchronization priority . . . . > *MEDIUM       *SAME, *LOW, *MEDIUM, *HIGH
Tracking space . . . . . . . . . > 100           0-100, *SAME
FlashCopy type . . . . . . . . .   *SAME         *SAME, *COPY
Persistent relationship  . . . .   *SAME         *SAME, *YES, *NO




                                                                        Bottom
F3=Exit   F4=Prompt   F5=Refresh   F10=Additional parameters   F12=Cancel
F13=How to use this display        F24=More keys
```

Figure 5-10 Example of Change ASP session command (change session attributes)

For more information about performance of geographic mirroring see 8.3.1, “Geographic mirroring” on page 144.

5.4 Switched disk for local HA and geographic mirroring for DR

Switched disk high-availability solutions are based on the concept of switching entire independent auxiliary storage pools (IASPs) from one system to another.

5.4.1 Switched disks between logical partitions

In this environment the hardware is switched between two logical partitions on the same physical system. When configuring this type of environment, you need to group all of the switchable hardware resources between partitions into an I/O pool using the Hardware Management Console (HMC). This allows for greater flexibility in defining what will be switched because you no longer have to move all the hardware in an entire tower when doing the switching. With this environment you can select which pieces of hardware are switched and which are not.

Note: Use of high-speed link (HSL) loop connections for IASP switching is no longer supported. POWER6 servers were the last servers supported for switched IASPs between servers.

5.4.2 Combining geographic mirroring and switched disks

Switchable IASPs offer a local HA solution to address system failures involving a particular LPAR. Adding any kind of remote replication to an existing switched disk environment instantly transforms it into a DR environment and provides coverage for disasters involving an entire site. As you can see in Figure 5-11, redundant LPARs provide protection from LPAR outage that might be due to a select software outage or a select planned outage.

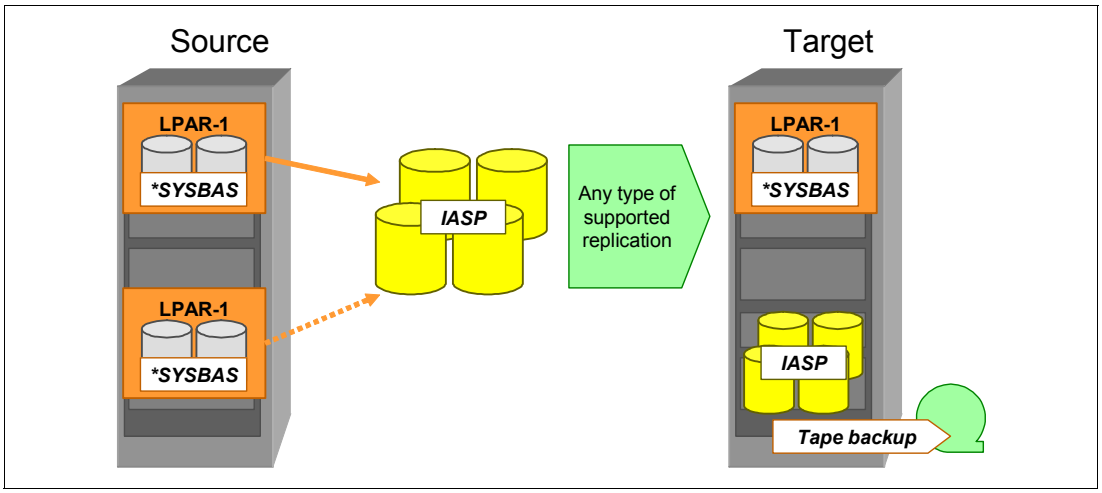


Figure 5-11 Switched IASP combined with replication technologies

Adding replication to a remote site server provides additional levels of protection, including disaster recovery or off-line tape backups. This combination can be achieved using internal disks because PowerHA supports geographic mirroring with internal disks, and the replication can be either synchronous or asynchronous.

Note: Geographic mirroring must first be suspended before IASP can be backed up to tape. PowerHA tracks changes and requires time for partial resynch when backup is complete, and hence affects your recovery point objective (RPO).

As shown in Figure 5-12, use the switched disk environment locally to achieve high availability, whereas the system using geographic mirroring can sit in a remote location and therefore also help to achieve disaster recovery. All three nodes are part of the recovery domain. Node 1 is the primary node, node 2 is the backup node one, and node 3 is backup node two. If just node 1 fails, then node 2 becomes the primary node (by switching and attaching the IASP to that node), and node 3 becomes backup node one. Geographic mirroring would still be active. If the entire site that is hosting both node 1 and node 2 fails, then node 3 becomes the primary node and work can continue there.

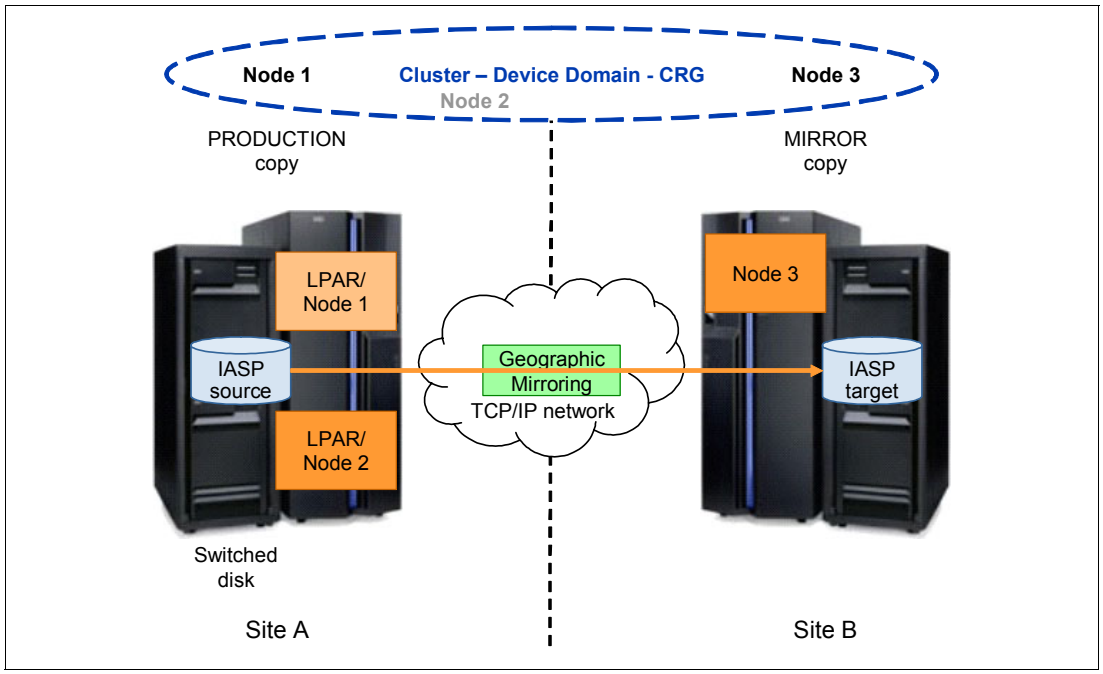


Figure 5-12 Combination of geographic mirroring and switched disk

6

DS8000 Copy Services

In this chapter we describe the DS8000 Copy Services features, which can be used by IBM PowerHA SystemMirror for i. This chapter describes each feature and how it works. However, before talking about Copy Services, a brief description of DS8000 storage concepts is useful.

6.1 DS8000 storage concepts

In this section we describe the major concepts of the DS8000 storage system. The goal of this virtualization is to make sure that the host's operating systems will continue to work exactly like they do with the usual internal drives.

For additional information, especially when performance considerations apply, refer to *IBM System Storage DS8000: Architecture and Implementation*, SG24-8886.

6.1.1 Hardware overview

Before presenting the virtualization concepts, we provide a brief section about the hardware components. At the time of writing, there are two models available as a DS8000. They are the DS8700 and the DS8800. These are the main hardware components:

- Base and expansion frames
  - The DS8700 can have up to four expansion frames. A fully populated five-frame DS8700 contains 1024 3.5" disk drives.
  - The DS8800 can have up to two expansion frames. A fully populated three-frame DS880 contains 1056 2.5" disks.
- Processors

  Both DS8700 and DS8800 use POWER6 processors, the same as those that are installed in the IBM Power Systems 570 servers (based on POWER6).

  For redundancy purposes, there are two processors complexes installed in the base frame. They normally are named server 0 and server 1. The total installable memory is 384 GB, each server using half of the installed memory. This memory is used to provide IO cache capabilities:

  - The DS8700 can have 2-way or 4-way POWER6 processors, running at 4.7 GHz.
  - The DS8800 can have 2-way or 4-way POWER6+™ processors, running at 5.0 GHz.

  Both DS8700 and DS8800 use the Peripheral Component Interconnect Express (PCIe) infrastructure to access disk subsystems and host connections.
- I/O enclosures

  I/O enclosures are installed in both the base frame and the first expansion frame with a maximum of eight enclosures. Each enclosure can have a maximum of four Fibre Channel 4-port adapters in the DS8700 and two 8-port adapters in the DS8800. The maximum number of ports is 128 4-Gbit/s ports. 8-Gbit/s ports are supported, but installation restrictions apply.
- Device adapters

  Installed in the I/O enclosures, device adapters connect to the installed disks drives through Fibre Channel interfaces card (running at 2 Gbps on DS8700 and 8 Gbps on DS8800) installed in the disks enclosures. Working by pair, one attached to each server, they are responsible for disks drives RAID protections.
- Disks enclosures

  Disks enclosures contain disks drives modules, Fibre Channel Interface Cards to connect to the device adapters and to the disks.

Table 6-1 lists available disk formats.

Table 6-1 Available disk formats

Disks drive formats	Availability and connection
300 GB 15 K rpm	Fibre Channel interface on DS8700
450 GB 15 K rpm	Fibre Channel interface on DS8700
600 GB 15 K rpm	Fibre Channel interface on DS8700
146 GB 15 K rpm	SAS interface on DS8800
450 GB 10 K rpm	SAS interface on DS8800
600 GB 10 K rpm	SAS interface on DS8800
2 TB 7,2 K rpm	SATA interface on DS8700
300 GB SSD	SAS interface on DS8800
600 GB SSD	Fibre Channel interface on DS8700

- Power and batteries

  Power supplies, batteries, and cooling are highly redundant.

  There are two redundant power supplies in each frame. Each server (processor complex) has two power supply units. Each disk enclosure has two power supply units. Each I/O enclosure has two power supply units.

  The battery backup (BBU) assemblies help protect data in the event of a loss of external power. In the event of a complete loss of AC input power, the battery assemblies are used to maintain power to the processor complexes and I/O enclosures for a sufficient period of time, to allow the contents of NonVolatileStorage memory (modified data not yet destaged to disk from cache) to be written to a number of disk drives internal to the processor complexes.

- Hardware Management Console (HMC)

  All base frames ship with an integrated HMC, which is used by these people:

  – IBM representatives to perform all maintenance operations, like replacing a failing item

  – The user to perform configuration and management through command-line interfaces and graphical user interfaces

6.1.2 Array site

An array site is a group of eight disk drives. Those disk drives are already assigned to array sites at the installation by the DS8000, and you do not have to deal with them. They are selected from two disk enclosures on two different loops (Figure 6-1). The loops are the internal links between the DS8000 RAID device adapters connected to the DS8000 processors and the disks drive modules.

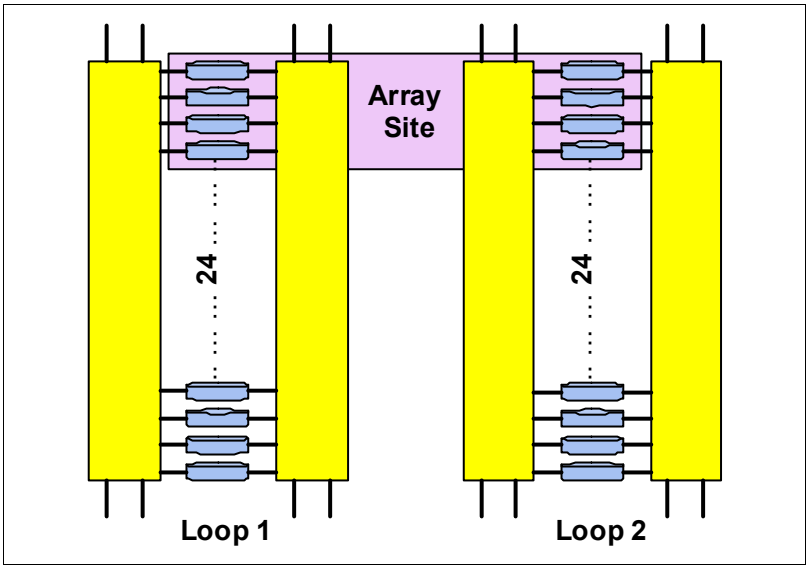


Figure 6-1 An array site

All disk drives in an array site have the same capacity and the same speed.

6.1.3 Array

DS8000 logical storage configuration starts at this point. An array is a group of eight disk drives created from an *array site*, defining the disk protection that you want to use. Available disks protections are RAID10, RAID5, and RAID6. When creating the array, specify the array site to use and the protection to apply. At this time, the DS8000 selects a spare disk to put in the array, if needed according to sparing algorithm. None, one, or two spare disks per array can exist.

Note: Spare disks, while they remain spare, are not a member of any RAID parity set even if they are a member of the array. They become a member of the RAID parity set if they replace a failing drive. These operations are automatically handled by the DS8000 when needed.

Several types of arrays exist, depending on the existence of spares and the selected disks protection:

4+4 RAID10 protection with no spare, four disks mirrored with four other disks.

3+3+2S RAID10 protection with two spares, three disks mirrored with three other disks.

6+P+Q RAID6 protection with no spare, two parity disks.

5+P+Q+S RAID6 protection with one spare, two parity disks.

7+P RAID5 protection with no spare, one parity disk.

6+P+S RAID5 protection with one spare, one parity disk (Figure 6-2). In Figure 6-2, the DS8000 reserves one entire disk for spare and distributes parity space and data space on each of the other disks (on the right side, Dx stands for data space and P for parity space).

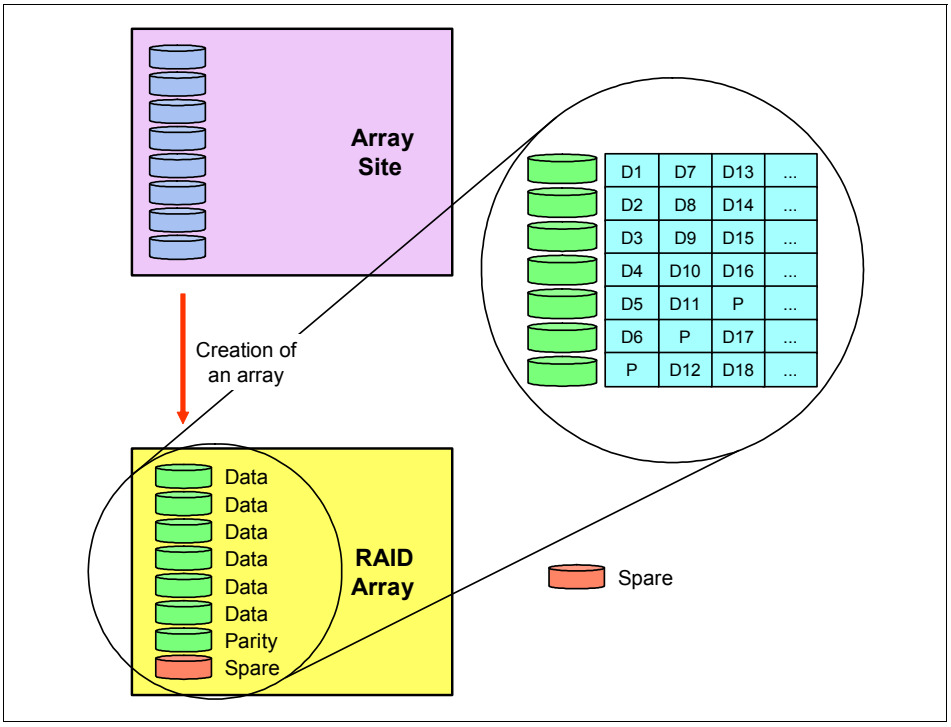


Figure 6-2 A RAID5 array with spare

A *one-to-one* relationship exists between an array and an array site.

6.1.4 Rank

The next configuration step is to dedicate the array to one of the two disk formats provided by the DS8000. It supports, at the same time, connections from both System z® servers and Windows/Linux/Unix and IBM i operating systems, but disk formats are not the same for both types. System z servers make use of count key data (CKD) format, while all the others make use of fixed block (FB) format.

The *rank* defines the way that the entire array is formatted for CKD usage or FB usage. At rank creation time, the array is also divided into extents. One extent is the smallest disk space that can be allocated to DS8000 host operating systems.

For CKD usage, one extent is equivalent to 1113 cylinders, which was the capacity of an original 3390 model 1 disk. System z administrators are used to deal with cylinders as disks size units.

For FB usage, one extent is equivalent to 1 GB (more precisely GiB, being equal to 2^{30} bytes).

Figure 6-3 shows an example of a fixed block rank with 1 GB extents. In Figure 6-3, one square represents one extent. The DS8000 distributes every extent to all disks members of the array.

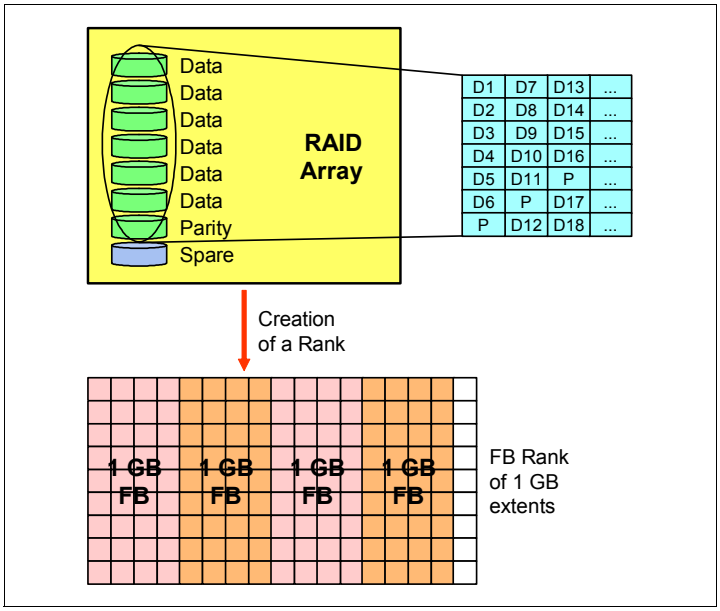


Figure 6-3 A fixed block rank

Note: From a DS8000 standpoint, the IBM i operating system and Windows/Unix/Linux operating systems are members of the same community, the fixed block community. This means that, for example, on the same rank there can coexist extents used by an IBM i operating system and others used by a Windows operating system. You have to make sure that you understand possible performance impacts with such a configuration.

A one-to-one relationship exists between a rank and an array.

6.1.5 Extent pools

An extent pool aggregates extents of the same usage (CKD or FB) from one or more ranks into a single object. The extent pool provides also the affinity to one of the two DS8000 servers through its rank group parameter, which can be set to 0 for server 0 or to 1 for server 1. Affinity means preferred server usage during normal operations. For performance reasons, it is important to properly balance servers usage. At the minimum, two extent pools must exist on a DS8000, one for server 0 and the other for server 1.

Figure 6-4 shows an example of two CKD extent pools and two FB extents pools:

- The CKD0 extent pool has an affinity to server 0 and makes use of two ranks.
- The CKD1 extent pool has an affinity to server 1 and makes use of a single rank.
- The FBtest extent pool has an affinity to server 0 and makes use of two ranks.
- The FBprod extent pool has an affinity to server 1 and makes use of three ranks.

Note: Server affinity applies only during normal operations. If a DS8000 server fails, the other is able to manage extent pools previously managed by the failing server.

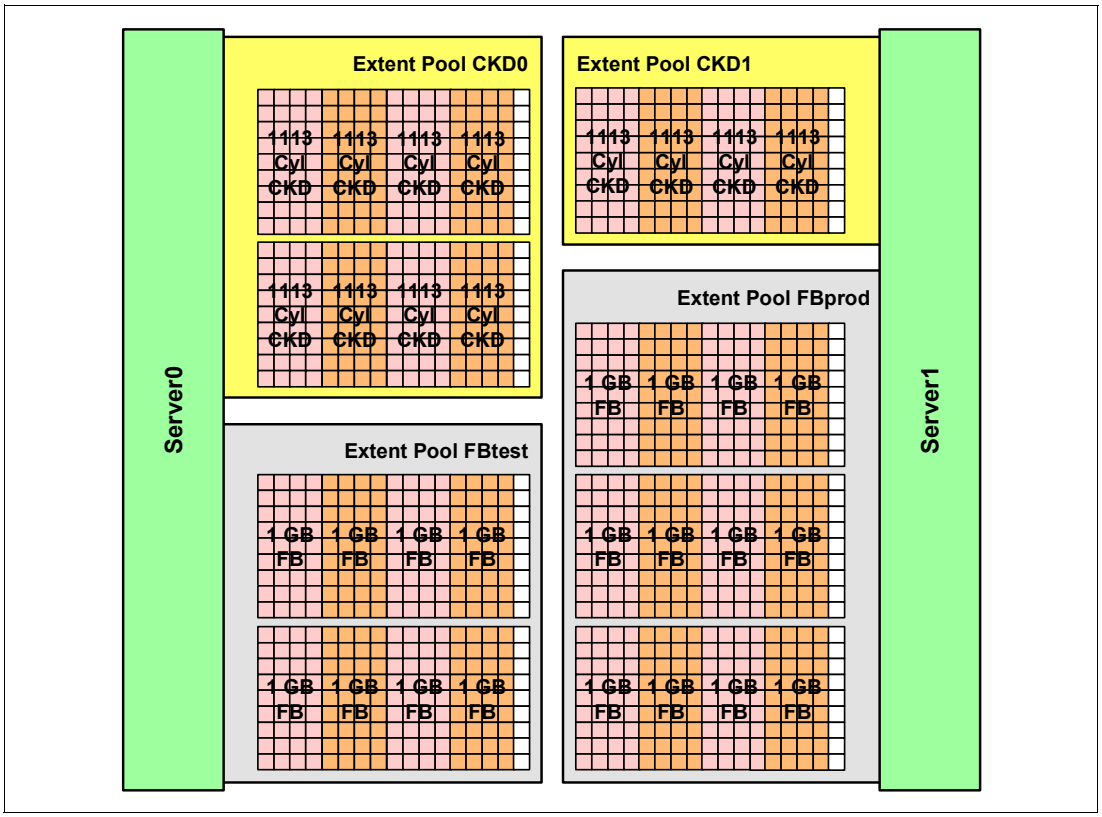


Figure 6-4 Extent pools

Note: Performance considerations apply whether you decide to assign several ranks or a single rank to an extent pool.

A one-to-n relationship exists between an extent pool and a rank.

6.1.6 Volumes

Until this step, the DS8000 is not yet ready for providing any storage resources to host operating systems. The next step is to create *volumes*, or logical volumes, by using a specific extent pool. This one provides the volume, at creation time, with the desired number of extents, located on the associated ranks. The volume does not take care of the extents' physical location. Volume creation commands are not the same for CKD extent pools and for FB extent pools because of parameter differences.

Creating a CKD volume requires the number of cylinders. Creating a FB volume (also known as a LUN for logical unit) requires the number of GB. A volume cannot span multiple extent pools. After it is created, the volume gets a type/model related to the host operating system.

Figure 6-5 shows the process of creating a 2.9 GB FB volume in an extent pool spanned on two ranks. Three extents (so 3 GB) remain available in the extent pool. Other extents are in use by other LUNs. Creating the 2.9 GB volume succeeds but allocates the three remaining extents, leaving 100 MB unused.

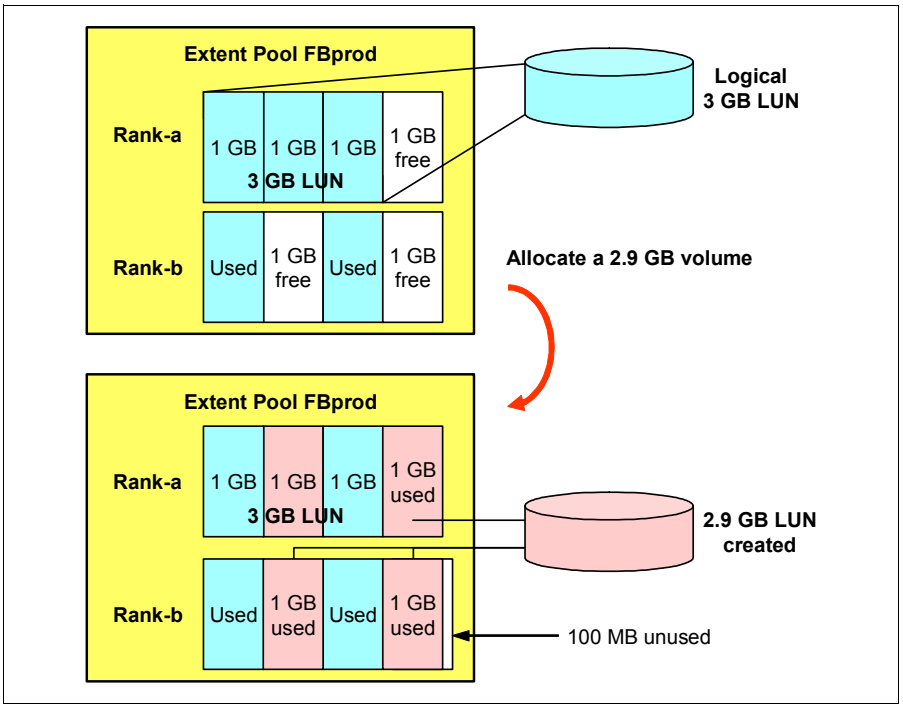


Figure 6-5 Creation of a FB volume

IBM i volumes are fixed block volumes, so they are also composed of a number of 1 GB extents. However, for DS8000 native or NPIV attachment, the IBM i operating system supports only certain fixed volumes sizes that are the same as supported physical disks drives. For example, supported volume sizes are 8.58 GB, 17.54 GB, 35.16 GB, 70.54GB, and so on. The other specific item for IBM i volumes is the disk protection that the operating system wants to achieve. If you want to use host-based IBM i mirroring, IBM i must see unprotected volumes. Otherwise it will be not possible to define and configure the mirrioring. The protection flag provided by the DS8000 for an IBM i volume is only a way to make IBM i believe that the volume is not protected. In reality, all volumes are RAID protected by the DS8000 storage system, as discussed in 6.1.3, “Array” on page 84.

The DS8000 provides the IBM i operating system with a specific serial number for every volume, just like physical disks drives have.

The DS8000 model determines the first two digits of the serial number (for example, DS8100, DS8300, DS8700 report 50). They are the same for all volumes. The next four digits are the volume ID. The last three digits are known as the OS/400 serial number suffix, which is set up at the DS8000 storage image level. This is the same for all volumes, and it points you to the appropriate DS8000, if your installation does not have a unique one.

Figure 6-6 shows an example of a disk serial number. In 50-EE902B8, *50* is related to the DS8000 model, *EE90* is the volume ID inside the DS8000 configuration, and *2B8* is the os400serial at the DS8000 storage image level.

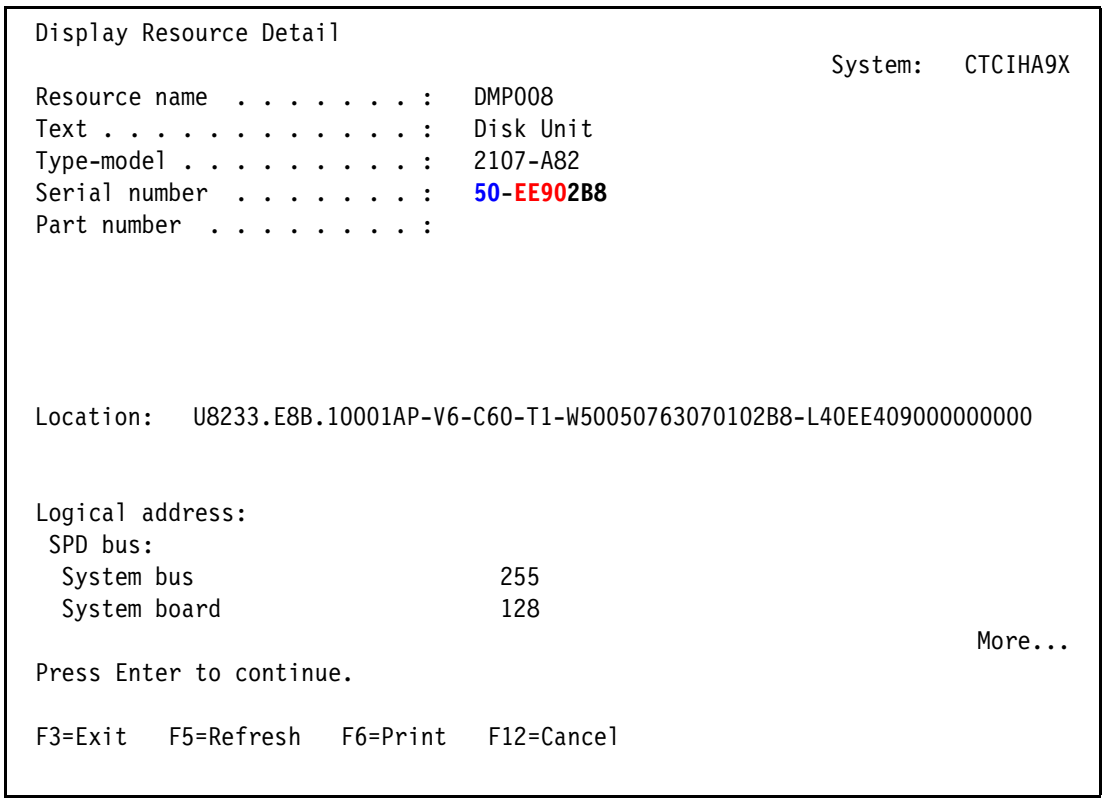


Figure 6-6 Disk serial number seen by IBM i operating system

The following figures show the related information on the DS8000 side. Figure 6-7 shows the storage image os400serial in bold. By default, these three characters are the last three one of the storage image world wide node name, but they can be changed.

Caution: Any change to os400serial requires a DS8000 reboot to be applied.

```
dscli> showsi IBM.2107-75AY031
Date/Time: September 15, 2011 1:48:29 PM CDT
Name             -
desc             -
ID               IBM.2107-75AY031
Storage Unit     IBM.2107-75AY030
Model            9B2
WWNN             5005076307FFC2B8
Signature        232c-a1ec-939a-bb44
State            Online
ESSNet           Enabled
Volume Group     V0
os400Serial      2B8
NVS Memory       2.0 GB
Cache Memory     53.9 GB
Processor Memory 62.7 GB
MTS              IBM.2424-75AY030
numegsupported   0
ETAutoMode       -
ETMonitor        -
IOPMmode         -
```

Figure 6-7 DS8000 storage image settings

Figure 6-8 shows the volume settings. The volume ID is marked in bold. We can also see that, from the DS8000 stand point, deviceMTM is the same parameter as type-model from the IBM i operating system stand point (2107-A82 in our case).

```
dscli> showfbvol EE90
Date/Time: September 15, 2011 1:48:43 PM CDT
Name            HA9x_sys
ID              EE90
accstate        Online
datastate       Normal
configstate     Normal
deviceMTM       2107-A82
datatype        FB 520U
addrgrp         E
extpool         P2
exts            17
captype         iSeries
cap (2^30B)     16.3
cap (10^9B)     17.5
cap (blocks)    34275328
volgrp          V41
ranks           1
dbexts          -
sam             Standard
repcapalloc     -
eam             rotateexts
reqcap (blocks) 34275328
realextents     17
virtualextents  0
migrating       0
perfgrp         -
migratingfrom   -
resgrp          -
```

Figure 6-8 Fixed-block volume settings

An n-to-one relationship exists between volumes and an extents pool.

6.1.7 Volume groups

As its name states, a *volume group* is a set of volumes. The volumes are presented together by the DS8000 to the host operating system through the volumes group. This one has a type that relates to the operating system that makes use of it. The types of volumes included into a volume must match the volume group type. For example, you cannot include volumes that do not have an IBM i type model (such as 2107-A82) into an IBM i volume group.

An n-to-n relationship exists between volumes and volumes groups. However, if a volume belongs to more than one volume's group, it means that it is potentially accessed by several hosts. In this case, the host operating system is responsible for ensuring data integrity. he IBM i operating system does not support sharing a volume with another operating system.

6.1.8 Host connections

A host connection is an object that represents the Fibre Channel adapter port of a host on DS8000. The main attributes of a host connection are its type, related to the volume group type, and its world-wide port name (WWPN), which is a unique 16-hexadecimal-character address. After creating a host connection, we are almost ready for this host to get disk resources to work with. We only have to assign the host connection a volume group. At this time, the host will be able to work with the volumes.

Figure 6-9 shows an example of host connection relationships to volume groups. IBM i partition number 1 has two Fibre Channel adapters, with WWPN 1 and WWPN 2. IBM i partition number 2 has two Fibre Channel adapters, with WWPN 3 and WWPN 4. Inside the DS8000, WWPN 1 and WWPN 2 are assigned the same volume group, volume group 1, and WWPN 3 and WWPN 4 are assigned the same volume group, volume group 2. With this configuration, IBM i partition 1 makes use of volumes included in volume group 1 and IBM i partition 2 makes use of volumes included in volume group 2. Both the partitions have a dual path to the volumes due to dual Fibre Channel attachments and assignment to the same volume group.

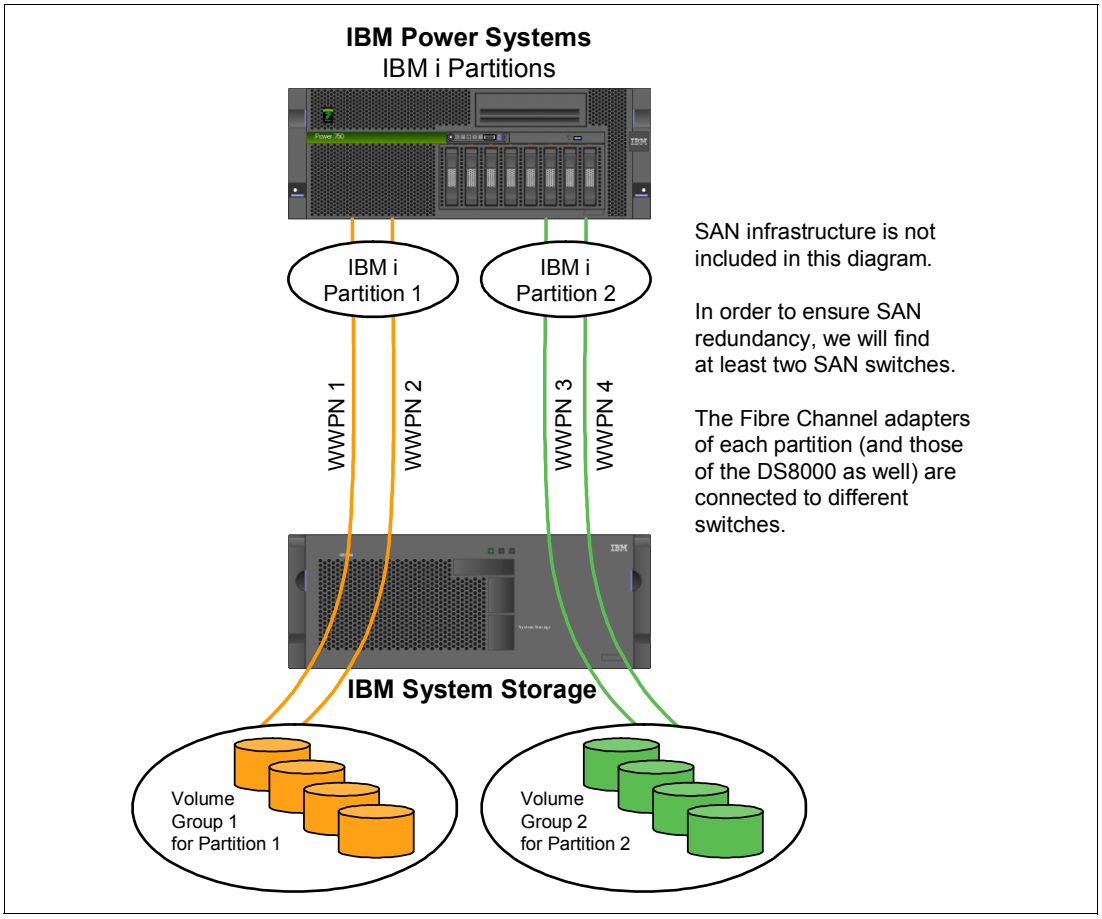


Figure 6-9 Example of host connection relationship to volume groups

A 1-to-n relationship exists between a volume group and host connections.

6.1.9 Logical subsystems

Another logical group of volumes is the *logical subsystem*. It is automatically created the first time a volume with a new logical subsystem id represented by the first two characters of the volume id is created. Therefore all volumes whose ID starts with the same two characters belong to the same logical subsystem (LSS). LSS have an affinity with one of the servers just like the extents pools. Even-numbered LSS belong to server 0 and odd-numbered LSS belong to server 1. Therefore, when we create a volume, affinity related to its extents pool and affinity related to its id through the LSS, must match.

LSS play a lest important role for FB volumes than they do for CKD volumes. For more information on LSS in a CKD context, refer to *IBM System Storage DS8000: Architecture and Implementation*, SG24-8886. In FB context, LSS are important for some of Copy Services usage which are detailed in next sections.

6.2 Metro Mirror

In this section, we introduce the DS8000 Metro Mirror feature when used in IBM i environments. For a complete information, refer to Part 4, “Metro Mirror” in *IBM System Storage DS8000: Copy Services in Open Environments*, SG24-6788.

6.2.1 Metro Mirror overview

Metro Mirror provides real-time copy of volumes between two DS8000. It is a synchronous copy where write operations are completed on both DS8000 before write acknowledgements are sent back to the source operating system. It ensures that no data is lost in the event of a failure. Due to the synchronous copy, distance considerations between the local and the remote DS8000 apply. No more than 300 km are allowed but, even below, write performance becomes more and more affected when this distance increases.

Figure 6-10 illustrates this write sequence when a host requests a write operation:

1. Data is stored in the source DS8000 cache and Non Volatile Storage, to be later destaged to the source volume.
2. Same data is sent to target DS8000, where it is stored in the target DS8000 cache and Non Volatile Storage, to be later destaged to the target volume.
3. Write acknowledgement is sent by target DS8000 to source DS8000.
4. Write acknowledgement is sent by source DS8000 to the host.

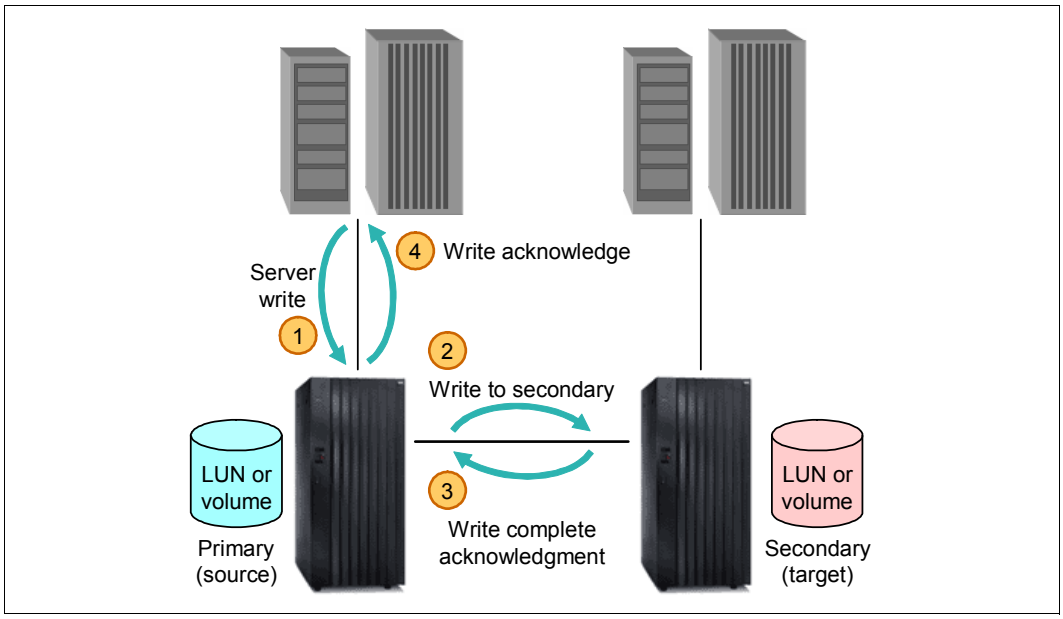


Figure 6-10 Metro Mirror write sequence

Note: For all IBM i volumes involved in a Metro Mirror relationship, both source and target volumes must have the same type-model. They must have the same capacity and the same protection flag.

6.2.2 Metro Mirror operations

This is a short review of basic operations that we can perform with Metro Mirror.

Establishing a Metro Mirror pair

This operation establishes the copy relationship between the source (or local) volume and the target (or remote) volume. Normally, those two volumes reside on two DS8000s, but it is possible to use the same DS8000 for tests purposes. During the initial copy, the target volume is in a *target copy pending* state until all tracks are copied. At the end of the initial copy, the target volumes is in a *target full duplex* state.

Some options are possible at establishment time:

No copy	No data is copied from the source volume to the target volume. You have to make sure that data is really synchronized.
Target read	With this option, the target volumes can be read by a host while replication is active. This option is not supported by IBM i, which must be able to write to any disk at any time.
Suspend after synchronization	As soon as the target volume goes into the *target full duplex* state, the replication is suspended.
Reset reserve on target	The Metro Mirror replication starts even if the target volume is reserved by a host.

Note: A PPRC path must exist from the source to the target site before creating the Metro Mirror pair.

Suspending a Metro Mirror pair

This operation stops sending data from the source volume to the target volume. However, the source DS8000 keeps a record of updated tracks on the source volume.

Resuming a Metro Mirror pair

This operation releases sending data from the source volume to the target volume. It makes use of updated track records during suspend time, to send only updated data.

Terminating a Metro Mirror pair

This operation ends the replication between the source volume and the target volume and removes this relationship. If you ever need the replication again, you have to restart it from the beginning.

Note: After terminating the Metro Mirror pair, you can delete the related PPRC path, depending of path usage for other pairs.

Failover and failback

Failover is related to the actions that you take to activate target volumes in case of a switch from a production site to a backup site. From the DS8000 standpoint, actions are the same for a planned switch and for an unplanned switch. Failover operations start after a failure or maintenance on the initial production site. After a failover, production can run on initial target volumes.

Failback is related to the actions that you take to come back to the nominal situation, with production running on initial source volumes. Failback operations can start after the initial production site failure is fixed or maintenance is finished.

On the backup site, failover function performs three steps in a single task at the volume level:

1. Terminate the original Metro Mirror relationship.
2. Establish a new relationship in the opposite direction, from the initial target to the initial source (from the backup site to the production site).
3. Before any update to the new source volume, suspend the new Metro Mirror relationship.

After these steps, the state of the original source volume is preserved (depending on the failure, nobody might be able to connect to the original source DS8000), and the state of

original target volumes becomes source suspended. All updates on new source volumes are tracked.

When we are ready to switch back to the nominal situation, we can start failback. Still on the backup site, the failback function performs various actions depending on the original source volume state. Failback also works at the volume level:

- If the original source volume is no longer in any Metro Mirror relationship, a new Metro Mirror relationship between the original target volume and the original source volume is established, and all data is copied back from the backup volume to the production volume.
- If the original source volume is still a member of the appropriate Metro Mirror relationship and has no updated tracks, only updated data on the original target volume is copied back to the original source volume.
- If the original source volume is still a member of the appropriate Metro Mirror relationship and has some updated tracks, updated data on the original target volume and the same tracks as the ones that were updated on the original source volume are copied back to the original source volume.

After the failback, the Metro Mirror direction is from the original target volume to the original source volume, and both volumes are synchronized.

The last operation to run to recover the original environment is an another failover and failback to perform, this time on the original production site. Figure 6-11 summarizes all these tasks.

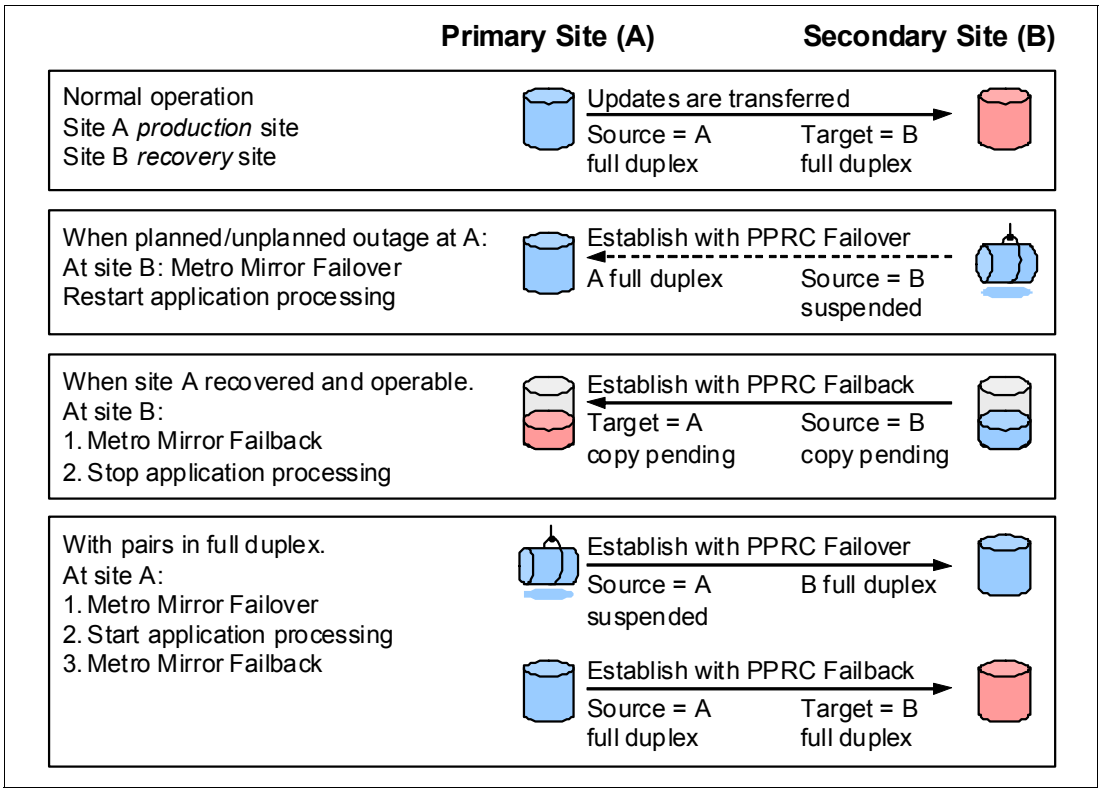


Figure 6-11 Metro Mirror failover and failback sequence

Note: For a planned switchover from site A to site B, and to keep data consistency at site B, the application at site A has to be quiesced before the Metro Mirror failover operation at site B.

6.2.3 PPRC paths and links

The *peer-to-peer remote copy (PPRC) path* is a logical construct on the DS8000 that defines which physical links are used by a Metro Mirror or Global Copy relationship from a DS8000 to another. This path is built at the logical subsystem (LSS) level. It means that Metro Mirror relationships for all volumes in the same LSS will use the same physical links. At the DS8000 standpoint, the physical link is from one of its Fibre Channel adapters to one of the target DS8000's Fibre Channel adapters and can include supported channel extenders or routers.

The path creates an *unidirectional* association, on the source DS8000, between:

- The target DS8000 world wide node name
- The source LSS
- The target LSS
- The source IO port
- The target IO port

For bandwidth and redundancy purposes, a path can have up to eight links. The DS8000 handles failures and balances the workload across available paths. As shown on Figure 6-12, a link can handle several paths.

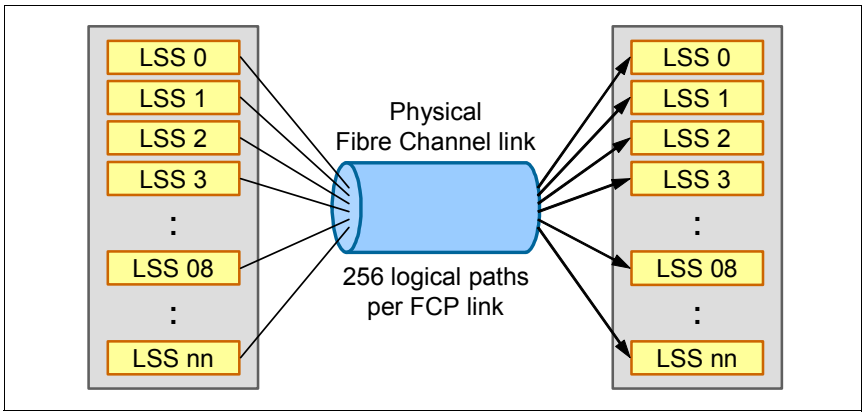


Figure 6-12 PPRC paths and links

Tip: Do not forget to create a path in both directions, from production to backup DS8000 and from backup to production DS8000. Otherwise, failback operation is not possible. Paths from backup to production can use any of the available links. There is no constraint to use the same paths from production to backup.

Note: Although a DS8000 IO port can handle Metro Mirror traffic and host connections at the same time, for performance reasons, we recommend not mixing them. We prefer dedicating I/O ports to replication traffic or host connections.

6.2.4 Metro Mirror and IBM PowerHA SystemMirror for i

Within the IBM PowerHA SystemMirror for i product, Metro Mirror is used to replicate data from an IASP to another IASP in a remote site.

Classical installation consists of a local IBM i partition and a remote IBM i partition. Both are in a cluster, and each partition makes use of a DS8000 external storage for IASP disks. There is no mandatory rule about SYSBAS disks. They can be internal drives or external storage.

Note: Metro Mirror relationships apply *only* to IASP volumes when used with IBM PowerHA SystemMirror for i.

Figure 6-13 shows a setup overview of a Metro Mirror installation for an IASP. Both IBM i partitions are members of a cluster, and Metro Mirror replication is in place for IASP volumes.

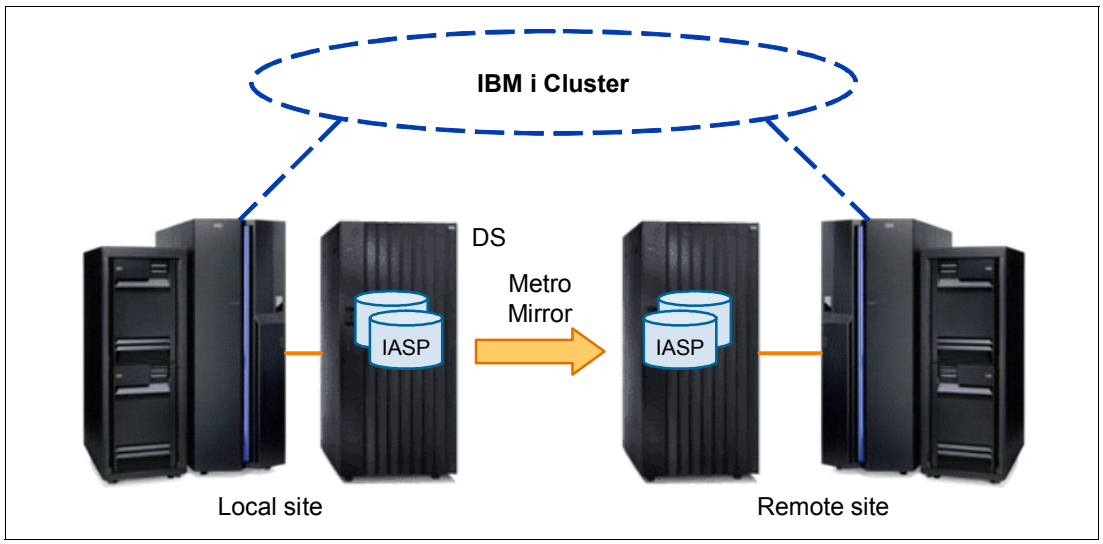


Figure 6-13 Metro Mirror implementation

When a switch occurs, either for a planned outage or an unplanned one, the target IASP volumes become available due to Metro Mirror failover operations, and the remote partition can use them for production activity.

When coming back to a normal situation is possible, failback operations can start, and the original source partition can start production activity.

As you see in 13.1.1, “Configuring IBM i DS8000 Metro Mirror (GUI and CL commands)” on page 264, IBM PowerHA SystemMirror for i handles all setup, switch, and switch back tasks but one. It is not able to establish a Metro Mirror relationship between source and target volumes. This DS8000 setup is done either through DSCLI or the Storage Manager GUI.

6.3 Global Mirror

In this section, we introduce the DS8000 Global Mirror feature when used in IBM i environments. For a complete information, refer to Part 6, “Global Mirror,” in *IBM System Storage DS8000: Copy Services in Open Environments*, SG24-6788.

6.3.1 Global Mirror overview

Global Mirror takes place when the distance from the source site to the target site is higher than that supported for Metro Mirror replication or when synchronous replication is not an option considering application performance impacts. However, DS8000 asynchronous replication is not able to handle data consistency on the target site. To ensure consistency, Global Mirror uses Flash Copy. For more information about the FlashCopy technique, see 6.5, “FlashCopy” on page 107.

Basically, Global Mirror combines two DS8000 techniques, Global Copy and Flash Copy:

- Global Copy itself is not covered in this book, but it is almost exactly the same technique as Metro Mirror, with two main differences:
  - The replication is *asynchronous*. There is some delay before writes are effective on the target site. RPO is not zero in this case. It mainly depends on network bandwidth and latency between source and target sites. Figure 6-14 illustrates how Global Copy operates when a host requires a write operation:
    i. Data is stored in the source DS8000 cache and Non Volatile Storage, to be later destaged to the source volume.
    ii. Write acknowledgement is sent by the source DS8000 to the host.
    iii. At a later time (that is, in an asynchronous manner), the source DS8000 sends the necessary data so that the updates are reflected on the target volumes. The updates are grouped in batches for efficient transmission.
    iv. The target DS8000 returns write complete to the source DS8000 when the updates are written to the target DS8000 cache and NVS.

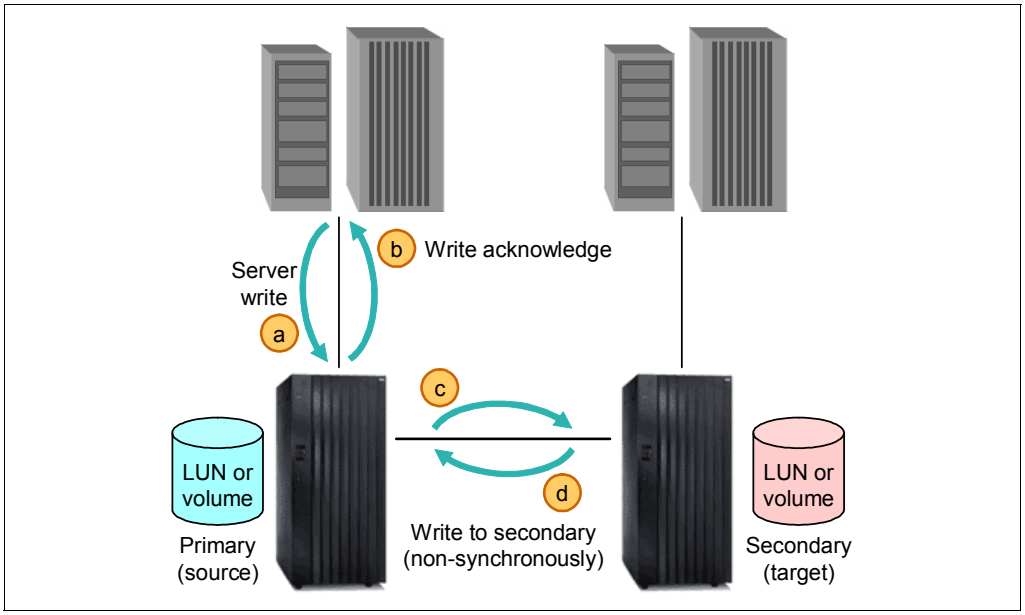


Figure 6-14 Global Copy write sequence

  - There is no guarantee that the write sequence on the target site is the same as it is on the source site. Therefore, *data is not consistent* on the target site. It becomes consistent when the application or host using the source volumes is quiesced. So data migration is the main usage of Global Copy.

- FlashCopy takes place to help maintain consistency volumes, as described in Figure 6-15. These are specific FlashCopy attributes required for Global Mirror:
  - Inhibit target write: Protect FlashCopy target volume C from being updated by anyone other than Global Mirror.
  - Enable change recording: Apply changes only from the source volume to the target volume that occurred to the source volume in between FlashCopy establish operations, except for the first time when FlashCopy is initially established.
  - Make relationship persistent: Keep the FlashCopy relationship until explicitly or implicitly terminated.
  - Nocopy: Do not initiate background copy from source to target, but keep the set of FlashCopy bitmaps required for tracking the source and target volumes. These bitmaps are established the first time that a FlashCopy relationship is created with the nocopy attribute. Before a track in the source volume B is modified, between Consistency Group creations, the track is copied to target volume C to preserve the previous point-in-time copy. This includes updates to the corresponding bitmaps to reflect the new location of the track that belongs to the point-in-time copy. Note that each Global Copy write to its target volume within the window of two adjacent consistency groups can cause FlashCopy I/O operations.

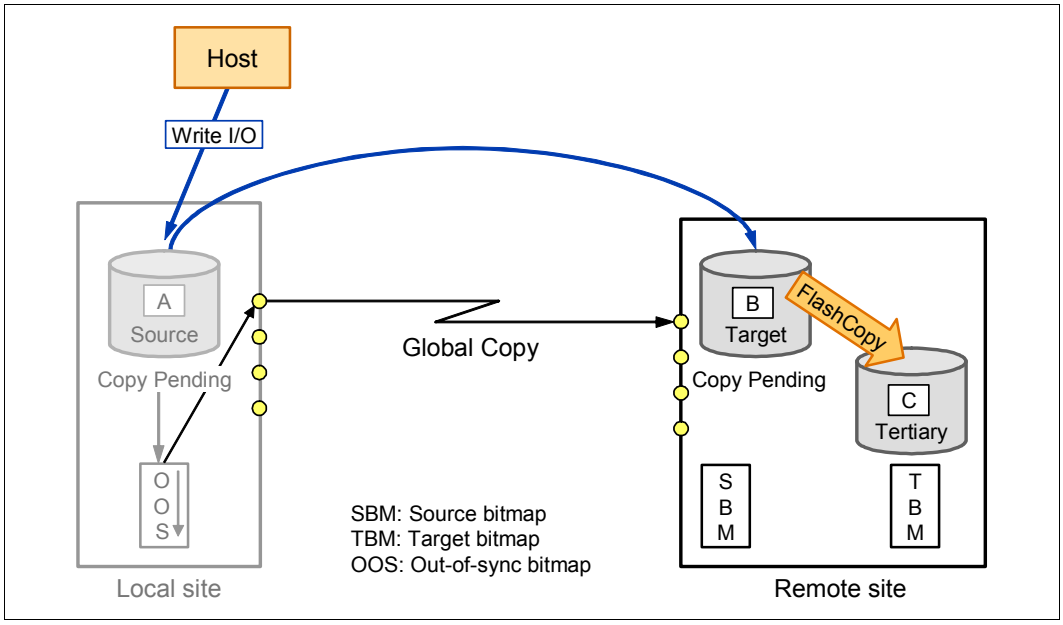


Figure 6-15 Global Mirror implementation

- Consistency group creation requires three steps automatically processed by DS8000 (Figure 6-16). After step 3 is complete, C volumes represents the consistency group and on these volumes, data are consistent.

  When consistency group creation is triggered, always by the source site, three steps occur:

  a. Serialize all Global Copy source volumes. This imposes a brief hold on all incoming write I/Os to all involved Global Copy source volumes. After all source volumes are serialized, the pause on the incoming write I/O is released and all further write I/Os are now noted in the change recording bitmap. They are not replicated until step 3 is done, but application write I/Os can immediately continue.

  b. Drain includes the process to replicate all remaining data that is indicated in the out-of-sync bitmap and still not replicated. After all out-of-sync bitmaps are empty, the third step is triggered by the microcode from the local site.

  c. Now the B volumes contain all data as a quasi point-in-time copy, and are consistent due to the serialization process in step 1 and the completed replication or drain process in step 2. Step 3 is now a FlashCopy that is triggered by the local system's microcode as an inband FlashCopy command to volume B as the FlashCopy source, and volume C as the FlashCopy target volume. Note that this FlashCopy is a two-phase process: First, the FlashCopy command to all involved FlashCopy pairs. Then the master collects the feedback and all incoming FlashCopy completion messages. When all FlashCopy operations are successfully completed, the master concludes that a new Consistency Group has been successfully created.

     FlashCopy applies here only to data changed since the last FlashCopy operation. This is because the enable change recording property was set at the time when the FlashCopy relationship was established. The FlashCopy relationship does not end due to the nocopy property, which is also assigned at FlashCopy establish time. Note that the nocopy attribute results in that the B volumes are not fully replicated to the C volumes by a background process. Bitmaps are maintained and updated instead.

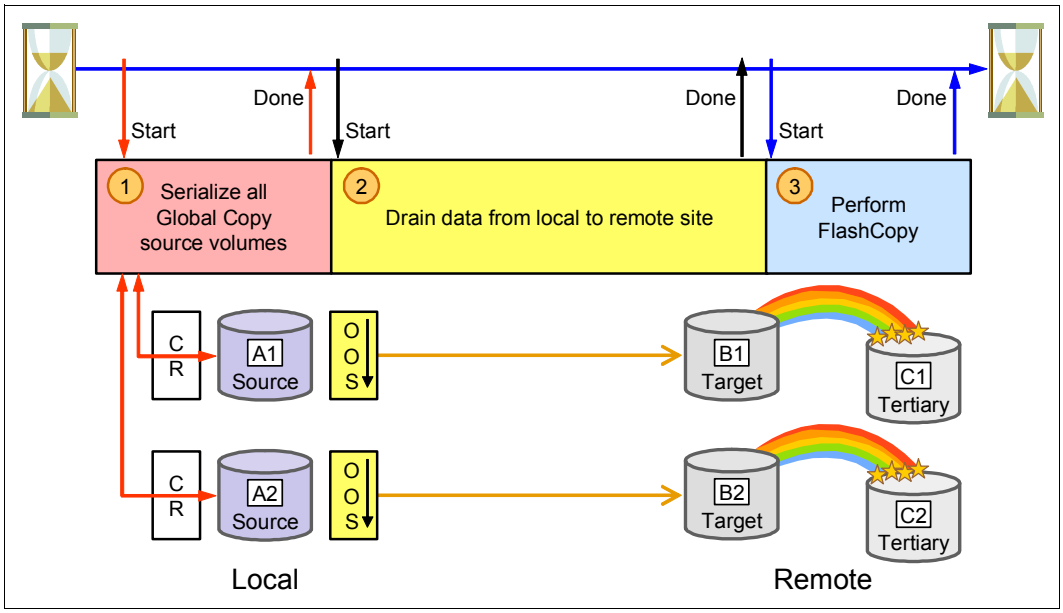


Figure 6-16 Global Mirror consistency group formation

There are several parameters used for Global Mirror tuning. See the section "Consistency Group interval time" in *IBM System Storage DS8000: Copy Services in Open Environments*, SG24-6788. It determines how long to wait before starting with the formation of a new

consistency group (for example, performing the three steps described above). The default is zero seconds. This means that consistency group creation happens constantly. As soon as a consistency group is created, the creation of a new one starts immediately.

6.3.2 Global Mirror operations

This is a brief review of basic operations that we can perform with Global Mirror.

Establishing Global Copy and FlashCopy relationships

These operations are needed before Global Mirror activation.

Global Copy

This operation establishes the copy relationship between the source (or local) volume and the target (or remote) volume. Normally, those two volumes reside on two DS8000s, but it is possible to use the same DS8000 for tests purposes. During the copy, the target volume gets a *target copy pending* state. For Global Mirror purposes, the Global Copy command with default parameters values can be used.

Note: A PPRC path must exist from the source to the target site before creating the Global Copy pair.

FlashCopy

Before creating the FlashCopy relationship on the target site, we recommend waiting until the end of the Global Copy initial copy. Because of the asynchronous nature of Global Copy, you cannot rely on the status, which will always be copy pending. However, as soon as the initial copy step is finished you will see the *out of sync tracks* indicator close to zero.

Specific FlashCopy settings (Target inhibit, Record, No copy, and Persistent (which is automatically set with Record)) are to be specified.

Creating a Global Mirror session

The session is an object that contains a reference to LSS, which is part of a Global Mirror relationship through its volumes. The session creation operation must be done for each LSS using an unique session ID for the entire storage image.

Adding Global Copy volumes to the Global Mirror session

Global Copy source volumes can be added at any time to the Global Mirror session. However, as we always reference the LSS with the session, added volumes must match LSS.

Starting a Global Mirror for a session

Starting a Global Mirror requests the source DS8000 to start creating the consistency groups.

Ending a Global Mirror for a session

Ending a Global Mirror requests the source DS8000 to stop creating the consistency groups. This operation tries to keep the last consistency group available.

Pausing and resuming a Global Mirror for a session

Pausing and resuming a Global Mirror can be used in complex situations, for example, if you want to run tests on target sites and keep normal business on source sites. This kind of scenario involves using a fourth set of volumes (called D volumes) on the target site.

Monitoring a Global Mirror for a session

Commands and GUI panels exist to help us monitor the Global Mirror status. For example, **showgmiroos** shows how many tracks are not synchronized from the source to the target.

Terminating a Global Mirror environment

This operation requires several steps to be requested in the proper order:

1. End Global Mirror.
2. Remove Global Copy volumes from the session.
3. Remove the Global Mirror session.
4. Terminate the FlashCopy pairs.
5. Terminate the Global Copy pairs.
6. Remove the paths.

Failover

Failover operations for Global Mirror involve more steps than for Metro Mirror. The very first one is the same as it is to reverse the Global Copy direction, so to make the target volumes (B volumes) as source suspended. But, those volumes are not consistent. We have to play with FlashCopy target volumes (C volumes), which are consistent. Therefore, these are the overall operations for a failover:

1. Global Copy failover: The situation becomes as shown in Figure 6-17.

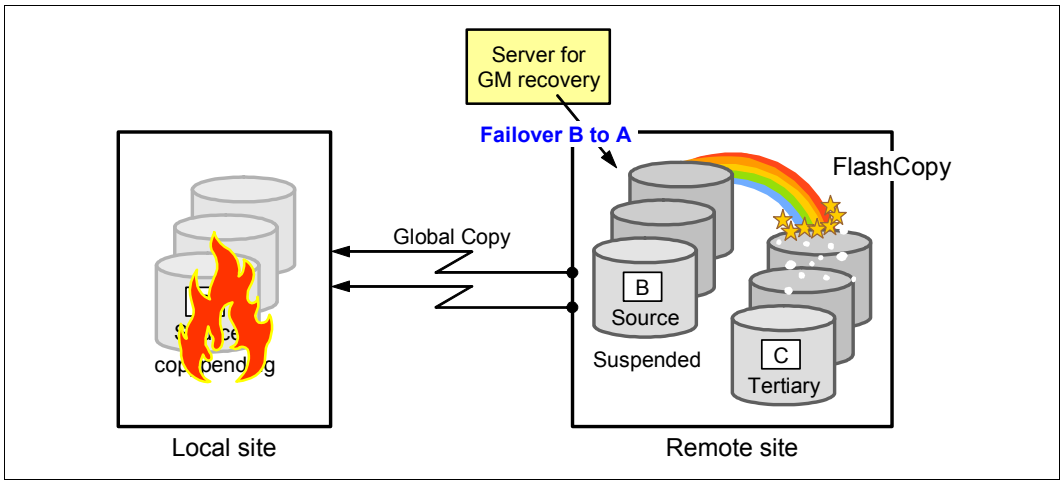


Figure 6-17 Global Mirror Global Copy failover

2. Set consistent data on B volumes.

   To make B volumes consistent, we use FlashCopy reverse, which copies content back to B from C volumes. This is a background copy that updates all changes on B since the last consistency group formation with the previous data they had stored on C volumes. This operation terminates the FlashCopy relationship between B and C volumes. The situation becomes as shown in Figure 6-18, but at the end there are no longer FlashCopy relationships.

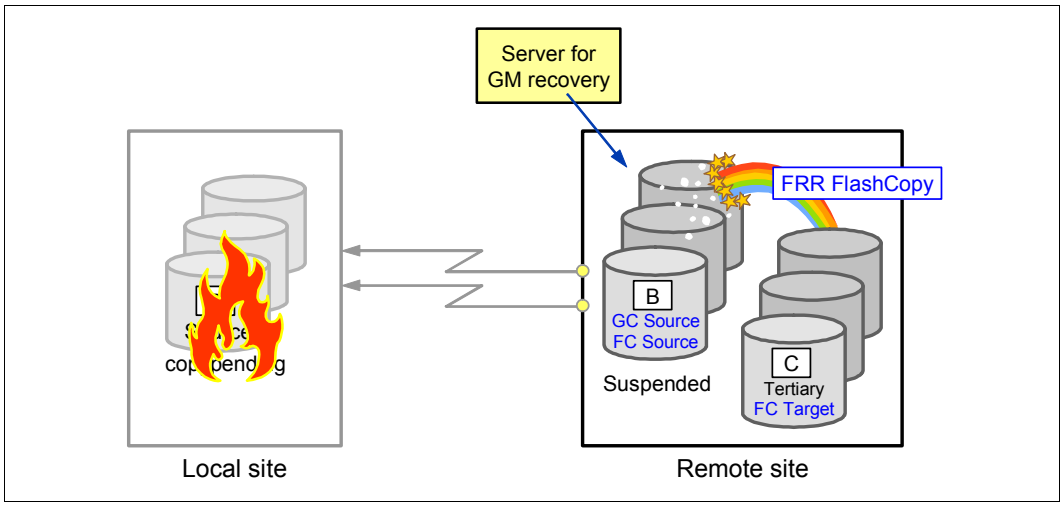


Figure 6-18 Global Mirror reverse FlashCopy

3. Re-establish a FlashCopy relationship between volumes B and C.

   Because of the termination of the FlashCopy relationship, we have to recreate it with the same settings that we used at the initial creation. At the end of this step we are back in the same situation as after the first step, but with consistent data on B volumes, and we can start host connections to these volumes.

Failback

When the source site becomes available, in a Metro Mirror environment, the first step is to run a failback operation from the target B volumes to synchronize the content of volume A with the content of volume B.

The second step is to make volume A source back and to start synchronizing B volumes with A volumes.

After Global Copy and FlashCopy are working again correctly, the last operation is to restart the Global Mirror session.

Note: Each time that a failover operation is done, applications can be quiesced before to have consistent data with no difference between the source and target data.

6.3.3 Global Mirror and IBM PowerHA SystemMirror for i

Within IBM PowerHA SystemMirror for i product, Global Mirror is used to replicate data from an IASP to another in a remote site, when the distance does not allow synchronous replication with Metro Mirror.

As with Metro Mirror replication, classical installation consists of a local IBM i partition and a remote IBM i partition. Both are in a cluster, and each partition makes use of a DS8000

external storage for IASP disks. There is no mandatory rule about SYSBAS disks. They can be internal drives or use external storage.

Note: Global Mirror relationships apply only to IASP volumes when used with IBM PowerHA SystemMirror for i.

Figure 6-19 shows a setup overview of a Global Mirror installation for an IASP. Both IBM i partitions are members of a cluster, and Global Mirror replication is in place for IASP volumes.

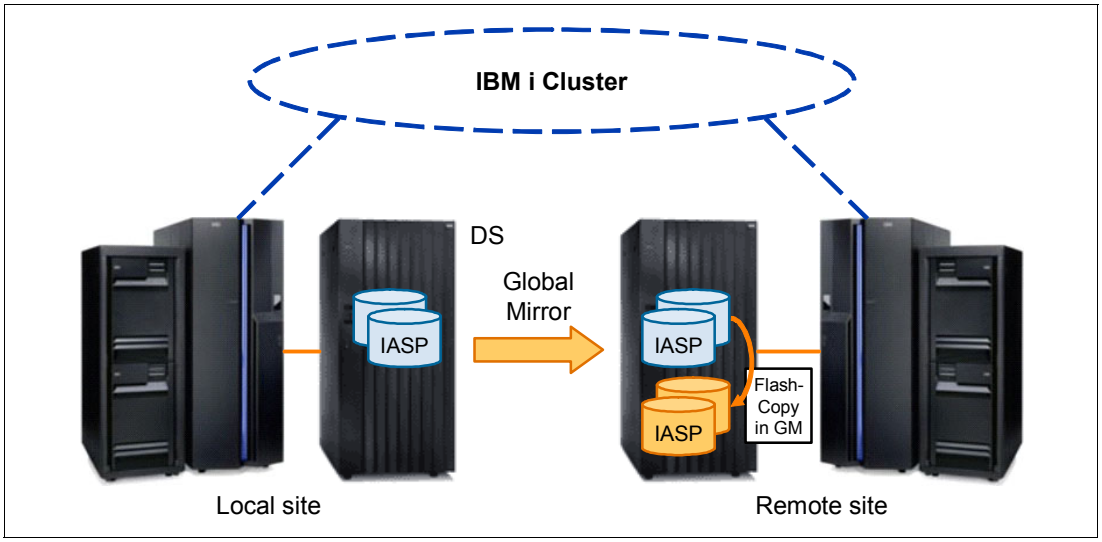


Figure 6-19 Global Mirror implementation

When a switch occurs, either in case of a planned outage or an unplanned one, the target IASP volumes become available due to Global Mirror failover, and the remote partition can use them for production activity.

When coming back to a normal situation is possible, failback operations can start, and the original source partition can start back production activity.

As you see in 13.1.3, “Configuring IBM i DS8000 Global Mirror (CL commands)” on page 311, IBM PowerHA SystemMirror for i handles all setup, switch, and switchback tasks but one. It is not able to establish the Global Mirror configuration between source and target volumes. The DS8000 setup is done either through DSCLI or Storage Manager GUI.

6.4 LUN-level switching

LUN-level switching is a new function provided by IBM PowerHA SystemMirror for i in IBM i 7.1 that allows for a local high availability solution with IBM System Storage DS8000 or DS6000 series similar to what used to be available as switched disks for IBM i internal storage.

With LUN-level switching single-copy, IASPs managed by a cluster resource group device domain located in IBM System Storage DS8000 or DS6000 series can be switched between IBM i systems in a cluster (Figure 6-20).

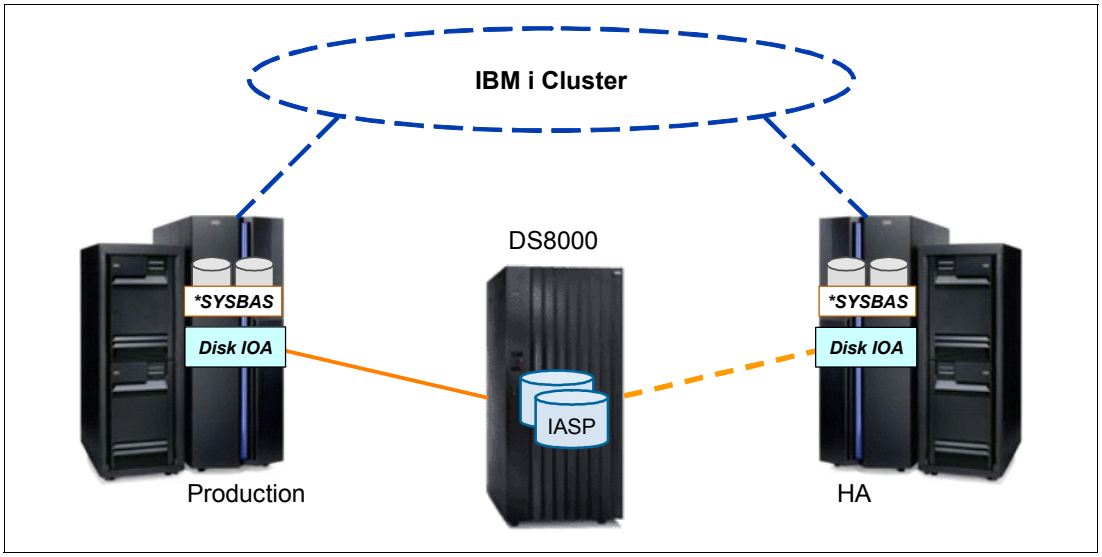


Figure 6-20 LUN-level switching between servers

A typical implementation scenario for LUN-level switching is where multi-site replication using Metro Mirror or Global Mirror is used for disaster recovery and protection against possible storage subsystem outages, while additional LUN-level switching at the production site is used for local high-availability protection, eliminating the requirement for a site-switch in case of potential IBM i server outages.

For implementation of LUN-level switching, an ASP copy description needs to be created for each switchable IASP using **ADDASPCPYD**, which has been enhanced with recovery domain information for LUN-level switching (Figure 6-21).

```
                    Add ASP Copy Description (ADDASPCPYD)

Type choices, press Enter.

Logical unit name:               LUN
  TotalStorage device  . . . . .                *NONE
  Logical unit range . . . . . .
                         + for more values
  Consistency group range  . . .
                         + for more values
Recovery domain:                 RCYDMN
  Cluster node . . . . . . . . .                *NONE
  Host identifier  . . . . . . .
                         + for more values
  Volume group . . . . . . . . .
                         + for more values
                         + for more values




                                                                        Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 6-21 IBM i ADDASPCPYD enhancement for LUN-level switching

An ASP session is not required for LUN-level switching, as there is no replication for the IASP involved.

You can see a more detailed scenario in 13.1.4, “Configuring IBM i DS8000 LUN-level switching” on page 317.

Note: Setting up an ASP copy description for LUN-level switching is only supported from the green-screen interface.

For LUN-level switching, the backup node host connection on the DS8000 or DS6000 storage system must not have a volume group (VG) assigned. PowerHA automatically unassigns the VG from the production node and assigns it to the backup node at site switches or failovers.

6.5 FlashCopy

In this section we introduce the DS8000 FlashCopy feature when used in IBM i environments. For complete information, see Part 3, “FlashCopy,” in *IBM System Storage DS8000: Copy Services in Open Environments*, SG24-6788.

FlashCopy overview

FlashCopy allows you to create a point-in-time copy of the logical volume. By doing a FlashCopy, a relationship is established between a source volume and a target volume. Both are considered to form a FlashCopy *pair*. As a result of the FlashCopy, either all physical blocks from the source volume are copied or, when using the *nocopy* option, only those parts are really copied that have changed in the source data since the FlashCopy has been established. The target volume needs to be the same size or bigger than the source volume whenever FlashCopy is used to flash an entire volume.

Typically, large applications such as databases have their data spread across several volumes, and their volumes should all be FlashCopied at exactly the same point-in-time. FlashCopy offers consistency groups, which allows multiple volumes to be FlashCopied at exactly the same instance.

Establishing a FlashCopy relationship

When the FlashCopy is started, the relationship between source and target is established within seconds by creating a pointer table, including a bitmap for the target.

If all bits for the bitmap of the target are set to their initial values, it means that no data block has been copied so far. The data in the target is not modified during setup of the bitmaps. At this first step, the bitmap and the data look as illustrated in Figure 6-22.

The target volume in the following figures can be a normal volume or a virtual volume (space efficient volume). In both cases the logic is the same.

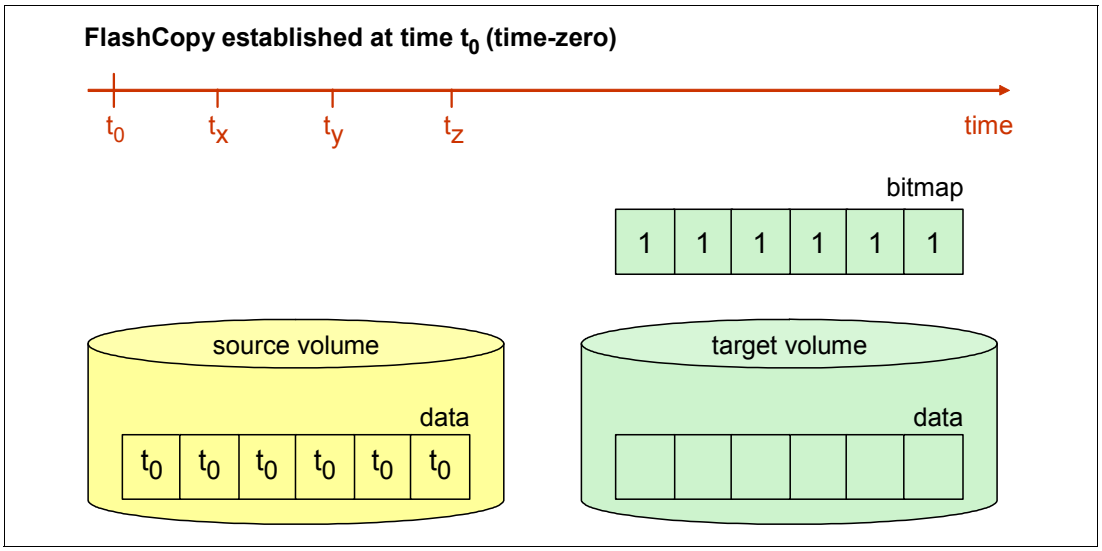


Figure 6-22 FlashCopy at time t0

After the relationship has been established, it is possible to perform read and write I/Os on both the source and the target. Assuming that the target is used for reads only while production is ongoing, the process will work as illustrated in Figure 6-23.

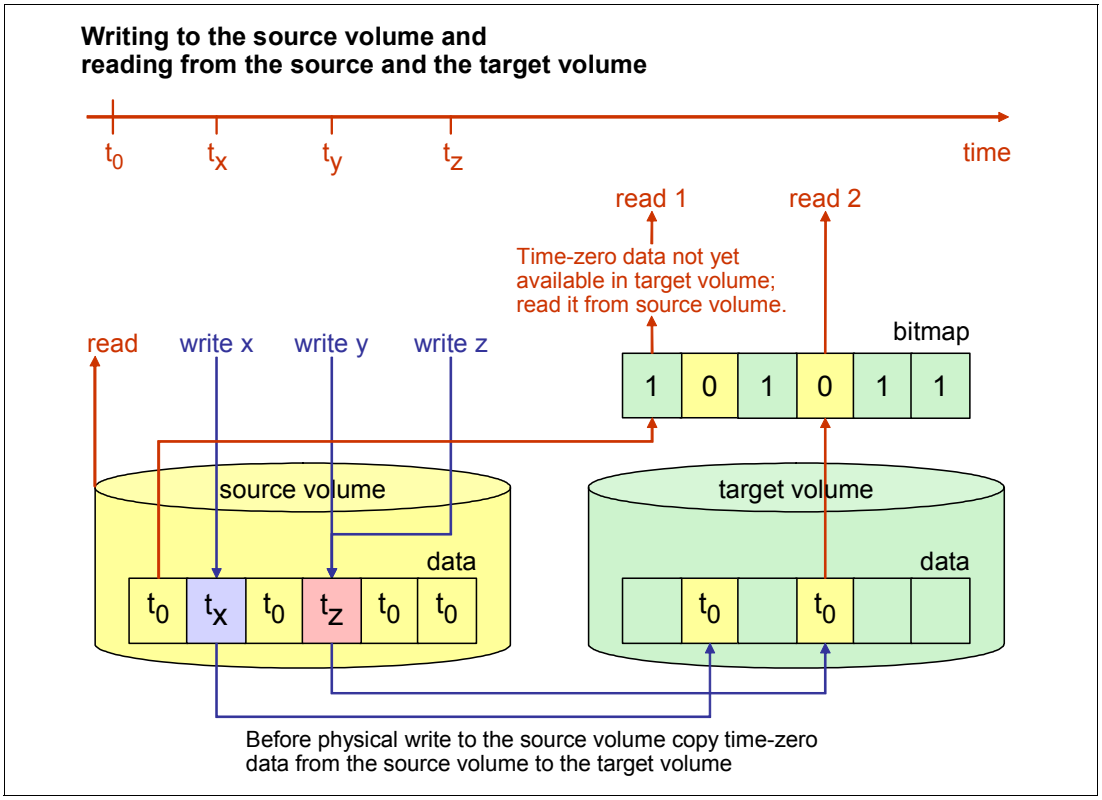


Figure 6-23 Reads from source and target volumes and writes to source volume

Reading from the source

The data is read immediately (Figure 6-23).

Writing to the source

Whenever data is written to the source volume while the FlashCopy relationship exists, the Storage system makes sure that the time-zero-data is copied to the target volume prior to overwriting it in the source volume. When the target volume is a space-efficient volume, the data is written to a repository.

To identify whether the data of the physical track on the source volume needs to be copied to the target volume, the bitmap is analyzed. If it identifies that the time-zero data is not available on the target volume, then the data will be copied from source to target. If it states that the time-zero data has already been copied to the target volume then no further action is taken.

It is possible to use the target volume immediately for reading data and also for writing data.

Reading from the target

Whenever a read request goes to the target while the FlashCopy relationship exists, the bitmap is used to identify whether the data has to be retrieved from the source or from the target. If the bitmap states that the time-zero data has not yet been copied to the target, then the physical read is directed to the source. If the time-zero data has already been copied to the target, then the read will be performed immediately against the target (Figure 6-23).

Writing to the target

Whenever data is written to the target volume while the FlashCopy relationship exists, the storage subsystem makes sure that the bitmap is updated. This way the time-zero data from the source volume never overwrites updates done directly to the target volume. Figure 6-24 illustrates he concept of writes to the target volume.

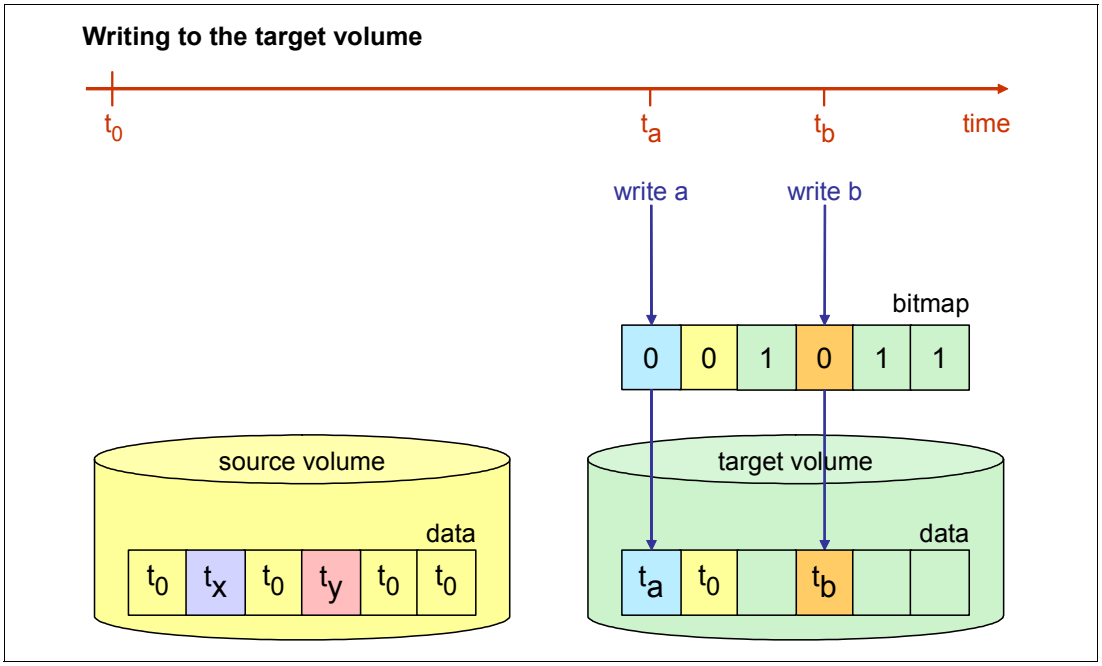


Figure 6-24 Writes to target volume

Terminating the FlashCopy relationship

The FlashCopy relationship is *automatically ended* when all tracks have been copied from the source volume to the target volume, if the FlashCopy option was to copy the data. The relationship can also be *explicitly withdrawn* by issuing the corresponding commands.

A FlashCopy space-efficient relationship ends when it is withdrawn. When the relationship is withdrawn there is an option to release the allocated space of the space efficient volume.

Full volume copy

When the *copy* option is invoked and the establish process completes, a background process is started that copies all data from the source to the target. If not explicitly defined as *persistent*, the FlashCopy relationship ends as soon as all data is copied.

Only the classical FlashCopy allows a full copy. FlashCopy SE has no such function. But remember, both features can coexist.

Nocopy option

If FlashCopy is established using the *nocopy* option, then the result will be as shown in Figure 6-22 on page 108, Figure 6-23 on page 109, and Figure 6-24.

The relationship will last until it is explicitly withdrawn or until all data in the source volume has been modified. Blocks for which no write occurred on the source or on the target will stay as they were at the time when the FlashCopy was established. If the *persistent* FlashCopy option was specified, the FlashCopy relationship must be withdrawn explicitly.

6.6 FlashCopy SE

PowerHA for SystemMirror for i with IBM i 7.1 newly supports space-efficient FlashCopy of the IBM System Storage DS8000 series.

The IBM System Storage DS8000 series FlashCopy SE licensed feature allows creation of space-efficient FlashCopy target volumes that can help to reduce the required physical storage space for the FlashCopy target volumes. These volumes are typically needed only for a limited time (such as for the duration of a backup to tape).

A space-efficient FlashCopy target volume has a virtual storage capacity reported to the host matching the physical capacity of the fully provisioned FlashCopy source volume, but no physical storage space is ever allocated. Physical storage space for space-efficient FlashCopy target volumes is allocated in 64 KB track granularity. This is done on demand for host write operations from a configured repository volume shared by all space-efficient FlashCopy target volumes within the same DS8000 extent pool (Figure 6-25).

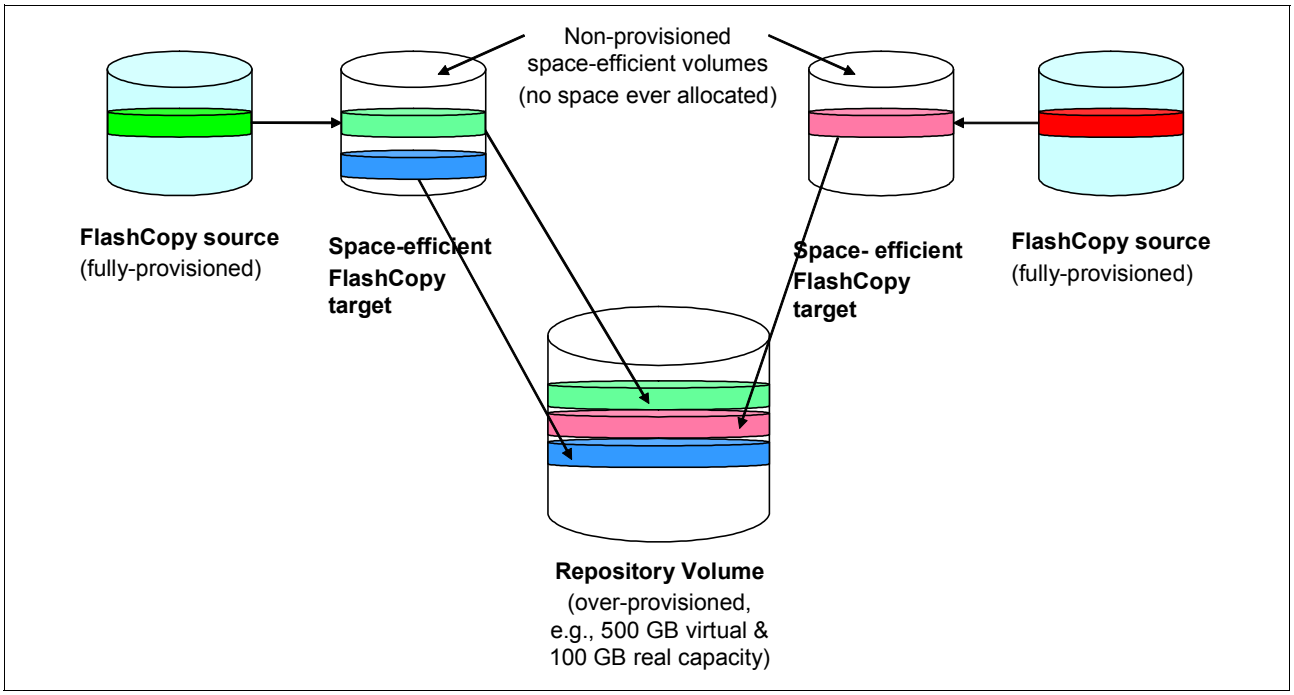


Figure 6-25 DS8000 space-efficient FlashCopy

From a user perspective, the PowerHA setup (not the DS8000 FlashCopy setup) for space-efficient FlashCopy is identical to the setup for traditional FlashCopy with the no-copy option. The reason for this is that PowerHA SystemMirror for i internally interrogates the DS8000 to determine the type of FlashCopy relationship and makes sure that it uses the corresponding correct DSCLI command syntax. The syntax check is done for either traditional FlashCopy or FlashCopy SE when issuing **`mkflash`** and **`rmflash`**.

For further information about using IBM System Storage DS8000 FlashCopy SE with IBM i, see IBM Redbooks publication *IBM System Storage Copy Services and IBM i: A Guide to Planning and Implementation*, SG24-7103.

7

Storwize V7000 and SAN Volume Controller Copy Services

This chapter provides an overview of the IBM Storwize V7000 and SAN Volume Controller storage concepts and their Copy Services functions Metro Mirror, Global Mirror, and FlashCopy.

7.1 Storwize V7000/SAN Volume Controller storage concepts

IBM Storwize V7000 and IBM System Storage SAN Volume Controller both provide in-band block-level storage virtualization. The key concept of storage virtualization is to decouple the logical storage representation as seen by the application servers from the underlying physical storage by abstracting the physical location of the data. This abstraction is achieved by the SVC/V7000 using a storage virtualization-layer, which also enables application server transparent advanced functions for virtualized volumes like volume mirroring, volume migration, FlashCopy creation, or thin provisioning. These advanced functions do not rely in any way on the functionality provided by the underlying disk storage system.

7.1.1 Hardware overview

The *SAN Volume Controller (SVC)* is a highly scalable storage virtualization appliance that provides the capability to virtualize external SAN-attached storage. Apart from optional internal Solid State Drives (SSDs), the SVC has no integrated internal storage. An SVC appliance consists of at least two and up to eight SVC nodes, each with their own dedicated uninterruptible power supply (UPS) building an SVC cluster. Within the cluster there are always two nodes, each paired into an *I/O group* that describes a node-pair with redundant write cache responsible for processing the I/O requests for the volumes associated with that particular I/O group. The latest hardware generation of the SVC, IBM device type 2145, consists of model CG8 nodes based on IBM System x3550 M3 quad-core servers, each providing 24 GB cache, four 8 Gb FC ports, and optionally either up to four SSDs or a dual-port 10 Gb Ethernet card.

The *Storwize V7000* is a modular storage system that includes the capability to virtualize external SAN-attached storage in addition to its own internal storage. The V7000 is built upon the IBM SAN Volume Controller (SVC) technology base using corresponding storage concepts and management interfaces with generally the same command set. From a hardware perspective, the V7000 IBM device type 2076 consists of two dual-active controllers with 8 GB of cache each, eight 8 Gb FC ports plus four optical 10 Gb FCoE ports, up to 24 disks in the controller enclosure, and up to nine SAS-chain attached expansion enclosures with up to 24 disks each, for a supported maximum of 240 internal small form factor (SFF) disk drives. With the release of V6.2 microcode, a V7000 clustering option was introduced to increase the V7000 scalability by allowing you to combine two V7000 systems into a V7000 cluster with two I/O groups.

7.1.2 Storage virtualization

SVC and V7000 share the same storage virtualization concept as that shown in Figure 7-1. They implement an indirection, or *virtualization*, layer in a Fibre Channel fabric. The virtualization layer mediates between the physical storage in RAID controllers presented as SCSI LUNs known as *managed disks* or *MDisks* to the SVC/V7000 and *virtual disks* (*VDisks*) or *volumes* presented to the application servers or hosts. The SVC/V7000 pools the managed physical storage and provides parts of that *storage pool* or *MDisk group* to application servers, in a way that makes the servers think that they are just using disks in the normal way.

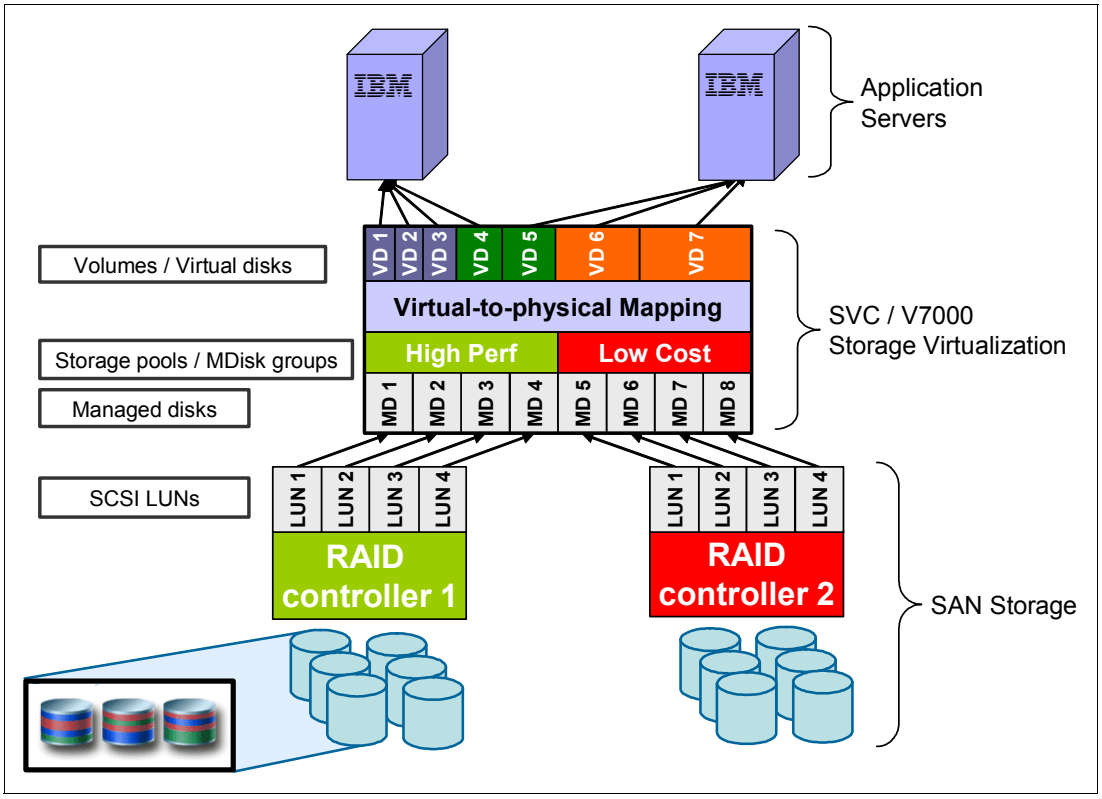


Figure 7-1 SVC and V7000 storage virtualization

All virtual disks seen by the application servers are mapped to the physical storage in the pools using a virtualization map that is fully transparent to the servers. Using this virtualization map, the SVC/V7000 can duplicate the server data to make instant copies or copies at a remote site, mirror data onto multiple physical disks to improve availability, migrate data concurrently between physical storage systems, or over allocate its physical storage to provide thin-provisioned virtual disks to the application servers and save physical storage cost.

Note: IBM i supports SVC/V7000 thin-provisioned or space-efficient volumes only for FlashCopy target volumes.

The SCSI LUNs created on the physical storage systems typically configured as a single LUN per RAID array are discovered as MDisks by the SVC/V7000 and assigned by the storage administrator to storage pools/MDisk groups usually based on common performance and availability device characteristics. Up to 128 storage pools with up to 128 MDisks each and up to 4096 MDisks per cluster from up to 64 storage systems are supported. At creation of the storage pool or MDisk group, an extent size ranging from 16 MB - 8 GB has to be specified, which is used by the SVC/V7000 to address the storage pool and determines the maximum

managed storage capacity in the cluster (from 64 TB to 32 PB) but is not relevant for the cluster's storage performance itself. Virtual disks (VDisks) are created from the extents of a storage pool with up to 2048 VDisks supported per I/O group, that is, cluster node pair.

VDisks are created as either *image mode*, *sequential mode*, or *striped mode* VDisks (Figure 7-2). By default, VDisks are created in striped mode, meaning that the extents for the VDisks are allocated in a round-robin fashion from the available MDisks within the MDisk group. That is, the strip size is identical to the extent size. Sequential mode VDisks have their extents allocated sequentially from the available MDisks, meaning that another MDisk is used for VDisk extent allocation only if the previous MDisk is already fully allocated. Image mode VDisks represent a one-to-one mapping between MDisks and VDisks and are typically used for migrating storage to (or from) the SVC/V7000.

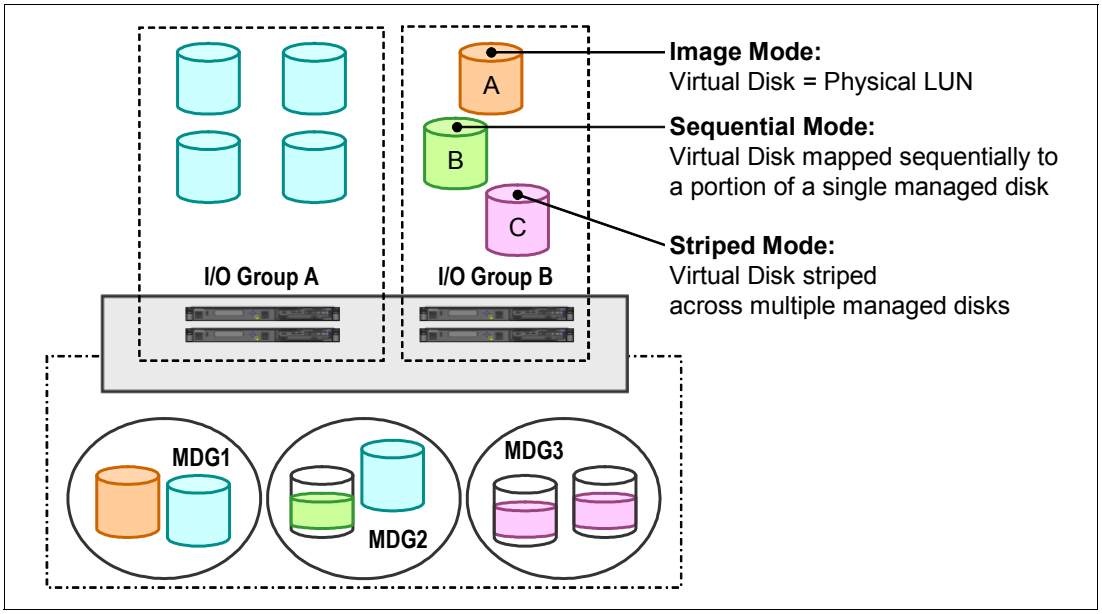


Figure 7-2 SVC and V7000 VDisk modes

7.1.3 I/O processing

As previously mentioned, the SVC/V7000 processes host I/O for a particular volume (VDisk) only within an I/O group consisting of two cluster nodes, which also serve as a backup for each other in case one cluster node becomes unoperational, for example, during microcode updates or failure conditions. Under normal conditions, I/O for a specific volume is always processed by the same node designated as the *preferred node*. When VDisks are created for an I/O group, they are automatically evenly distributed in a round-robin fashion between both nodes of the I/O group.

For host write I/O processing, both nodes of an I/O group are involved (Figure 7-3 on page 117) such that the host sends its write I/O request down the preferred path to the preferred node, and the SVC/V7000 preferred node stores the write data in its UPS backed-up write cache and propagates a second copy of the write data via the SAN fabric to the alternate (non-preferred) node in the I/O group. After the write data has been stored in both nodes' write cache, it is acknowledged by an I/O complete message to the host and asynchronously destaged by the preferred node to the physical disks. In case one of the nodes of an I/O group becomes unoperational, the remaining node goes into *write-through mode*, meaning that because the write cache redundancy has been lost, the host write I/O is not acknowledged until it has been written to the physical disks with corresponding host write I/O performance implications.

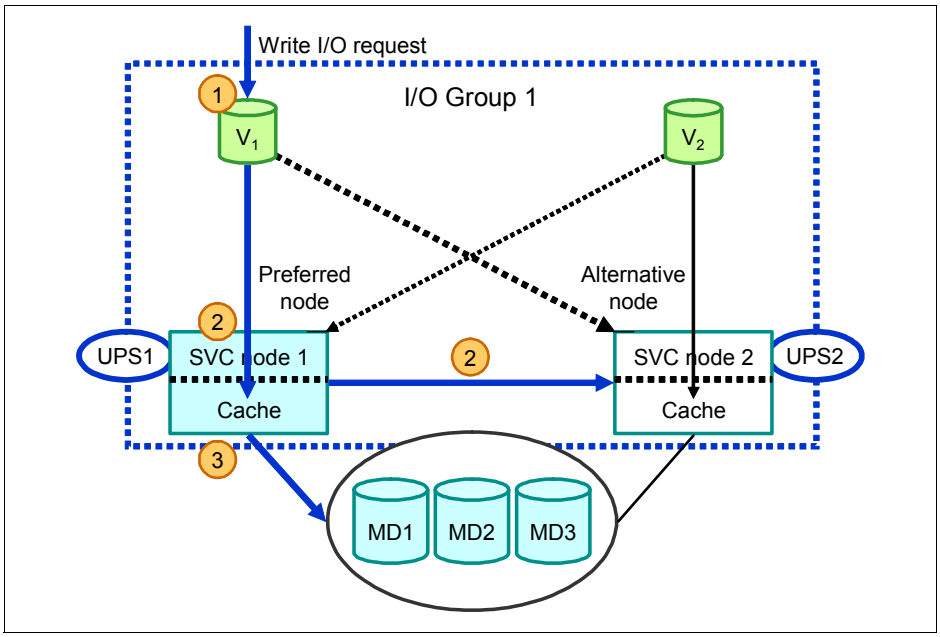


Figure 7-3 SVC and V7000 I/O processing

Because both nodes of an I/O group are responsible for I/O processing, it is important that the host systems are always zoned to both nodes of an I/O group and have the SDDPCM multi-pathing software for SVC/V7000 installed.

One of the advanced functions enabled by the SVC/V7000 virtualization-layer is *volume (VDisk) mirroring* for which two copies[1] of extents are kept for a volume (VDisk), but the volume itself is still represented only once to the application servers. For availability reasons, the two copies are typically allocated on MDisks in different MDisk groups, which are ideally from two different storage systems. As volume mirroring is implemented in the SVC/V7000 I/O stack below the cache and Copy Services, neither the host application servers nor Copy Services functions are aware that the volume is mirrored so that any Copy Services or migration functions work the same for mirrored as for non-mirrored volumes. Important from an I/O processing point of view is that under normal conditions (that is, both volume copies are available), host I/O writes are directed to both copies, but read I/O is directed only to the primary volume, which is by default the first created copy. The location of the primary volume can also be changed by the user to either account for load-balancing or possibly different performance characteristics for the storage of each copy.

Regarding the I/O to the disk storage system, the SVC/V7000 uses an integrated multi-path driver within its microcode to direct the I/O for a specific LUN to a single storage system port using a round-robin algorithm to distribute the workload from different LUNs to different storage system ports. All LUNs that are part of an MDisk group or storage pool (that is, that are *managed* MDisks) must be available. Otherwise, the complete MDisk group is taken offline, except for an unavailable image mode MDisk, which does not impact the availability of other MDisks.

Due to these I/O processing characteristics of the SVC/V7000, certain considerations are required to ensure a proper SAN switch zoning without compromising redundancy or performance. See 8.1.5, “Storage considerations” on page 139, for further information about IBM PowerVM Virtual I/O Server requirements for SVC/V7000 attachment.

[1] In SVC/V7000 terminology, a *copy* refers to one instance of an entity like a volume, that is, a mirrored volume has two copies.

7.1.4 Copy Services

The SVC/V7000 offers a common platform and single point of control not only for regular provisioning and management of heterogeneous storage but also for advanced functions, such as Copy Services, that are enabled by the SVC/V7000 virtualization-layer also between storage systems of different architectures or from different vendors. Copy Services functions available for the SVC/V7000 are Metro Mirror for synchronous remote replication, Global Mirror for asynchronous remote replication, and FlashCopy for point-in-time volume copies that are each discussed in more detail in the following sections.

The SVC/V7000 Copy Services functions are licensed on a used versus installed capacity basis with a single combined license for Metro Mirror and Global Mirror. FlashCopy is included in the V7000 base license (5639-VM1), whereas a separate license for FlashCopy is required for the SVC based on the virtual capacity only of the source volumes.

Both intra-cluster (that is, within the same cluster (and within the same I/O group only, rather than used for testing or migration purposes)) and inter-cluster remote copy relationships (that is, between two SVC or V7000 systems) are supported by the SVC/V7000.

For inter-cluster remote Copy Services, two SVC or V7000 clusters are connected to each other over a Fibre Channel fabric. The *maximum supported distance* for remote Copy Services is determined by the maximum supported round-trip latency of 80 ms. Up to 10 km Fibre Channel distance is supported with long-wave SFPs. Extended distance support is provided by using extenders, routers, or DWDM/CWDM hardware with up to seven hop counts supported.

The source and target volumes for Copy Services relationships need to be of the same virtual size.

Interoperability between SVC and V7000 Copy Services is available with the latest SVC/V7000 V6.3 microcode and is also supported by 5799-HAS PRPQ PowerHA SystemMirror for i. Allowing a V7000 system to participate in a SVC remote copy cluster partnership requires its cluster layer property to be changed from storage to appliance.

For further details about SVC and V7000 refer to these IBM Redbooks publications:

- *Implementing the IBM System Storage SAN Volume Controller V6.1*, SG24-7933
- *Implementing the IBM Storwize V7000*, SG24-7938

7.2 Metro Mirror

Metro Mirror is a synchronous remote copy relationship between two SVC/V7000 volumes (VDisks) of equal (virtual) size. When a remote copy relationship (either Metro Mirror or Global Mirror) is established, the preferred primary volume is designated as the *master* volume and the preferred secondary volume as the *auxiliary* volume. As long as the secondary volume is available, every host write I/O sent to the Metro Mirror primary volume is acknowledged back to the host only after it has been committed to the write cache of the primary SVC/V7000 and the secondary SVC/V7000 system (Figure 7-4).

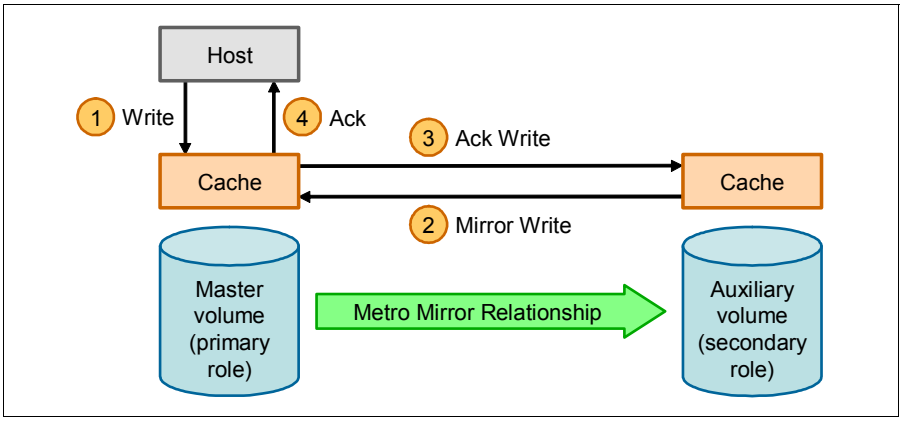


Figure 7-4 Metro Mirror SVC/V7000 write I/O processing

The *role* of a master or auxiliary volume is either primary or secondary, depending on the direction or failover state of the current remote copy relationship. Up to 2048 remote copy relationships are supported in a two-node SVC/V7000 cluster.

With SVC/V7000, establishing a Copy Services relationship is done in two phases of creating the relationship first before starting it in a second step. This is different from, for example, IBM System Storage DS8000 Copy Services, for which establishing a Copy Services relationship is done in a single step by creating the out-of-sync bitmaps and starting the relationship automatically at its creation.

When creating the Metro Mirror relationship, the user can specify whether the auxiliary volume is already in sync with the master volume, and the background copy process is then skipped. The in-sync option (path 1a in Figure 7-5 on page 120) is intended to be used when the volumes were created with the format option should not be used for IBM i volumes because IBM i specially formats the volumes itself when they are configured (that is, added to an IBM i ASP). Hence, using the SVC/V7000 format option at volume creation and therefore the in-sync option when creating a Metro Mirror relationship does not make sense.

7.2.1 Bandwidth thresholds

At initial synchronization, data is copied in data chunks, called *grains*, of 256 KB by a background copy process from the primary to secondary remote copy volume with a default bandwidth limit of 50 MBps between both SVC/V7000 clusters. This *partnership bandwidth limit* is evenly divided by the nodes in the cluster and is an attribute of the remote copy partnership between both SVC/V7000 systems. The bandwidth should be less than (when still accounting for host write I/O updates during synchronization) or equal to the available replication link bandwidth between both systems, when no relevant host write update are expected during synchronization. Also, this overall cluster partnership bandwidth should be chosen deliberately to not exceed the capabilities of the primary and secondary storage

systems to prevent performance impacts for the foreground host I/O. Additionally, there is also a *relationship bandwidth limit* for the maximum background copy rate for each remote copy relationship hat defaults to 25 MBps and is an attribute of the SVC/V7000 cluster configuration.

7.2.2 Remote copy relationship states

A Metro Mirror or Global Mirror remote copy volume relationship can be in one of these states:

- Consistent stopped
- Inconsistent stopped
- Consistent synchronized
- Inconsistent copying
- Idling

Figure 7-5 shows an overview of these states as they apply to a connected remote copy relationship and the conditions that cause a state transition.

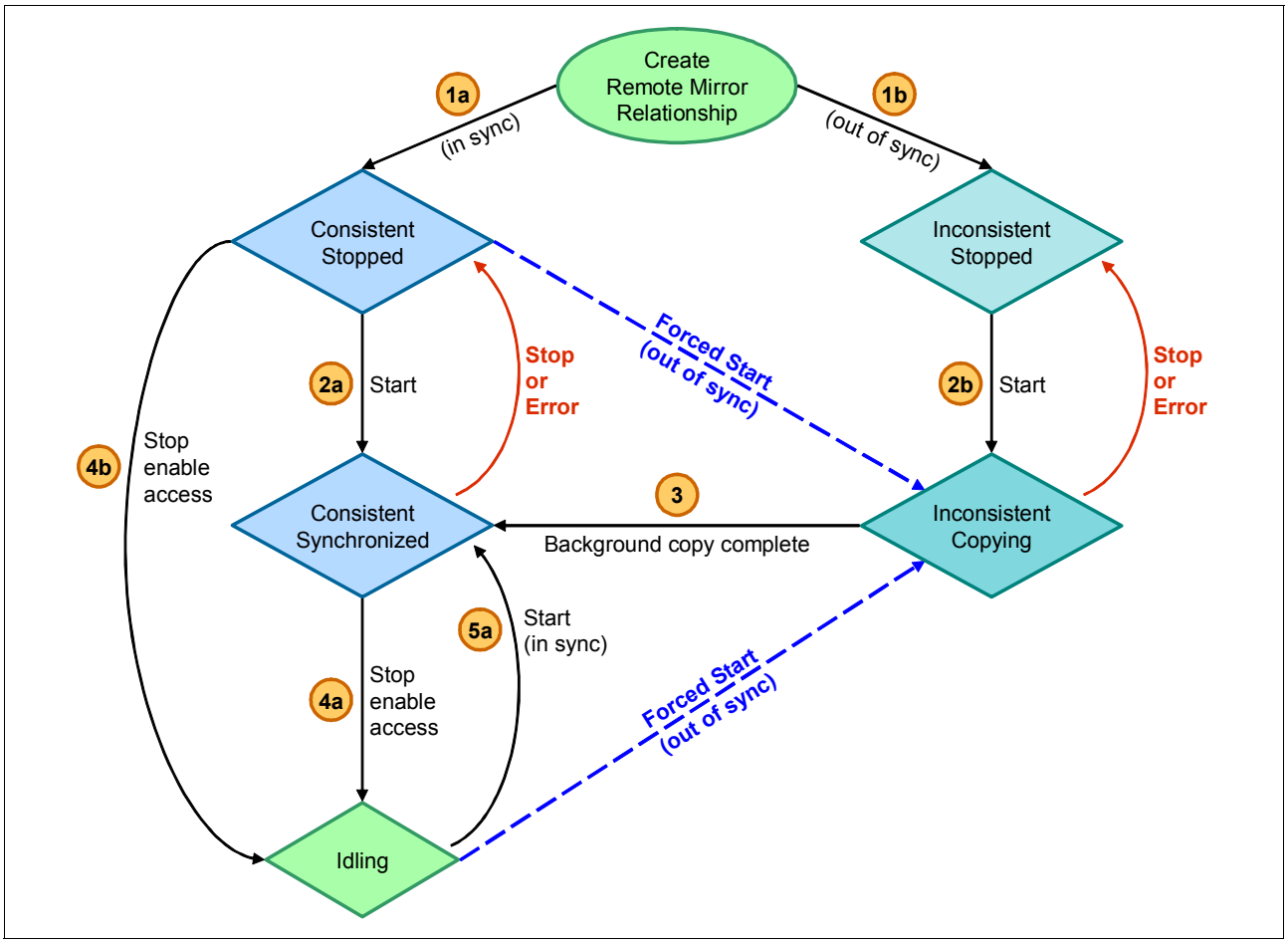


Figure 7-5 SVC/V7000 remote copy volume states and transitions

The remote copy states can be described as follows:

Inconsistent stopped State after creating a remote copy relationship (without using the in sync option) or after a failure condition that occurred while the relationship was in inconsistent copying state. Secondary volume data is not consistent with primary volume data and due to the risk of (undetected) data inconsistency should not be accessed by an application server.

Inconsistent copying State after starting an inconsistent stopped or idling relationship with changes to be synchronized with a background copy process running to copy data from the primary to the secondary volume. The primary is accessible for read and write I/O, but the secondary is offline (that is, not accessible for either read or write I/O), while the background copy is running and the relationship is not consistent.

Consistent synchronized State of an inconsistent copying relationship after completion of the background copy process or of a restarted consistent stopped relationship. The primary volume is accessible for read and write I/O, but the secondary volume is accessible only for read I/O. A switch of the remote copy direction does not change this state.
Stopping the relationship takes it to the consistent stopped state.
Stopping the relationship with the -access parameter takes it to the idling state.
Switching the relationship leaves it in the consistent synchronized state but reverses the primary and secondary roles.

Consistent stopped State after stopping a consistent synchronized relationship or after it encountered an error that forced a consistency freeze. The secondary contains a consistent image but might be out-of-date with respect to the primary, which might have received write updates from the host after the relationship entered this state. Restarting on a non-synchronized relationship that has had changes requires the -force CLI command parameter.

Idling State after stopping a consistent synchronized relationship with enabling write access to the secondary volume. Both master and auxiliary volumes operate in the primary role so that both master and auxiliary volumes are accessible for read and write I/O.

Changes are tracked for both the master and auxiliary volumes so that when starting the remote copy relationship again in the desired direction, specified by the required -primary CLI command parameter, only a partial synchronization for the changed grains is needed. Restarting on a non-synchronized relationship that has had changes requires the -force CLI command parameter.

In addition to these states that are valid for a connected remote copy relationship (that is, one where the primary system is still able to communicate with the secondary system), there is also a disconnected state of remote copy relationships where the primary system can no longer communicate with the secondary system. When the clusters can communicate again, the relationships automatically become connected again.

If the relationship or consistency group becomes disconnected, the primary volumes transition to *inconsistent disconnected*. The master side transitions to *idling disconnected*.

The SVC/V7000 logs informational events like remote copy relationship changes, loss of synchronization, or remote cluster communication errors in an error log for which SNMP traps, e-mail notification, or syslog server messages can be configured to trigger either automation or alert the user for manual intervention.

7.2.3 Consistency groups

Consistency is a concept of the storage system ensuring write I/O processing in exactly the same order in which write updates are received by the host and maintaining this order even with Copy Services relationships. It applies to a single relationship, but can also be applied to a set of relationships spanning multiple volumes by using consistency groups.

For Metro Mirror and Global Mirror remote copy relationships, maintaining the correct write order processing requires that in case of an error event causing loss of replication for only a subset of remote copy relationships of an application servers, remote write update processing for the non-affected remote copy relationships is automatically stopped to ensure application server data consistency at the secondary site.

For FlashCopy point-in-time volume copies, maintaining consistency and correct write order processing requires that write I/O to all volumes of an application server in a consistency group is temporarily put on hold until all FlashCopy volume relationships for the consistency group have been started. The storage system depends on the concept of dependent writes having been implemented in the application logic to ensure consistency across multiple volumes in a consistency group (for example, that a journal is updated with the intended database update before the database itself is actually updated). This application logic write dependency ensures that when a SCSI queue full status is set as part of the consistency group formation for a volume, further dependent application writes are put on hold by the application so that the storage system can proceed setting SCSI queue full status for all remaining volumes and therefore guarantee dependent write data consistency for all volumes in the consistency group. This write dependency concept still applies for IBM i with its single-level storage architecture, as IBM i SLIC storage management would hold off all I/O to a disk unit in an SCSI queue full condition but would not stop the I/O to other disk units that are still available for I/O operations.

Up to 127 FlashCopy consistency groups with up to 512 FlashCopy volume relationships in a consistency group are supported in a SVC/V7000 system. For Metro Mirror and Global Mirror, up to 256 remote mirror consistency groups are supported with no limit imposed for the number of either Metro Mirror or Global Mirror remote copy relationships other than the limit of 2048 volumes supported per I/O node pair.

PowerHA SystemMirror for i with the 5799-HAS PRPQ inherently uses consistency groups for SVC/V7000 FlashCopy relationships and requires them to be configured for Metro Mirror or Global Mirror relationships. Due to the IBM i single-level storage architecture, which stripes the data across all disk units of an ASP, consistency groups should be defined on an IASP group level.

Note: Stand-alone volume copy relationships and consistency groups share a common configuration and state model. That is, all volume copy relationships in a consistency group that is not empty have the same state as the consistency group.

7.3 Global Mirror

Global Mirror is an asynchronous remote copy relationship between two SVC/V7000 volumes (VDisks) of equal (virtual) size. When a remote copy relationship (either Metro Mirror or Global Mirror) is established, the preferred primary volume is designated as the *master* volume and the preferred secondary volume as the *auxiliary* volume. Every host write I/O sent to the Global Mirror primary volume is acknowledged back to the host after it has been committed to the write cache of both nodes for the corresponding I/O group of the primary SVC/V7000 system. At some later point in time (that is, asynchronously), this write update is sent by the primary SVC/V7000 to the secondary SVC/V7000 system (Figure 7-6). Global Mirror therefore provides the capability to perform remote copy over long distances, up to the maximum supported round-trip latency of 80 ms, exceeding the performance-related limitations of synchronous remote copy without host write I/O performance impacts caused by remote replication delays.

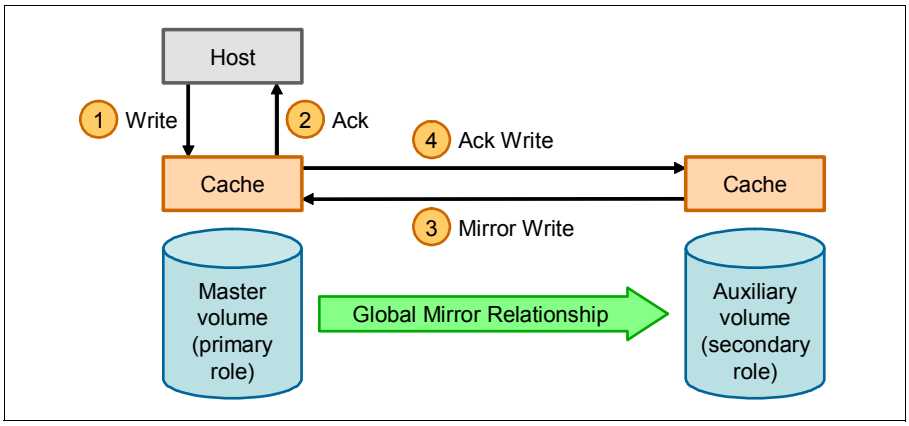


Figure 7-6 Global Mirror SVC/V7000 write I/O processing

Though the data is sent asynchronously from the primary to the secondary SVC/V7000, the write ordering is maintained by sequence numbers assigned to acknowledged host write I/Os and with the secondary applying writes in order by their sequence number. Consistency of the data at the remote site is maintained at all times. However, during a failure condition the data at the remote site might be missing recent updates that have not been sent or that were in-flight when a replication failure occurred, so using journaling to allow for proper crash consistent data recovery is of key importance, just like we generally recommend ensuring crash consistent data recovery even when not using remote replication to be able to recover from a loss of transient modified data in IBM i main store after a potential sudden server crash due to a severe failure condition.

Global Mirror volume relationship states and transitions are identical to those for Metro Mirror, as described previously in 7.2.2, “Remote copy relationship states” on page 120.

A log file is utilized by the SVC/V7000 for Global Mirror to maintain write ordering and help prevent host write I/O performance impacts when the host writes to a disk sector that is either currently in the process of being transmitted or due to bandwidth limits is still waiting to be transmitted to the remote site. The SVC/V7000 also makes use of shared sequence numbers to aggregate multiple concurrent (and dependent) write I/Os to minimize its Global Mirror processing overhead.

Global Mirror link tolerance

Global Mirror uses a special *link tolerance* parameter defined at the cluster level that specifies the duration with a default of 300 seconds, for which inadequate intercluster link performance with

write response times above 5 ms is tolerated. If this tolerated duration of degraded performance where the SVC/V7000 needs to hold off writes to the primary volumes with the effect of synchronous replication-like degraded performance is exceeded, it stops the most busy active Global Mirror relationship consistency group to help protect the application server's write I/O performance and logs an event with error code 1920. The link tolerance can be disabled by the user setting its value to 0. However, this provides no protection for the application server's write I/O performance anymore in cases where there is congestion on either the replication link or the secondary storage system. While the link tolerance setting allows you to define a period of accepted performance degradation, it is still important to properly size the remote copy replication bandwidth for the peak write I/O throughput and possible re-sync workload, and to help prevent longer production workload performance impacts.

The concept of consistency groups to guarantee write-dependent data consistency applies to Global Mirror the same as previously described for Metro Mirror. Consistency groups are required to be configured for PowerHA SystemMirror for i with the 5799-HAS PRPQ.

7.4 FlashCopy

The SVC/V7000 FlashCopy function provides the capability to perform a point-in-time copy of one or more volumes (VDisks). In contrast to the remote copy functions of Metro Mirror and Global Mirror, which are intended primarily for disaster recovery and high-availability purposes, FlashCopy is typically used for online backup or creating a clone of a system or IASP for development, testing, reporting, or data mining purposes.

FlashCopy is supported also across heterogeneous storage systems attached to the SVC/V7000, but only within the same SVC/V7000 system. Up to 4096 FlashCopy relationships are supported per SVC/V7000 system, and up to 256 copies are supported per FlashCopy source volume. With SVC/V7000 V6.2 and later, a FlashCopy target volume can also be a non-active remote copy primary volume, which eases restores from a previous FlashCopy in a remote copy environment by using the FlashCopy reverse function.

PowerHA SystemMirror i with the 5799-HAS PRPQ supports these SVC/V7000 FlashCopy functions, which are discussed in more detail below:

- FlashCopy no-copy and background copy
- Thin-provisioned (space-efficient) FlashCopy targets
- Incremental FlashCopy
- Reverse FlashCopy
- Multi-target FlashCopy (by using separate ASP copy descriptions for each target)

7.4.1 I/O indirection

The FlashCopy *indirection layer*, which is logically located below the SVC/V7000 cache, acts as an I/O traffic director for active FlashCopy relationships. To preserve the point-in-time copy nature of a FlashCopy relationship, the host I/O is intercepted and handled according to whether it is directed at the source volume or at the target volume, depending on the nature of the I/O read or write and whether the corresponding grain has already been copied. Figure 7-7 and Figure 7-8 illustrate the different processing of read and write I/O for active FlashCopy relationships by the indirection layer.

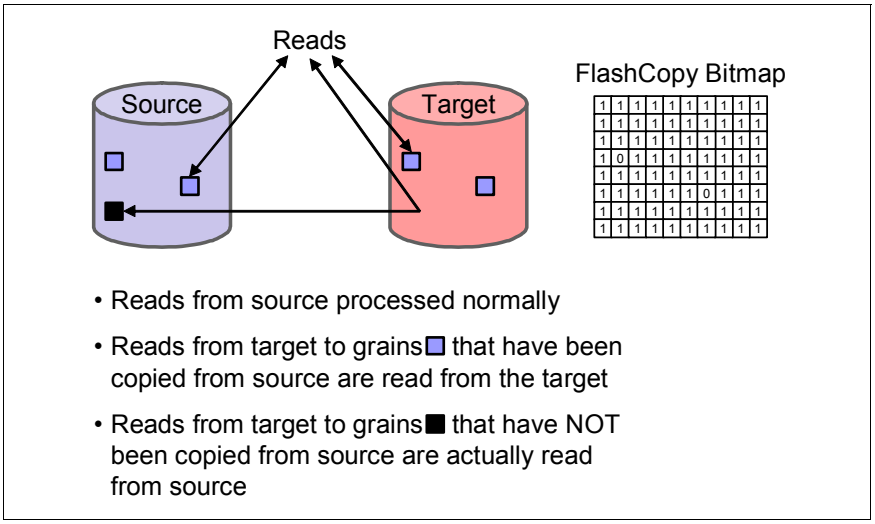


Figure 7-7 SVC/V7000 FlashCopy read processing

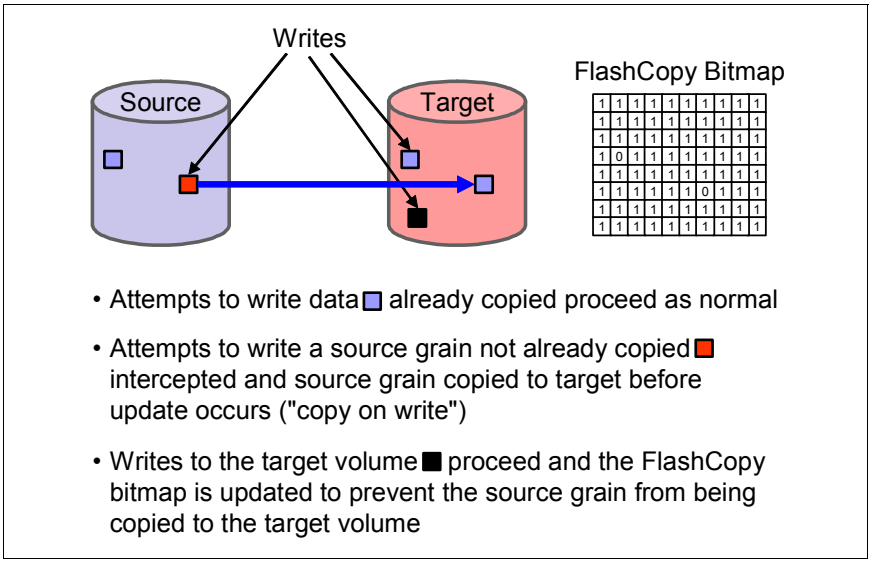


Figure 7-8 SVC/V7000 FlashCopy write processing

While a fixed grain size of 256 KB is used for remote mirror volume relationships for FlashCopy, the user can choose from the default grain size of 256 KB or alternatively from the smaller grain size of 64 KB as the granularity for tracking and managing out-of-sync data of a FlashCopy relationship.

The concept of consistency groups to ensure that dependent write data consistency across multiple volume copy relationships applies to FlashCopy (see 7.2.3, “Consistency groups” on page 122).

7.4.2 Background copy

A FlashCopy relationship can either be a no-copy or a background copy relationship. With a *background copy* relationship, any grain from the source volume is copied to the target volume. By default (that is, if not specifying the autodelete option), the relationship is retained even after all grains have been copied. For a *no-copy* relationship, only grains that have been modified on the source after starting the FlashCopy relationship are copied from the source volume to the target volume prior to the source grain being allowed to be updated (*copy on write* processing), provided that the corresponding grain on the target volume has not been updated already by the host accessing the target volumes.

An option for FlashCopy is creating an *incremental FlashCopy* relationship, which uses background copy to copy all of the data from the source to the target for the first FlashCopy and then only the changes hat occurred since the previous FlashCopy for all subsequent FlashCopies being started for the relationship.

When creating a FlashCopy relationship, the user can specify a desired *copy rate* for the background copy process, which can either be 0 (meaning that a FlashCopy no-copy relationship without a background copy is established) or any value from 1 - 100, which translates to the desired background copy throughputs (Table 7-1).

Table 7-1 FlashCopy background copy rates

Copy rate value	Data copied	256 KB grains/s	64 KB grains/s
1 - 10	128 KBps	0.5	2
11 - 20	256 KBps	1	4
21 - 30	512 KBps	2	8
31 - 40	1 MBps	4	16
41 - 50[a]	**2 MBps**	**8**	**32**
51 - 60	4 MBps	16	64
61 - 70	8 MBps	32	128
71 - 80	16 MBps	64	256
81 - 90	32 MBps	128	512
91 - 100	64 MBps	256	1024

a. Default value

A FlashCopy relationship is established on a SVC/V7000 system in three steps:

1. Creating a FlashCopy relationship

   This triggers the internal creation of a FlashCopy out-of-sync bitmap used by the SVC/V7000 for tracking the grains needing to be copied.

2. Preparing a FlashCopy relationship or consistency group

   This achieves consistency for the volumes by destaging the source volume's modified data to disk, putting it in write-through mode, discarding the target volume's cache data, and rejecting any I/O to the target volume.

3. Starting a FlashCopy relationship or consistency group

   This briefly pauses the I/O to the source volumes until all reads and writes below the SVC/V7000 cache layer have completed and starts the actual FlashCopy relationship. That is, the logical dependency between the source and target volume is established in the SVC/V7000 indirection layer.

7.4.3 FlashCopy relationship states

A FlashCopy volume relationship can be in any of the following states:

- Idle
- Copied
- Copying
- Stopped
- Stopping
- Suspended
- Preparing
- Prepared

The FlashCopy states are described here:

Idle or copied Mapping between source and target volume exists, but the source and the target behave as independent volumes.

Copying Background copy process is copying grains from the source to the target. Both the source and the target are available for read and write I/O, but the target depends on the source for grains not yet copied yet.

Stopped FlashCopy relationship was stopped either by the user or by an I/O error. The source volume is still assessable for read and write I/O, but the target volume is taken offline as data integrity is not provided. From a stopped state, the relationship can either be started again with the previous point-in-time image or lost or deleted if it is not needed anymore.

Stopping Relationship is in the process of transferring data to a depend relationship. The source volume remains accessible for I/O, but the target volume remains online if the background copy process completed or is put offline if the background copy process has not completed while the relationship was in copying state. Depending on whether the background copy completed, the relationship moves either to the idle/copied state or the stopped state.

Suspended State of a relationship when access to metadata used by the copy process got lost. Both the source and the target are taken offline and the background copy is put on hold. When metadata becomes available again, the relationship returns to the copying state or the stopping state.

Preparing Source volume is placed in write-through mode and modified data of the source volume is destaged from the SVC/V7000 cache to create a

consistent state of the source volume on disk in preparation for starting the relationship. Any read or write data associated with the target volume is discarded from the cache.

Prepared Relationship is ready to be started with the target volume having been placed in offline state. Write performance for the source volume can be degraded, as it is in write-through mode.

7.4.4 Thin-provisioned FlashCopy

In addition to regular, full-provisioned volumes (VDisks), which at their creation have the full physical storage capacity allocated corresponding to their volume capacity, the SVC/V7000 also supports *thin-provisioned* or *space-efficient* volumes, which are created with a virtual capacity reported to the host higher than the actual physical capacity pre-allocated to the volume. If the thin-provisioned volume is created with the *autoexpand* option, additional extents up to the virtual capacity of the volume are automatically allocated from the storage pool (MDisks group) as needed when the currently allocated physical capacity gets exhausted.

Note: Using thin-provisioned SVC/V7000 volumes for IBM i is only supported for FlashCopy *target* volumes, not for volumes directly assigned to an IBM i host.

Using thin-provisioned volumes also does not make sense for remote copy secondary volumes, which become fully allocated at initial synchronization. Similarly, using thin-provisioned volumes for FlashCopy targets only makes sense for FlashCopy no-copy relationships that are used for a limited duration and therefore have limited changes only.

For optimal performance of thin-provisioned FlashCopy, the grain size for the thin-provisioned volume used as the FlashCopy target should match the grain size of the FlashCopy relationship. Additionally, for space-efficiency reasons, to help minimize physical storage allocations on the thin-provisioned target, consider also using the small grain size of 64 KB.

7.4.5 Multi-target and reverse FlashCopy

As previously mentioned, a single FlashCopy source volume supports up to 256 target volumes. Creating and maintaining multiple targets from a FlashCopy source volume from different times might be useful, for example, to have multiple points to restore from by using the reverse FlashCopy function (Figure 7-9).

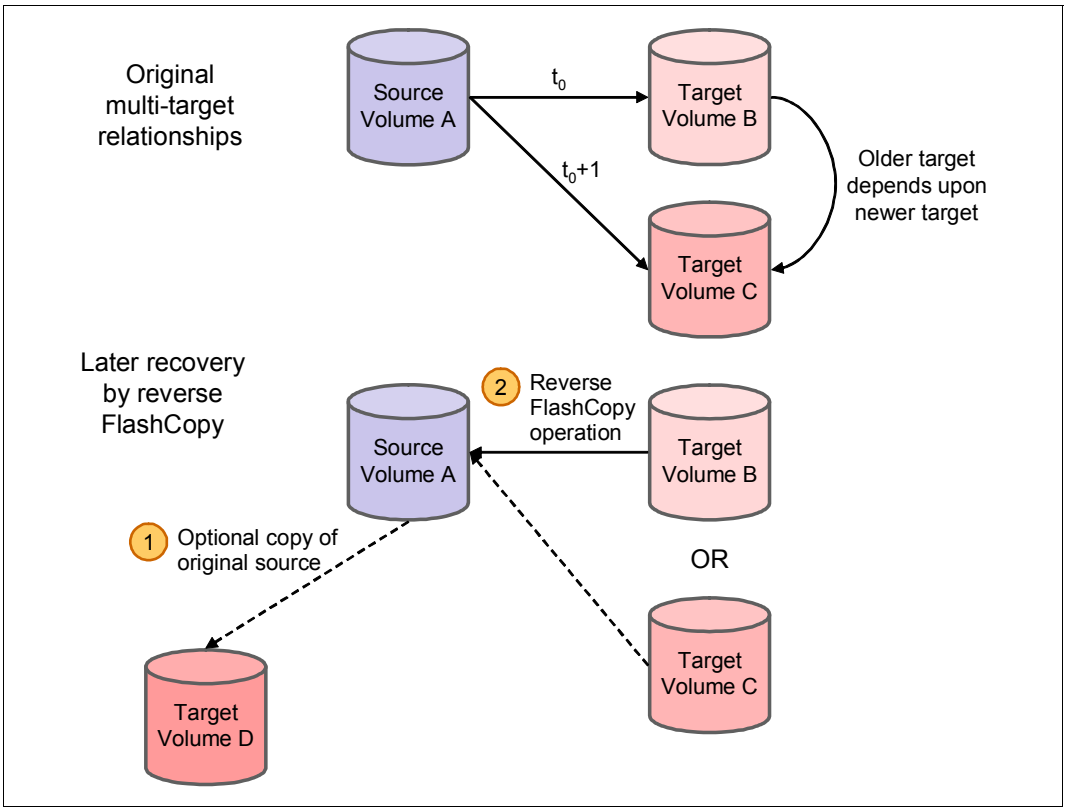


Figure 7-9 SVC/V7000 multi-target and reverse FlashCopy

A key advantage of the SVC/V7000 reverse FlashCopy function is that it does not destroy the original source and target relationship, so any processes using the target (such as tape backup jobs) can continue to run uninterrupted. It does not require a possible background copy process to have completed, and regardless of whether the initial FlashCopy relationship is incremental, the reverse FlashCopy function only copies data from the original target to the source for grains that were modified on the source or target.

Consistency groups cannot contain more than one FlashCopy relationship with the same target volume, so they need to be reversed by creating a set of new *reverse* FlashCopy relationships and adding them to the new reverse consistency group.

The SVC/V7000 performs the *copy on write* processing for multi-target relationships in a way that data is not copied to all targets, but only once from the source volume to the newest target so that older targets refer to newer targets first before referring to the source. Newer targets together with the source can therefore be regarded as composite source for older targets.

8

Planning for PowerHA

Good planning is essential for the successful setup and use of your server and storage subsystems. It ensures that you have met all of the prerequisites for your server and storage subsystems and everything that you need to gain advantage from best practices for functionality, redundancy, performance, and availability.

In this chapter, we discuss important planning considerations and all the prerequisites that you need to use IBM PowerHA SystemMirror for i in your environment.

8.1 Requirements for PowerHA

In this section we discuss the licensing information for IBM PowerHA SystemMirror for i and PRPQ 5799-HAS and the requirements for Power System, Virtual I/O Server, and Storage System.

8.1.1 Licensing considerations

Before installing the IBM PowerHA SystemMirror for i license product (5770-HAS), check whether the following is in place:

- IBM i 7.1 is installed on all system nodes (servers or logical partitions) that will be part of your high-availability or disaster-recovery solution.
- HA Switchable Resources (Option 41 of 5770-SS1) are installed on all system nodes (server or logical partitions) that will be part of your high-availability or disaster-recovery solution. As HA Switchable Resources is included in 5770-HAS PowerHA SystemMirror for i 7.1, you do not have to order it separately, but it still has to be installed separately.

The licensed program 5770-HAS must be licensed for all processor cores in the partitions (nodes) that you want to be part of your cluster.

There are two editions of PowerHA System Mirror for i, the Standard edition and the Enterprise edition (Table 8-1). The standard edition, which is PID 5770-HAS, is targeted at a data center HA solution. The Enterprise edition, which is a feature of 5770-HAS, adds support for a multi-site HA and DR solution. There is no charge to upgrade from PowerHA 6.1 to 7.1 Standard or Enterprise Edition.

Table 8-1 PowerHA SystemMirror for i Editions

IBM i HA/DR clustering	Standard edition	Enterprise edition
Centralized cluster management	✓	✓
Cluster resource management	✓	✓
Centralized cluster configuration	✓	✓
Automated cluster validation	✓	✓
Cluster admin domain	✓	✓
Cluster device domain	✓	✓
Integrated heartbeat	✓	✓
Application monitoring	✓	✓
IBM i event/error management	✓	✓
Automated planned fail over	✓	✓
Managed unplanned fail over	✓	✓
Centralized Flash Copy	✓	✓
LUN-level switching	✓	✓
Geomirror Sync mode	✓	✓
Geomirror Async mode		✓

IBM i HA/DR clustering	Standard edition	Enterprise edition
Multi-Site HA/DR management		✓
DS8000/SVC/V7000 Metro Mirror		✓
DS8000/SVC/V7000 Global Mirror		✓

The pricing of PowerHA SystemMirror 7.1 for i is based on a *per-core basis* and is broken down among small-tier, medium-tier, and large-tier Power server for each edition.

You can order a Power System *Capacity BackUp (CBU)* model for your disaster recovery and high availability when ordering a Power System or doing a model upgrade. The terms for the CBU for i allow a primary system processor's optional i license entitlements and 5250 Enterprise Enablement license entitlements (or optional user entitlements in the case of the Power 520 and 720 CBU) to be temporarily transferred to the CBU Edition when the primary system processor cores are not in concurrent use. For an IBM i environment, you must register the order for a new CBU at *IBM Capacity Backup for Power Systems* site:

http://www-03.ibm.com/systems/power/hardware/cbu/

A minimum of one IBM i processor entitlement or enterprise enablement is required on each primary machine and CBU machine at all times, including during the temporary move period. For machines that have IBM i priced per user, the minimum quantity of IBM i user entitlements for each machine is required on each primary machine and CBU machine at all times, including during the temporary move period.

The same applies to PowerHA SystemMirror entitlements. A minimum of one entitlement on the primary machine and one on the CBU machine is required at all times, including during the temporary move period.

In Figure 8-1, there are eight active cores with IBM i and PowerHA SystemMirror entitlements on the primary server. On the CBU box, we have one IBM i entitlement, one PowerHA entitlement, and a temporary key (good for two years) with a usage limit of eight cores for PowerHA SystemMirror.

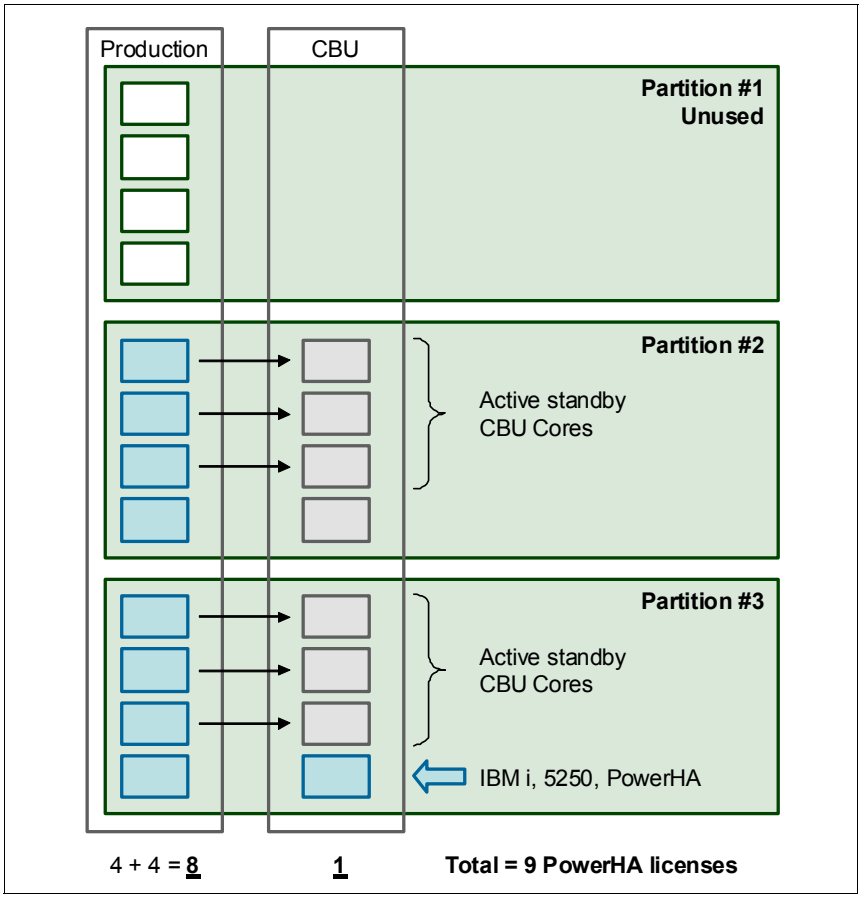


Figure 8-1 PowerHA licensing for CBU

8.1.2 PRPQ ordering information

You need to order 5799-HAS Program Request Pricing Quotation (PRPQ) for IBM PowerHA SystemMirror for i to obtain these enhancements in PowerHA:

- PowerHA GUI.
- SVC/V7000 remote Copy Service support Metro Mirror, Global Mirror, and FlashCopy.
- New CL commands:
  - **CFGDEVASP** to create an IASP
  - **CFGGEOMIR** to configure geographic mirroring

This PRPQ (5799-HAS) is available in the English language (2924) only but can be installed on other language systems. You must install the PRPQ with the LNG parameter as 2924 when you install PRPQ with **RSTLICPGM**. PTF SI44148 for LPP 5770-HAS is a prerequisite for the PRPQ.

The order for the product should be placed via the ordering system particular to a location (for example, AAS or WTAAS). Specify product ID 5799-HAS. See the following list for additional information by geography:

- United States

  PRPQs in the United States are ordered by the CSO via US ordering processes. If you have any problems ordering the PRPQ, contact the Special Product Marketing representative.

- EMEA

  For EMEA ordering information, see SWORDERINFO (EPLC) on HONE. For price information see the HONE SW Price application. For further assistance with EMEA ordering information contact SDFMAIL@dk.ibm.com.

- Asia Pacific

  In Asia Pacific, submit your request via Access GFS at:

  `http://cqrfa.vienna.at.ibm.com:8080/agfs/access/indexservlet`

  In Japan assistance is available from Process Fulfillment at EB49010@jp.ibm.com.

- Canada

  Assistance is available from Canada RPQ/Austria/IBM.

- Latin America

  To order PRPQs in Latin America, contact the responsible fulfillment account representative who normally orders your equipment via AAS. If you have any problems ordering the PRPQ, contact your sales manual team in Vienna.

8.1.3 Power Systems requirements

IBM PowerHA SystemMirror for i requires IBM i 7.1. If the hardware is not supported by IBM i 7.1, IBM PowerHA SystemMirror for i is not supported. The following platforms supported by IBM i 7.1 are able to use IBM PowerHA SystemMirror for i:

- Power Systems servers and blades with POWER7® processors
- Power Systems servers and blades with POWER6/6+ processors
- System i servers with POWER6 processors
- System i servers with POWER5/5+ processors

When using external storage-based remote Copy Services together with IBM PowerHA SystemMirror for i, make sure to use separate Fibre Channel adapters for SYSBAS and for each IASP group. When using NPIV, you do not need separate physical adapters, but you still must use separate virtual adapters for SYSBAS and each IASP group.

For using the independent ASP, consider the disk arms of System ASP for good application performance. This is particularly important when it comes to temporary storage in a system configured with independent disk pools. All temporary storage is written to the SYSBAS disk pool. You must also remember that the operating system and basic functions occur in the SYSBAS disk pool. As a starting point use the guidelines provided in Table 8-2.

Table 8-2 Disk arms for SYSBAS guidelines

Disk arms in iASPs	Arms for SYSBAS: Divide iASP arms by:
Less than 20	3
20 - 40	4
Greater than 40	5

Note: Every situation is going to be different. The rule of thumb above is presented as a starting point. Your own requirements might vary.

Also, install the latest PTFs related with IBM PowerHA SystemMirror for i to all nodes in the cluster. The latest known PTFs fix known problems and provide the enhanced benefits in an IBM PowerHA environment. Periodically check and install the latest known PTF in “Recommended Fixes for High Availability for Release 7.1" at the following IBM i Support Recommended fixes website:

http://www-912.ibm.com/s_dir/slkbase.nsf/recommendedfixes

IBM PowerVM

IBM PowerVM is a virtualization technology for AIX®, IBM i, and Linux environments on IBM POWER processor-based systems. The PowerVM Virtual I/O Server included in the PowerVM provides a virtualization environment, in that the storage and network I/O resources are shared. This feature is required for IBM i to attach to DS8000 via VSCSI or NPIV and to SVC/V7000.

Virtualization technology is offered in three editions on Power Systems:

- PowerVM Express Edition
- PowerVM Standard Edition
- PowerVM Enterprise Edition

They provide logical partitioning (LPAR) technology by using either the Hardware Management Console (HMC) or the Integrated Virtualization Manager (IVM), Dynamic LPAR operations, Micro-Partitioning®, and Virtual I/O Server capabilities, and N_Port ID Virtualization (NPIV). You can choose the editions of IBM PowerVM depending on what features you want (Table 8-3).

Table 8-3 IBM PowerVM Editions

Features	Express	Standard	Enterprise
Maximum VMs	3/Server	1000/Server	1000/Server
Management	VMControl, IVM	VMControl, IVM, HMC	VMControl, IVM, HMC
Virtual I/O Server	✓	✓(Dual)	✓(Dual)
Suspend/resume		✓	✓
NPIV	✓	✓	✓

Features	Express	Standard	Enterprise
Shared Processor Pools		✓	✓
Shared Storage Pools		✓	✓
Thin Provisioning		✓	✓
Active Memory™ Sharing			✓
Live Partition Mobility			✓

8.1.4 Virtual I/O Server considerations

Virtual I/O Server (VIOS) is virtualization software that runs in a separate partition of your Power System. Its purpose is to provide virtual storage and networking resources to one or more client partitions. The IBM i client partition can also be hosted by VIOS.

The Virtual I/O Server owns the physical I/O resources, such as Ethernet and SCSI (SAS)/FC adapters. It virtualizes those resources for its client LPARs to share them remotely using the built-in hypervisor services. These client LPARs can be quickly created, typically owning only physical memory and shares of processors without any physical disks or physical Ethernet adapters.

Traditionally, IBM i has been supporting 520 bytes per sector storage. There are restrictions to directly attach a common storage system to IBM i for this reason. However, VIOS brings to he IBM i client partition the ability to use 512 bytes per storage sector by using the sector conversion of IBM PowerVM Hypervisor, which converts IBM i traditional 8 x 520 byte sectors into 9 x 512 byte sectors of a 4 KB memory page. Therefore, open storage volumes (or logical units, LUNs) are physically attached to VIOS through a FC or a Serial-attached SCSI (SAS) connection and then made available to IBM i (Figure 8-2).

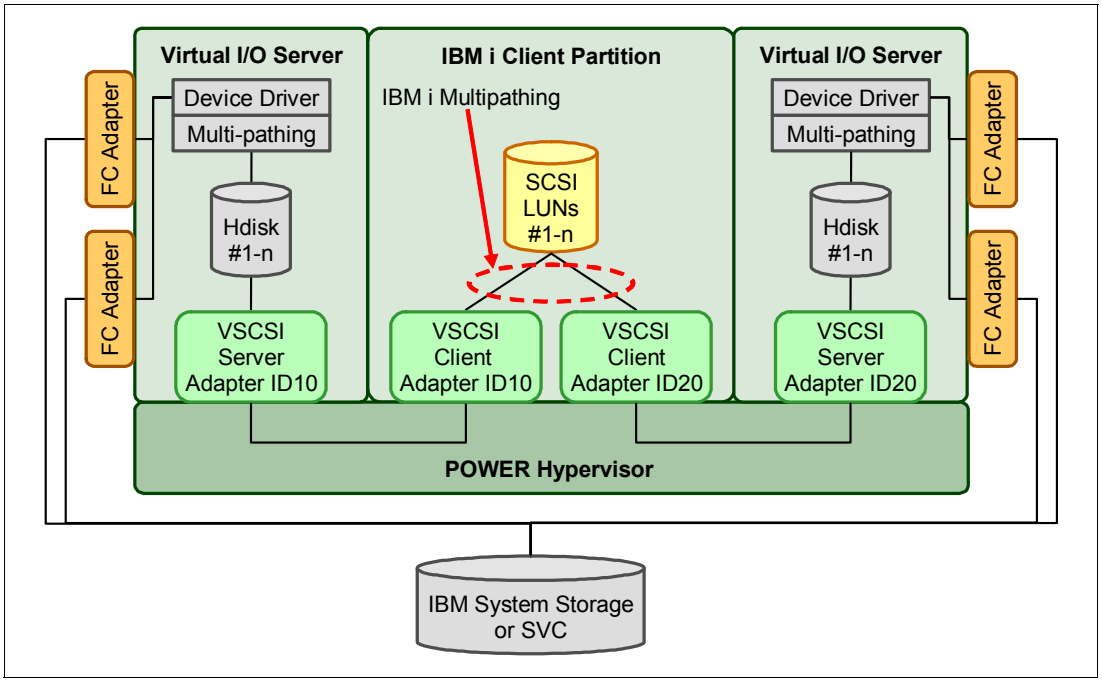


Figure 8-2 VSCSI connection via VIOS with multipathing

When you configure the IBM i client partition with VIOS, consider a dependency on VIOS. If the VIOS partition fails, IBM i on the client will lose contact with the virtualized open storage LUNs. The LUNs also become unavailable if VIOS is brought down for scheduled maintenance or a release upgrade. To remove this dependency, you can configure *Redundant VIOS* so that two or more VIOS partitions can be used to simultaneously provide virtual storage to one or more IBM i client partitions using IBM i multipathing (Figure 8-2 on page 137).

Virtual SCSI

From IBM i 6.1.1, IBM i VSCSI client driver supports MPIO through two or more VIOS partitions to a single set of LUNs. This multipath configuration allows a VIOS partition to fail or be brought down for service without IBM i loosing access to the disk volumes as the other VIOS partitions remain active (Figure 8-2 on page 137).

When setting up external storage with VIOS and VSCSI, there are several VIOS settings that should be changed. The first is the `fc_err_recov` and `dyntrk` attributes for each fscsi*X* device in VIOS partitions. To show the current fscsi*X* devices, use **`lspath`**.

Run the following command to set the fast fail and dynamic tracking attributes on these devices:

```
chdev–attr fc_err_recov=fast_fail,dyntrk=yes –perm –dev fscsiX
```

R to restart VIOS after these changes. The fast fail is designed for a multipath environment not to retry failed paths for a long time. In single path configuration, do not set fast_fail.

Note: In a dual VIOS environment using IBM i multipath, the SCSI reserve policy for each LUN (or hdisk) on both VIOS LPARs must be set to *no_reserve* to allow disk sharing.

This change must be made prior to mapping the LUNs to IBM i, and it does not require a restart of VIOS.

To show the current reserve policy settings, run the following command:

```
lsdev –dev hdiskX –attr reserve_policy
```

You can set the `no_reserve` attribute by running the following command:

```
chdev –dev hdiskX –attr reserve_policy=no_reserve
```

You need to install the *SVC Subsystem Device Driver Path Control Module* (SDDPCM) on VIOS to enable the management of multiple paths to the SAN Volume Controller or V7000 virtual Disks (VDisks). For more information about the Multipath Subsystem Device Driver, see the latest user's guides, available here:

`https://www-304.ibm.com/support/docview.wss?dc=DA400&rs=540&uid=ssg1S7000303&context=ST52G7&cs=utf-8&lang=en&loc=en_US`

N_Port ID Virtualization (NPIV) with IBM System Storage DS8000

N_Port ID Virtualization (NPIV) is an industry-standard FC protocol that allows VIOS to directly share a single FC adapter among multiple client LPARs, acting as a FC passthrough. Unlike VSCSI, NPIV does not map a LUN to a virtual target device in VIOS, which the client LPAR can then access as a generic SCSI disk. Instead, a port on the physical FC adapter is mapped to a virtual FC server adapter in VIOS, which in turn is connected to a virtual FC client adapter in IBM i (Figure 8-3 on page 139). When the virtual FC client adapter is created, two unique *world-wide port names (WWPNs)* are generated for it. Through the link to the server virtual FC adapter and then the physical adapter in VIOS, those WWPNs become

available on the SAN, and storage can be mapped to them as with any other FC host ports. Note that these WWPNs are unique, not just within the Power server, but globally on the SAN. When a virtual FC client adapter is deleted, the WWPNs are not reused. By default, the PowerVM Hypervisor is capable of creating 32,000 WWPNs for virtual FC client adapters. If additional WWPNs are required, clients can acquire an enablement code from IBM.

These are the requirements for IBM i DS8000 NPIV attachment:

- 8 Gb IOP-less Fibre Channel IOA (CCIN 577D) or 10 Gb FCoE PCI Express Dual Port Adapter (CCIN 2B3B)
- NPIV-capable SAN switch
- IBM i 6.1.1 or later
- HMC V7R3.5.0 or later
- POWER6 FW 350_038 or later
- VIOS 2.1 Fix Pack 22.1 or later

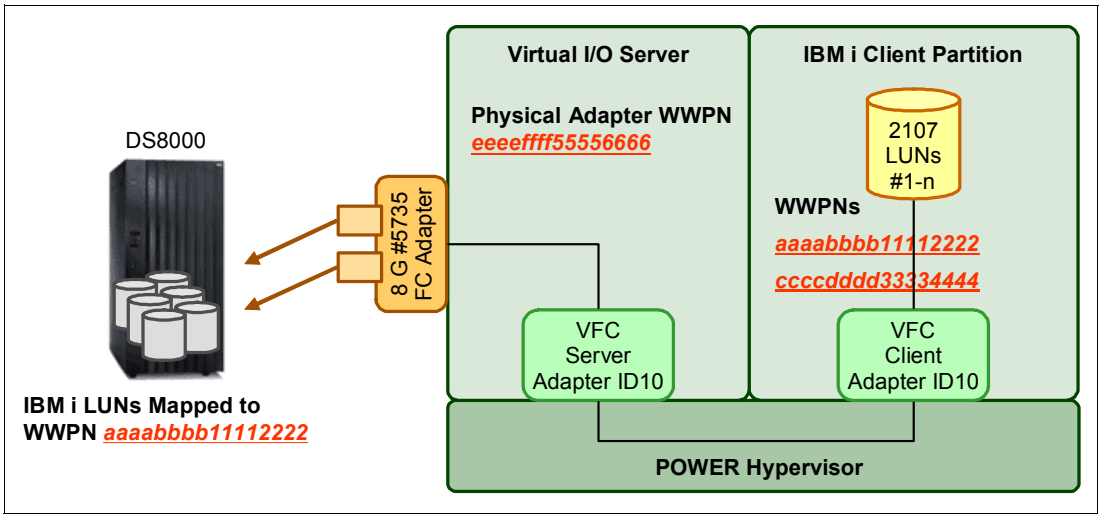


Figure 8-3 NPIV for IBM i

For more information about connecting IBM i through VIOS to DS8000 using VSCSI and NPIV, see *DS8000 Copy Services for IBM i with VIOS*, REDP-4584.

8.1.5 Storage considerations

You can check which of the external storages can attach directly to IBM i or serve for IBM i partitions via VIOS in 4.1, “PowerHA technologies” on page 46.

IBM Storage System DS6000/DS8000

For the IBM PowerHA SystemMirror for i to be able to communicate with the external Storage System you need to install the DS command-level interface (DS CLI) on all nodes in the cluster. The DS CLI software can be found and downloaded here:

http://www-947.ibm.com/support/entry/portal/Downloads/Hardware/System_Storage/Storage_software/Other_software_products/Copy_Services_CLI_%28Command_Line_Interface%29

For more information about using DS CLI with IBM i, see Chapter 8, "Using DS CLI with System i," in *IBM i and IBM System Storage: A Guide to Implementing External Disks on IBM i*, SF24-7120.

The support of DS8000 multiple Global Mirror sessions have been added in Version 6.1 of the DS8000 firmware. Prior to 6.1, only a single Global Mirror session was allowed to be active at a time on the DS8000. However, the Global Mirror session on the DS8000 is different from the PowerHA ASP session of type *GLOBALMIR. PowerHA does not support multiple Global Mirror sessions, as there are additional parameters that are required on the DSCLI commands. However, the legacy commands for Global Mirror from PowerHA work as long as the user only has a single Global Mirror session on the DS8000.

Fibre Channel-attached LUNs are identified as the storage unit device type of 2107 on the IBM i host system. You can specify 1 - 32 LUNs for each attachment to the IBM i Fibre Channel adapter feature 2766, 2787, or 5760. Also, you can specify 1 - 64 LUNs for each attachment to the IBM i Fibre Channel adapter feature 5749, 5774/5276, or 5735/5273.

Table 8-4 Capacity and Models of Disk Volumes for IBM i

Size	Type	Protected model	Unprotected model
8.5 GB	2107	A01	A81
17.5 GB	2107	A02	A82
35.1 GB	2107	A05	A85
70.5 GB	2107	A04	A84
141.1 GB	2107	A06	A86
282.2 GB	2107	A07	A87

SVC/V7000

Attaching SVC/V7000 to IBM i is possible using IBM PowerVM Virtual I/O Server (VIOS). IBM i runs in its own LPAR as a client of the VIOS server and accesses the SVC/V7000 storage via Virtual SCSI (vSCSI).

For this attachment, the minimum requirements for IBM PowerHA SystemMirror for i are SVC/V7000 Version 6.1.x and IBM VIOS 2.2.1. For a SVC-supported hardware list, device driver, firmware, and recommended software level, see the following web page:

https://www-304.ibm.com/support/docview.wss?uid=ssg1S1003697#_CodeLevel

To communicate between PowerHA and SVC/V7000 for issuing command-line interface (CLI), you must prepare for using the Secure Shell (SSH) TCP port 22. You can generate the SSH key pair from IBM i. For more information, see “Preparing for SSH connection between IBM i and SVC/V7000” on page 373.

Table 8-5 lists the configuration maximums for VIOS supporting an IBM i client attached to SVC.

Table 8-5 SVC configuration maximums for IBM i servers

Object	Maximum	Descriptions
Volume (HDisk)	512	The maximum number of volumes that can be supported by the SAN Volume Controller for a host running an IBM i operating system (per host object).
Paths per volume	8	The maximum number of paths to each volume. The suggested number of paths is 4.[a]

a. Subsystem device driver path-control module (SDDPCM) for AIX supports 16 paths per volume, but the SAN Volume Controller supports a maximum of only eight paths for a reasonable path-failover time.

There are also known issues and limitations with the SVC and IBM i host, so consider the following items when attaching SVC to a host that runs IBM i:

- A maximum of 16 disk virtual LUNs and 16 optical virtual LUNs is supported for each IBM i virtual I/O client SCSI adapter.
- SAN Volume Controller thin-provisioned volumes are supported for IBM i for use as FlashCopy targets only.

8.2 PowerHA Copy Services Support considerations

In this section we describe the considerations for using Copy Services support in various scenarios.

8.2.1 Global Mirror symmetrical and asymmetrical configurations

Global Mirror and IASPs offer a new and exciting opportunity for a highly available environment. They enables customers to replicate their environment over an extremely long distance without the use of traditional IBM i replication software. This environment comes in two types:

- Asymmetrical
- symmetrical

Asymmetrical configuration

With this configuration (Figure 8-4), Global Mirror can only be used from the production site to the disaster/recovery site. This type of configuration would be typical for a disaster recovery configuration where the production systems would run in the secondary location only if there was an unplanned outage of the primary location. Only one consistency group is set up, and it resides at the remote site. This means that you cannot do regular role swaps by reversing the Global Mirror direction (disaster recovery to production).

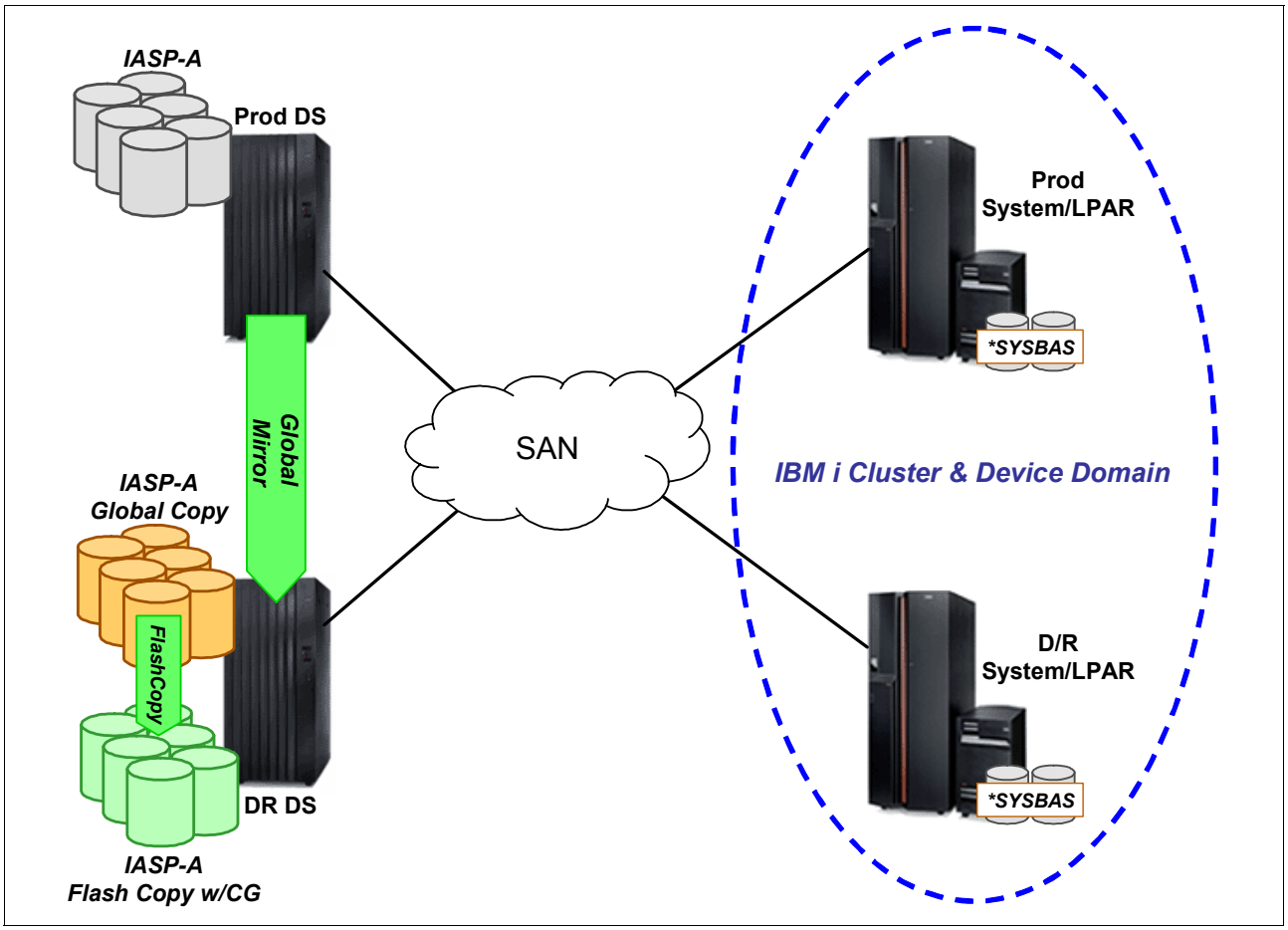


Figure 8-4 Global Mirror with asymmetrical configuration

After production workloads are moved to the recovery site, Global Copy must be used to return to the primary site. This is a manual procedure that is not supported by IBM PowerHA SystemMirror for i. For more detailed steps for Global Mirror switchback with an asymmetrical configuration, see in “Using CL commands for an asymmetrical Global Mirror switchback after a planned outage” on page 339. Because no disaster recovery capability would be provided in the reverse direction, it is unlikely that in this type of configuration we would choose to run for extended periods of time in the secondary location unless forced to by unavailability of the primary site.

Because Global Mirror uses two copies of data in the secondary location, there would be twice as many physical drives in this location as in the production location if the same size drives were used. In some situations, it might be cost effective to use larger drives in the secondary location. Spreading the production data over all these drives should provide equivalent performance in a disaster situation while reducing the overall cost of the solution.

Symmetrical configuration

With the configuration shown in Figure 8-5, an additional FlashCopy consistency group is created on the source site production DS model. It provides all the capabilities of asymmetrical replication, but adds the ability to do regular role swaps between the production and the disaster recovery sites using PowerHA.

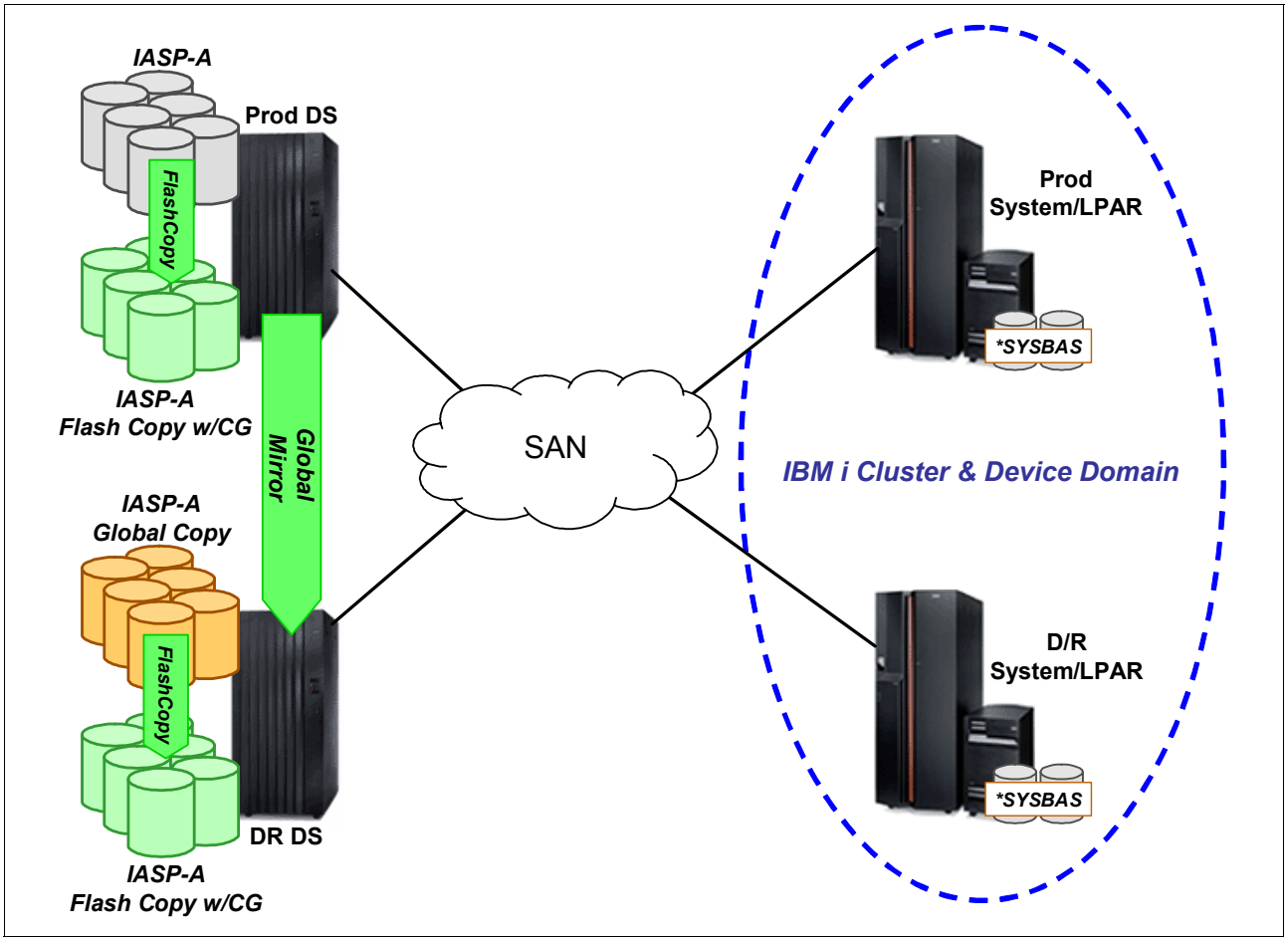


Figure 8-5 Global Mirror with symmetrical configuration

As we have FlashCopy capacity in both sites, it is possible to provide a disaster recovery solution using Global Mirror in both directions between the two sites. This type of configuration would typically be used where the production workloads might run for extended periods of time in either location.

8.2.2 FlashCopy NoCopy/full copy/incremental/reverse

In this section we discuss the various FlashCopy options.

Full copy versus NoCopy

There are two variants to the copy operation in FlashCopy. Whether you do a full copy or NoCopy depends on how you want to use the FlashCopy targets. In a real-world environment where you might want to use the target volume over a longer time and with high I/O workload, *full copy* is a better option to isolate your backup system I/O workload from your production workload when all data has been copied to the target. For the purpose of using FlashCopy for

creating a temporary system image for saving to tape during a low production workload, the *nocopy* option is recommended.

With the standard FlashCopy, *full copy* is the default, while the *nocopy* option is the default for FlashCopy SE.

Incremental and reverse FlashCopy

Incremental FlashCopy provides the capability to *refresh* a FlashCopy relationship, thus refreshing the target volume. When refreshing a target volume, any writes previously written to the target volume are always overwritten. The incremental FlashCopy is not available with FlashCopy SE.

To perform an incremental FlashCopy, you must first establish the FlashCopy relationship with the change data recording and persistent FlashCopy options enabled. You can do the incremental copy at any time, and you do not have to wait for the previous background copy to complete.

Usually you can use the incremental FlashCopy to minimize the amount of data that must be regularly copied and save the time for the physical copy of FlashCopy.

With *reverse FlashCopy*, the FlashCopy relationship can be reversed by copying over modified tracks from the target volume to the source volume. For DS6000 and DS8000, the background copy process must complete before you can reverse the order of the FlashCopy relationship to its original source and target relationship. The change data recording is a prerequisite for reverse restore.

The reverse restore function can only be used when a full copy relationship is completed. Therefore, it is not possible with FlashCopy SE. PowerHA supports reverse FlashCopy only to volumes that are not a part of a Metro Mirror or Global Mirror relationship.

It is possible to establish up to 12 FlashCopy relationships using the same source for DS8000. That is, a source volume can have up to 12 target volumes. However, a target volume can still only have one source. Also, a FlashCopy target volume cannot be a FlashCopy source volume at the same time.

For multi-relationship FlashCopy, an incremental FlashCopy relationship can be established with one and only one target. For each source volume, only one FlashCopy relationship can be reversed. With reversing one relationship of a DS8000 multi-relationship FlashCopy, all FlashCopy targets for which the background copy process has not completed are lost.

A persistent relationship is necessary to do an incremental FlashCopy or a reverse FlashCopy.

8.3 Sizing and performance considerations

In this section we discuss sizing and performance considerations on various environments.

8.3.1 Geographic mirroring

With geographic mirroring, IBM i does the replication. It is important to consider performance when planning to implement a geographic mirroring solution. While asynchronous geographic mirroring does allow a bit more flexibility regarding the distance between systems, there are still implications to undersizing the source, target, or the communications line between the two.

Minimizing the latency (that is, the time that the production system waits for the acknowledgement that the information has been received on the target system) is key to good application performance.

In the following sections we discuss how to size for good geographic mirroring performance.

General performance recommendations

When implementing geographic mirroring, different factors can influence the performance of systems involved in this HA solution. To maximize the performance of your applications that are used in this HA solution, several planning considerations must be taken into account. The factors discussed in this section provide general planning considerations for maximizing performance in a geographic mirroring environment.

There are two separate aspects to consider when sizing for a geographic mirroring environment. During the normal run time of the production environment, there will be some overhead added by geographic mirroring as the IBM i operating system is sending disk writes to the target system. The second aspect is the overhead and time required for synchronization, when the target IASP is reconnected to the source IASP and changes are pushed from the source to the target to make the two equivalent again.

Source and target comparison

Geographic mirroring consumes resources on both the source and the target resource. Especially for synchronous geographic mirroring, the best performance will be seen when the source and target systems are fairly equivalent in CPU, memory, and the disk subsystem.

CPU considerations

There is extra overhead on CPU and memory when doing geographic mirroring, and this has to be considered for both the source and the target system. Geographic mirroring increases the CPU load to the system processors on both the system owning the production copy of the IASP and the system owning the mirror copy of the IASP. There must be sufficient excess CPU capacity to handle this overhead, but there is no formula to calculate this exactly as it depends on many factors in the environment and the configuration. This CPU usage is needed for both systems to communicate and replicate data from the source IASP to the target IASP.

You might require additional processors to increase CPU capacity. As a general rule, the partitions that you are using to run geographic mirroring needs more than a partial processor. In a minimal CPU configuration, you can potentially see 5 - 20% CPU overhead while running geographic mirroring.

Regarding the backup system, be especially careful in sizing that system's processor. It should not be a small percentage of your production system, because this might slow down synchronization times considerably. If your backup system has fewer processors in comparison to your production system and there are many write operations, CPU overhead might be noticeable and affect performance.

Memory considerations

Geographic mirroring also requires extra memory in the machine pool. For optimal performance of geographic mirroring, particularly during synchronization, increase your machine pool size by at least the amount given by the following formula and then use **WRKSHRPOOL** to set the machine pool size:

```
Extra machine pool size = 300 MB + (0.3 * number of disk arms in the IASP)
```

This extra memory is needed particularly during the synchronization process on the system that owns the mirror copy of the IASP. However, you must add extra storage on every cluster

node involved in geographic mirroring (as defined in the cluster resource group). Any node in the cluster can become the primary owner of the mirror copy of the IASP if a switchover or failover occurs.

Important: The machine pool storage size must be large enough before starting the resynchronization. Otherwise, increasing memory is not taken into account as soon as the synchronization task is in progress, and the synchronization process can take longer.

If the system value QPFRADJ is equal to 2 or 3**,** then the system might make changes to the storage pools automatically as needed. To prevent the performance adjuster function from reducing the machine pool size take these steps:

1. Set the machine pool minimum size to the calculated amount (the current size plus the extra size for geographic mirroring from the formula) by using the Work with Shared Storage Pools (**WRKSHRPOOL**) command or the Change Shared Storage Pool (**CHGSHRPOOL**) command.
2. Set the Automatically adjust memory pools and activity levels (QPFRADJ) system value to zero, which prohibits the performance adjuster from changing the size of the machine pool.

Note: We recommend that you use **WRKSHRPOOL** for setting the machine pool size to the calculated minimum. Disabling Performance Auto Adjuster can have other performance implications in your environment.

Disk subsystem considerations

Disk unit and IOA performance can affect overall geographic mirroring performance. This is especially true when the disk subsystem is slower on the mirrored system. When geographic mirroring is in synchronous mode, all write operations on the production copy are gated by the mirrored copy writes to disk. Therefore, a slow target disk subsystem can affect the source-side performance. You can minimize this effect on performance by running geographic mirroring in asynchronous mode. Running in asynchronous mode alleviates the wait for the disk subsystem on the target side and sends confirmation back to the source side when the changed memory page is in memory on the target side.

System disk pool considerations

Similar to any system disk configuration, the number of disk units available to the application can have a significant affect on its performance. Putting additional workload on a limited number of disk units might result in longer disk waits and ultimately longer response times to the application. This is particularly important when it comes to temporary storage in a system configured with independent disk pools. All temporary storage (such as objects in the QTEMP library) is written to the SYSBAS disk pool. If your application does not use a lot of temporary storage, then you can get by with fewer disk arms in the SYSBAS disk pool.

Although disk writes associated with production data will be on the IASP, there will continue to be disk activity associated with the SYSBAS pool for the operating system and basic functions. As a starting point, you can use the guidelines shown in Table 8-6.

Table 8-6 Disk arms for SYSBASE guidelines

Disk arms in IASPs	Arms for SYSBAS: Divide IASP arms by:
Less than 20	3
20 - 40	4
Greater than 40	5

For example, if IASP contains 10 drives, then SYSBAS should have at least three. As another example, if IASP contains 50 drives, then SYSBAS should have at least 10.

Note: Disk pool sizing is very application dependent, and the above guidelines are only given as a starting point. You might find that fewer or more arms are required for your application environment. Understanding performance monitoring for your application environment is critical for sizing and capacity planning.

You will want to monitor the percent busy of the SYSBAS disk arms in your environment to ensure that you have the appropriate number of arms. If it gets above 40% utilization, then you must add more arms. Also, when possible, the disk assigned to the IASP should be placed on a separate I/O adapter from the SYSBAS disk to reduce any potential contention. It has also been found that IOA cache is very important and provides greater data integrity and improved performance.

Communications lines

When you are implementing a PowerHA solution using geographic mirroring, plan for adequate communication bandwidth so that the communications bandwidth does not become a performance bottleneck in addition to system resources.

Geographic mirroring can be used for virtually any distance. However, only you can determine the latency that is acceptable for your application. The type of networking equipment, the quality of service, the distance between nodes, the number, and the characteristics of data ports used can all affect the communications latency. As a result, these become additional factors that can impact geographic mirroring performance.

To ensure better performance and availability, we recommend that you take the following actions:

- To provide consistent response time, geographic mirroring should have its own redundant communications lines. Without dedicated communication lines, there might be contention with other services or applications that utilize the same communication line. Geographic mirroring supports up to four communication lines (data port lines), and a cluster heartbeat can be configured for up to two lines. However, we recommend utilizing Virtual IP addresses to provide redundancy to the cluster heartbeat.
- It is important to know that a round-robin approach is used to send the data across the lines. This implies that for best performance, when multiple dataport lines are configured, they should have close to equivalent performance characteristics. If one slow line is added, then this will gate the sending of the data to that line speed.
- Geographic mirroring replication should also be run on a separate line from the cluster heartbeating line (the line associated with each node in the cluster). If the same line is used, during periods of heavy geographic mirroring traffic, heartbeating could fail, causing

a false partition. For the cluster heartbeat we recommend using two different IP addresses for redundancy.

- From a high availability point of view, we recommend using different interfaces and routers connected to different network subnets for the four data ports that can be defined for geographic mirroring (Figure 8-6). It is better to install the Ethernet adapters in different expansion towers, using different System i hardware buses. Also, if you use multiport IOA adapters, use different ports to connect the routers.
- Virtual IP adapter (VIPA) can be used to define the geographic mirroring IP addresses.

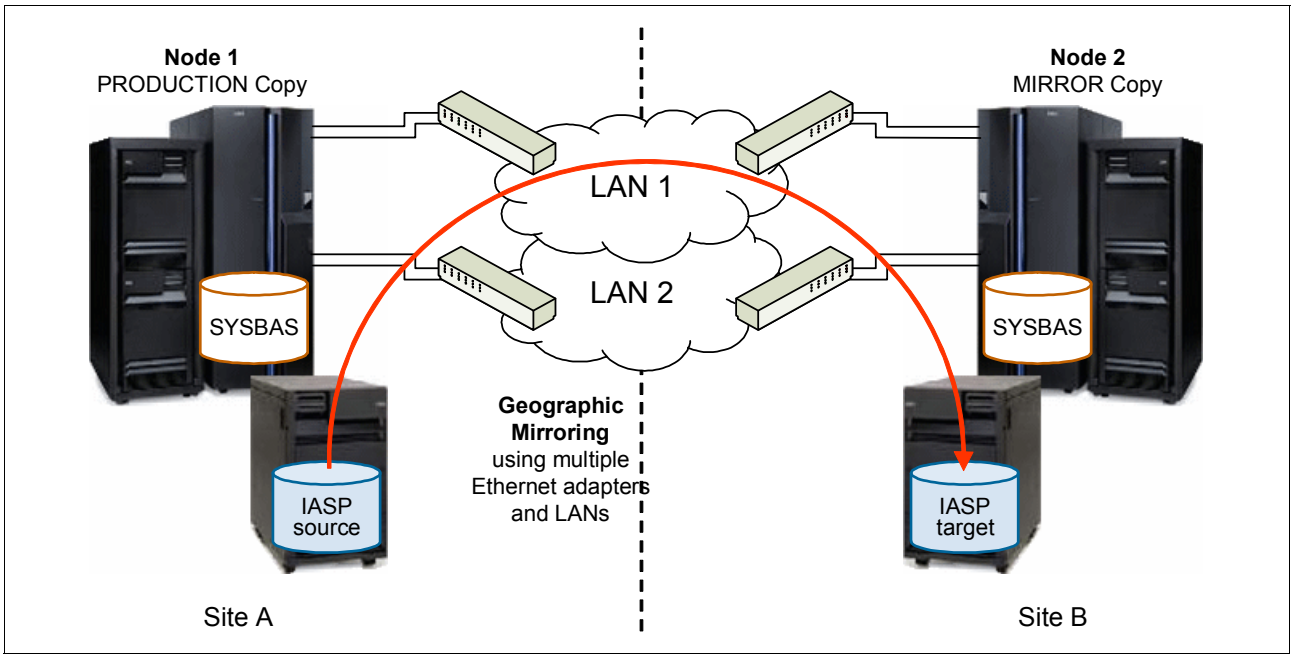


Figure 8-6 Recommended network configuration for geographic mirroring

Runtime environment

In this section we discuss the runtime environment.

Delivery and mode

When configuring geographic mirroring, there are two main parameters that affect geographic mirroring runtime performance. The DELIVERY parameter affects the performance of disk writes to the IASP. With synchronous delivery, the disk write will not complete until the affected page in storage has also been received on the target system. Asynchronous delivery allows the disk write on the source to complete after the write has been cached. The actual sending of the disk write to the target system happens outside the scope of the write on the source. For synchronous delivery, there is also a synchronous or asynchronous MODE. Synchronous mode ensures that the write has arrived at the disk cache on the target (essentially on disk at that point) before returning. Asynchronous mode only ensures that the write is on memory on the target.

Synchronous delivery and synchronous mode guarantee equivalent copies of the IASP on the source and the target while geographic mirroring is active. It also provides the added protection of a *crash-consistent* copy of the data in case of a target system failure, because all writes will have been received into the disk subsystem.

Synchronous delivery and asynchronous mode can be beneficial for customers running with a significantly slower disk subsystem on their target system. This allows the disk write on the

source to complete without waiting for the completion on the target. This delivery and mode still guarantee equivalent data on the source and target IASPs in the case of a failure of the source system.

With synchronous delivery, it is important to have the communications bandwidth available to support the number of disk writes at all peak periods throughout the day or night. The overhead of sending the data to the target will be added to the time for each disk write to complete, which can significantly affect production performance. Even with a very fast line, if the distance between the source and the target is too great, production performance will suffer. For this reason, asynchronous delivery for geographic mirroring was introduced in release 7.1.

Asynchronous delivery is best for those environments in which the source and target are separated by too long of a distance for acceptable synchronous response times, or for scenarios where the bandwidth cannot support the peak write rate.

Sizing for optimum performance

For the best runtime performance, it is important to know the write volume within the IASP. We only consider writes because those are the only I/O that is transferred to the target system. If the IASP has not yet been defined, the write volume in SYSBAS can be used as a rough estimate, understanding that this might result in excess communications capacity. Both the peak and average megabytes per second written should be collected, preferably over short intervals, such as 5 minutes.

For synchronous delivery, the bandwidth of the communications lines must be able to keep up with the peak write volume. If it cannot keep up, the writes will begin to stack up and production performance will suffer.

For asynchronous delivery, the bandwidth of the lines must still keep up at least to the average write volume. Because writes on the source are not waiting, it is acceptable for some queuing to occur, but if the line cannot handle the average write volume, then geographic mirroring will continue to get further and further behind.

It also is important to examine the variance of the write rate over time. If there is a large variance between peak and average, then it might be advisable to size more for the peak. Undersizing in this case affects the recovery point objective in the case of a source system failure during the peak write rate.

Communications transports speeds

Just how fast is a T1 line? A data T1 transfers information at about 1.544 megabits every second, or about 60 times more than the average conventional dialup modem. That translates to .193 MBps theoretical throughput for a T1 line. The absolute best that you can hope to get out of a T1 line is 70% effective throughput, and most network specialists say to plan for 30%. Therefore, the best that a T1 line can transfer is .135 MBps. If you have a 2 Gigabyte file to initially sync up, then that synch would take over 2 hours with nothing else running. As you scale to 2 TB, that same sync would take over 80 days. As you can see, most systems need more than a T1 line to achieve effective geographic mirroring transaction throughput.

T3 lines are a common aggregation of 28 T1 circuits that yield 44.736 Mbps total network bandwidth or 5.5 MBps with a best effective throughput of 70%, which equals 3.9 MBps and a planning number of 2 MBps.

The OC (the optical carrier fiber optic-based broadband network) speeds help you grow.

Table 8-7 provides other communication line speeds.

Table 8-7 Communication line speeds

Type	Raw speed (MBps)	Raw speed (MBps)	30% planning (MBps)	GB/hour during synch
T1	1.544	0.193	0.06	0.22
DS3/T3	44.736	5.5	2	7.2
OC-1	51.840	6.5	2.1	7.6
OC-3	155.52	19.44	6	21.6
OC-9	455.56	56.94	18	64.8
OC-12	622.08	77.76	24	86.4
OC-18	933.12	116.64	35	126
OC-24	1244	155.5	47	169
OC-36	1866	233.25	70	252
OC-48	2488	311	93	335
OC-192	9953	1244.12	373	1342
1 Gb Ethernet local	1000	125	38 (30% local)	225

Monitoring the runtime environment

When using asynchronous delivery, it might be useful to determine whether geographic mirroring is “keeping up” with disk writes. On **DSPASPSSN** on the source system, the Total data in transit field gives the amount of data in megabytes that has been sent to the target system, but not acknowledged as received. This field is only shown when the transmission delivery is *ASYNCH and the state is ACTIVE.

Synchronization

In this section we discuss synchronization.

Partial and full synchronizations

When you suspend mirroring for any planned activities or maintenance, any changes made on the production copy of the independent disk pool are not being transmitted to the mirror copy. So, when you resume geographic mirroring, synchronization is required between the production and mirror copies.

If geographic mirroring is suspended without tracking, then full synchronization occurs. This can be a lengthy process. If geographic mirroring is suspended with the tracking option, PowerHA will track changes up to the tracking space limit specified on the ASP session. When mirroring is resumed, the production and mirror copies are synchronized concurrently with performing geographic mirroring.

Tracking is available on both the source side and the target side. Target side tracking greatly reduces the need for a full synchronization. Usually a full synchronization is only required when either the source or target IASP does not vary off normally, such as from a crash or an abnormal vary-off.

While a synchronization is taking place, the environment is not highly available. This makes it essential to calculate the time required to do a full synchronization to understand whether the business can support that length of time exposed to an outage.

Tracking space

Tracking space is a reserved area within the IASP where the system tracks changed pages while in suspended status, and the changes need to be synchronized when resuming mirroring. Tracking space is needed only when the target copy of the IASP is suspended, detached, or resuming. The changes themselves are not contained within the tracking space, only a space-efficient indication of which pages require changes. The amount of tracking space allocated can be defined by the user. The maximum is 1% of the total space within the IASP. Using **CHGASPSSN**, a user can set the percentage of that 1%. For example, setting the field to 10% means that the tracking space would be 10% of 1% or .1% of the total IASP size. These parameters can be viewed using **DSPASPSSN**. The tracking space allocated is the percentage of the maximum (it would show 10% in the above example) and the tracking space used is the percentage of the available tracking space being used.

Note: If tracking space is exhausted (it reaches 100%), then no more changes can be tracked, and when you resume geographic mirroring, a full synchronization is required.

Monitoring synchronization

To track how much data is left to be synchronized, **DSPASPSSN** can be used on the source system. On the second screen, there are fields for Total data out of synch and Percent complete. These fields will display the megabytes of data that need to be resynchronized and how far the synchronization has progressed. Both of these fields are updated as the synchronization runs. Each time the a synchronization starts or is resumed, these fields will be reset. In the case of a resume, the percent complete will reset to 0, but you should also see a reduced total data out of synch.

Calculating full synchronization time

It is fairly simple to determine the bandwidth needed for day-to-day operations and initial synchronization. All write I/O will need to be sent from the primary system to the secondary system.

To determine the megabytes of writes per second for each interval, run the performance tools during a representative and peak period.

From the resulting `QADMDSK` file use these parameters:

- DSBLKW number of blocks written: A block is one sector on the disk unit. PD (11,0).
- INTSEC elapsed interval seconds: The number of seconds since the last sample interval. PD (7,0).

Then you take these steps:

1. Calculate disk blocks written per second:

   ```
   Disk blocks written per interval divided by the number of seconds in the
   interval ((QAPMDISK.QAPMDISK.DSBLKW / QAPMDISK.QAPMDISK.INTSEC)
   ```

2. Convert disk blocks to bytes. Multiply by 520 to get the number of bytes.
3. Divide by a million to get megabytes per second.
4. Divide by 2 to get geographic mirror traffic because the disk writes are doubled if the system is using mirrored disk.

The formula to calculate the amount of traffic expressed as megabytes written per second is as follows:

```
((QAPMDISK.QAPMDISK.DSBLKW / QAPMDISK.QAPMDISK.INTSEC)  * 520) / 1000000 / 2
```

For example, if you determine that the amount of traffic is 5 MBps and you want to use geographic mirroring for disaster recovery, then you need a pipe that can accommodate 5 MBps of data being transferred. If you are configuring two lines as data ports, then you need 2.5 MBps per line.

From Table 8-7 on page 150, we can see the following facts:

- A DS3/T3 allows 5.6 MBps theoretical throughput with a 2 MBps with a best practice at 30% utilization.
- An OC-3 line allows 19.44 MBps theoretical throughput with 6 MBps with a best practice at 30% utilization.

You can initially start with two DS3 lines, but you might need to go to two OC-3 lines to account and plan for growth.

To determine the time needed for initial synchronization, divide the total space utilized in the IASP by the effective communications capability of the chosen communications lines. Speed of the lines makes a big difference.

For example, if the IASP size is 900 GB and you are using 1 Gb Ethernet switches, then the initial synchronization time will be less than an hour. However, if you are using two T3/DS3 lines, each having an effective throughput of 7.2 GB/hour, it would take around 63 hours to do the initial synchronization. This was calculated by dividing the size of the IASP by the effective GB/hour, that is, 900 GB divided by 14 GBps. A full resynchronization might also be needed in the event of a disaster, so that must be factored into disaster recovery plans.

In most cases, the size of the data is used in the calculation, not the size of the IASP. An exception to this is a *NWSSTG in an IASP. An *NWSSTG object is treated as one file, so the size of the *NWSSTG is used instead of the amount of data within the `*NWSSTG` file.

To compute the initial synchronization time for *NWSSTG in an IASP, divide the size of the network storage space of the IBM i hosted partition by the effective speed of the communications mechanism.

For example, if the network storage space hosting IBM i was set up as 600 GB, it would take 42 hours to do the initial synchronization for a disaster recovery scenario using two DS3 lines.

To improve the synchronization time, a compression device can be used.

Synchronization priority

The synchronization priority setting (low, medium, or high) determines the amount of resources allocated to synchronization. Lower settings will gate synchronization, which will also allow more resources to be allocated to non-synchronized work.

Managing contention between run time and synchronization

Ideally, synchronization will run best when the system is quiet. However, most businesses cannot support this amount of quiesced time. Thus, synchronization will most likely be contending for system resources with normal production workload, in addition to the normal geographic mirroring runtime workload.

For the least effect on production work, a synchronization priority of low can be selected. However, this lengthens the amount of time required to complete the synchronization, also lengthening the amount of time without a viable target.

For additional information about best practices for high-availability environments, see Chapter 15, “Best practices” on page 397.

8.3.2 Virtual I/O Server

When you configure VIOS for IBM i client serving, dedicated VIOS processors will work better than shared processors. Virtual SCSI devices generally have higher processor utilization when compared with directly attached storage. The amount of processor required for Virtual SCSI Server is based on the maximum I/O rates required of it.

The sizing methodology used is based on the observation that the processor time required to perform I/O operating on the virtual SCSI server is fairly constant for a given I/O size. Table 8-8 provides the approximate cycles per second for both physical disk and logical volume operations on a 1.65 Ghz processor. For other frequencies, scaling by the ratio of the frequencies is sufficiently accurate to produce a reasonable sizing.

Table 8-8 Approximate cycle per second on a 1.65Ghz logical partition

Disk type	4 KB	8 KB	32 KB	64 KB	128 KB
Physical disk	45,000	47,000	58,000	81,000	120,000
Logical volume	49,000	51,000	59,000	74,000	105,000

For example, consider a Virtual I/O Server that uses two client logical partitions on physical disk-backed storage. The first client logical partition requires a maximum of 7,000 8-KB operations per second. The second client logical partition requires a maximum of 10,000 8-KB operations per second. The number of 1.65 Ghz processors for this requirement is approximately ((7,000 × 47,000 + 10,000 × 47,000) / 1,650,000,000) = 0.48 processors, which rounds up to a single processor when using a dedicated processor logical partition.

As a sizing general rule, a single dedicated POWER6 processor for VIOS is good enough for about 40,000 virtual SCSI I/O.

For the memory sizing of VIOS, we do not need to consider the additional requirement for VSCSI as similar to the processor because there is no data caching required by VSCSI. With large I/O configurations and very high data rates, a 1 GB memory allocation for the VIOS is likely to be sufficient. For low I/O rate situations with a small number of attached disks, 512 MB will most likely suffice.

The Virtual I/O Server requires a minimum of 30 GB of disk space with the Virtual I/O Server Version 2.2.0.11, Fix Pack 24, Service Pack 1, or later.

On the VIOS host the mapped logical drives are seen as hdisks and then assigned to an IBM i partition. We recommend that you assign entire hdisks to IBM i. VIOS supports logical volumes where you can subset an hdisk into multiple volumes to assign to a partition. This is not recommended for IBM i partitions.

IBM i has always been architected to perform best with more disk arms. This does not change with SAN disks. You need to create a good number of logical drives, not one large drive.

We strongly recommend that only FC or SAS physical drives are used to create LUNs for IBM i as a client of VIOS because of the performance and reliability requirements of IBM i production workloads.

As the LUNs are virtualized by VIOS, they do not have to match IBM i integrated disk sizes. The technical minimum for any disk unit in IBM i is 160 MB and the maximum is 2 TB, as measured in VIOS. Actual LUN size is based on the capacity and performance requirements of each IBM i virtual client partition and load source disk restrictions (17.5 GB minimum, 1.9 TB maximum). IBM i performs best with logical drives that are the same size.

When creating an open storage LUN configuration for IBM i as a client of VIOS, it is crucial to plan for both capacity and performance. As LUNs are virtualized for IBM i by VIOS instead of being directly connected, it might seem that the virtualization layer will necessarily add a significant performance overhead. However, internal IBM performance tests clearly show that the VIOS layer adds a negligible amount of overhead to each I/O operation. Instead, the tests demonstrate that when IBM i uses open storage LUNs virtualized by VIOS, performance is almost entirely determined by the physical and logical configuration of the storage subsystem.

It is possible to virtualize up to 16 LUNs to IBM i through a single VSCSI connection. Each LUN typically uses multiple physical disk arms in the open storage subsystem. If more than 16 LUNs are required in an IBM i client partition, an additional pair of VSCSI server (VIOS) and client (IBM i) adapters must be created.

8.3.3 Copy Service bandwidth

The SAN infrastructure between the local and remote storage system plays a critical role in how much data and how fast remote Copy Services is able to replicate. If the SAN bandwidth is too small to handle the traffic, then application write I/O response times will be longer.

Consider that the bandwidth can handle *peak write* workload requirements.

For IBM Storage System, collect the performance data to get the highest write rate and calculate the needed bandwidth as follows:

1. Assume 10 bits per byte for network overhead.
2. If the compression of devices for remote links is known, you can apply it.
3. Assume a maximum of 80% utilization of the network.
4. Apply a 10% uplift factor to the result to account for peaks in the 5-minute intervals of collecting data, and a 20 - 25% uplift factor for 15-minute intervals.

As an example, we show how to calculate the required bandwidth for a given write workload:

1. The highest reported write rate at an IBM i is 40 MBps.
2. Assume 10 bits per byte for network overhead:

   ```
   40 MBps * 1.25 = 50 MBps
   ```

3. Assume a maximum of 80% utilization of the network:

   ```
   50 MBps * 1.25 = 62.5 MBps
   ```

4. Apply a 10% uplift for 5-minute intervals:

   ```
   62.5 MBps * 1.1 = app 69 MBps
   ```

5. The needed bandwidth is 69 MBps.

SVC Global Mirror is much more sensitive to a lack of bandwidth than DS8000. Also, it is important to design and size a pipe that is able to handle normal and peak write I/O. You can calculate the link bandwidth for SVC Global Mirror with the peak write load for all servers, additional background copy rate, and the SVC intercluster heartbeat traffic size. The background copy rate is usually 10 - 20% of the maximum peak load. The intercluster heartbeat traffic size for the primary cluster and the secondary cluster can be found as a table in the SVC information center. As a example for the bandwidth calculation, see the following steps:

1. The peak write load for all servers is 10 MBps.
2. Add 10 - 20% for background copy rate:

   ```
   10 MBps * 1.15 = 11.5 MBps
   ```

3. Add the SVC intercluster heartbeat traffic. If there are two nodes in cluster 1 and four nodes in cluster 2, the size is 4 Mbps:

   ```
   11.5 MBps + 0.5 MBps = 12 MBps
   ```

   The needed bandwidth is 12 MBps.

A Recovery Point Object (RPO) estimation tool is available for IBM and IBM Business Partners. This tool provides a method for estimating the RPO in a DS8000 Global Mirror environment in relation to the bandwidth available and other environmental factors (see Techdocs Document ID: PRS3246 for IBM and IBM Business Partners). For more information, contact to IBM or IBM Business Partners.

8.3.4 FlashCopy space-efficient relation

With a FlashCopy *space-efficient* or *thin-provisioned* relation, disk space will only be consumed for the target when a write to source needs to be hardened on disk or when a write is directed to the target. For this reason, using FlashCopy SE requires less disk capacity than using standard FlashCopy, which can help lower the amount of physical storage needed.

FlashCopy SE is designed for temporary copies, so FlashCopy SE is optimized for use cases where a small percentage of the source volume is updated during the life of the relationship. If much more than 20% of the source is expected to change, there might be a trade-off in terms of performance versus space efficiency. Also, the copy duration should generally not last longer than 24 hours unless the source and target volumes have little write activity.

DS8000 FlashCopy SE

The DS8000 FlashCopy SE repository is an object within an extent pool and provides the physical disk capacity that is reserved for space-efficient target volumes. When provisioning a repository, storage pool striping will automatically be used with a multi-rank extent pool to balance the load across the available disks. FlashCopy SE is optimized to work with repository extent pools consisting of four RAID arrays. In general, we recommend that the repository extent pool contain between one and eight RAID arrays. Extent pools larger than eight RAID arrays are not recommended for the repository. It is also important that adequate disk resources are configured to avoid creating a performance bottleneck. It is advisable to use the same disk rotational speed or faster (10 K RPM or 15 K RPM) for the target repository as for the source volumes. We also recommend that the repository extent pool have as many disk drives as the source volumes.

After the repository is defined in the extent pool it cannot be expanded, so planning is important to ensure that it is configured to be large enough. If the repository becomes full, the FlashCopy SE relationships will fail. After the relationship fails, the target becomes unavailable for reads or writes, but the source volume continues to be available for reads and

writes. You can estimate the physical space needed for a repository by using historical performance data for the source volumes along with knowledge of the duration of the FlashCopy SE relationship. In general, each write to a source volume consumes one track of space on the repository (57 KB for CKD, 64 KB for FB). Thus, the following calculation can be used to come up with a reasonable size estimate:

```
IO Rate x (% Writes/100) x ((100 – Rewrite%)/100) x Track Size x Duration in
seconds x ((100+Contingency%)) = Repository Capacity Estimate in KB
```

Because it is critical not to undersize the repository, a contingency factor of up to 50% is suggested.

You can monitor and notify repository capacity and threshold using Simple Network Management Protocol (SNMP) traps. You can set notification for any percentage of free repository space with a default notification at 15% free and 0% free. Also, you can convert and send these messages to QSYSOPR messages using the ACS toolkit. For more detailed information, see in 10.2, “DS storage management” on page 178.

SVC/V7000 thin provisioning

For SVC/V7000, when you are using a fully allocated source with a thin-provisioned target, you need to disable the background copy and cleaning mode on the FlashCopy map by setting both the background copy rate and cleaning rate to zero. If these features are enabled, then the thin-provisioned volume will be either offline or as large as the source. You can select the grain size (32 KB, 64 KB, 128 KB, or 256 KB) for thin-provisioning. The grain size that you select affects the maximum virtual capacity for the thin-provisioned volume. If you select 32 KB for the grain size, the volume size cannot exceed 260,000 GB. The grain size cannot be changed after the thin-provisioned volume has been created. In general, smaller grain sizes save space and larger grain sizes produce better performance. For best performance, the grain size of the thin-provisioned volume must match the grain size of the FlashCopy mapping. However, if the grain sizes are different, the mapping still proceeds.

You can set the cache mode to readwrite for maximum performance when you create a thin-provisioned volume. Also, to prevent a thin-provisioned volume from using up capacity and getting offline, the autoexpand feature can be turned on.

9

PowerHA user interfaces

There are several interfaces available for you to set up and manage your HA environment, including the brand new PowerHA GUI, which effectively combines both the Cluster Resource Services GUI and the High Availability Solution Manager GUI.

In this chapter we introduce you to the new PowerHA GUI and discuss the other interfaces:

9.1 Command line

Although there are multiple GUIs available to manage your HA environment, you might still want to use the traditional 5250 interface for performing certain tasks either from a CL program or a command line.

Note: The current release of IBM PowerHA for i GUI only supports the command-line interface for working with SVC/V7000 copy services.

The IBM PowerHA for i licensed program provides IBM i command-line interfaces to configure and manage your high-availability solution.

These are the categories of various IBM PowerHA for i commands:

- Cluster administrative domain commands
- Monitored resource entry commands
- Cluster commands
- Commands and APIs for working with ASP copy descriptions and sessions
- Geographic mirroring and iASP configuration commands

9.1.1 The Work with Cluster (WRKCLU) command

The Work with Cluster (**WRKCLU**) command is a good starting point because it gives you one-stop access to most of the functions available within the HA environment. When you run this command, the Work with Cluster Menu displays (Figure 9-1).

```
                              Work with Cluster
                                                          System:   DEMOGEO1
Cluster  . . . . . . . . . . . . . . . . :   PWRHA_CLU

Select one of the following:

     1. Display cluster information
     2. Display cluster configuration information

     6. Work with cluster nodes
     7. Work with device domains
     8. Work with administrative domains
     9. Work with cluster resource groups
    10. Work with ASP copy descriptions

    20. Dump cluster trace



Selection or command
===>

F1=Help   F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel
```

Figure 9-1 Work with Cluster

The Work with Cluster page has menu options to display cluster-level information, such as version and summary, to view and manage configuration and tuning parameters, to work with the nodes, to work with the device and administrative domains, to work with CRGs and ASP copy descriptions, and also an option to trace and gather debug information.

9.1.2 The Configure Device ASP command

The new Configure Device ASP (**CFGDEVASP**) command is part of the base operating system and is available with PTF SI44141. The Configure Device ASP command can be used to create or delete an independent auxiliary storage pool (ASP).

When used with the *CREATE action, this command does as follows:

- Creates the independent ASP using the specified non-configured disk units
- Creates an ASP device description by the same name if one does not already exist

When used with the *DELETE action, this command does as follows:

- Deletes the independent ASP
- Deletes the ASP device description if it was created by this command

See 2.2, “Creating an IASP” on page 18, for more information about creating an independent ASP.

```
                      Configure Device ASP (CFGDEVASP)

Type choices, press Enter.

ASP device . . . . . > PWRHA_ASP1    Name
Action . . . . . > *CREATE       *CREATE, *DELETE
ASP type . . . . . .   *PRIMARY      *PRIMARY, *SECONDARY, *UDFS
Primary ASP device . . . . . .                Name
Protection . . . . . .   *NO           *NO, *YES
Encryption . . . . . .   *NO           *NO, *YES
Disk units . . . . . .   *SELECT       Name, *SELECT
               + for more values

                           Additional Parameters

Confirm  . . . . . .   *YES          *YES, *NO




                                                                        Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 9-2 Configure Device ASP (CFGDEVASP) command

Note: The Configure Device ASP (**CFGDEVASP**) command does not rely on SST/DST access. However, the user profile must still have *IOSYSCFG and *SERVICE special authorities.

9.1.3 Configure Geographic Mirroring command

The Configure Geographic Mirror (**CFGGEOMIR**) command (Figure 9-3) can be used to create a geographic mirror copy of an existing independent ASP. This command can also create ASP copy descriptions if they do not already exist and can start an ASP session. It performs all the configuration steps necessary to take an existing standalone independent ASP and create a geographic mirror copy.

```
                    Configure Geographic Mirror (CFGGEOMIR)

Type choices, press Enter.

ASP device . . . . . . . . . . . > PWRHA_ASP1    Name
Action . . . . . . . . . . . . . > *CREATE       *CREATE, *DELETE
Source site  . . . . . . . . . .   *             Name, *
Target site  . . . . . . . . . .   *             Name, *
Session  . . . . . . . . . . . .   PWRHA_SSN1    Name, *NONE
  Source ASP copy description  .     PWRHA_CPY1  Name
  Target ASP copy description  .     PWRHA_CPY2  Name
Transmission delivery  . . . . .   *ASYNC        *SYNC, *ASYNC
Disk units . . . . . . . . . . .   *SELECT       Name, *SELECT
               + for more values

                           Additional Parameters

Confirm  . . . . . . . . . . . .   *YES          *YES, *NO
Cluster  . . . . . . . . . . . .   *             Name, *
Cluster resource group . . . . .   *             Name, *
                                                                    More...
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 9-3 Configure Geographic Mirror (CFGGEOMIR) command

9.2 PowerHA GUI

The new PowerHA GUI combines key features from the existing Cluster Resource Services and High Availability Solutions Manager GUIs and is accessed via web-based IBM Systems Director Navigator for i. PowerHA GUI can be ordered via 5799 HAS PRPQ and is available for customers running 7.1. See 8.1.2, “PRPQ ordering information” on page 134, for more information.

This new GUI makes configuration and management of a high availability environment easy to use and intuitively obvious. The overall status of the high availability environment can be easily determined and supports the ability to drill down to detect and correct specific problems in the environment. The GUI enables basic functionality and allows configuration and management of a geographic mirroring or DS8000-based replication environment.

Note: We recommend that PowerHA GUI is used for managing your high-availability environment going forward, as the older Cluster Resource Services and High Availability Solutions Manager GUIs will be discontinued after all of their functionality is brought into the PowerHA GUI.

9.2.1 Accessing the PowerHA GUI

Using the new PowerHA GUI to create and mange your HA solution is straightforward and intuitive, as it combines both the solution-based HASM GUI features and the task-based CRS GUI features.

The PowerHA GUI is accessed via the IBM Systems Director Navigator for i from this website:

`http://<system_name>:2001`

You are then redirected automatically to port 2005, and your browser might issue a security certificate warning. Choose **Continue to this website** and enter your user ID and password to log in to IBM Systems Director Navigator for i. Expand the IBM i Management tasks tree to see the new PowerHA option (Figure 9-4).

Figure 9-4 IBM Systems Director Navigator for i: PowerHA option

When PowerHA is accessed for the first time in the current session, it will look for any existing configuration and inspect the HA environment (Figure 9-5).

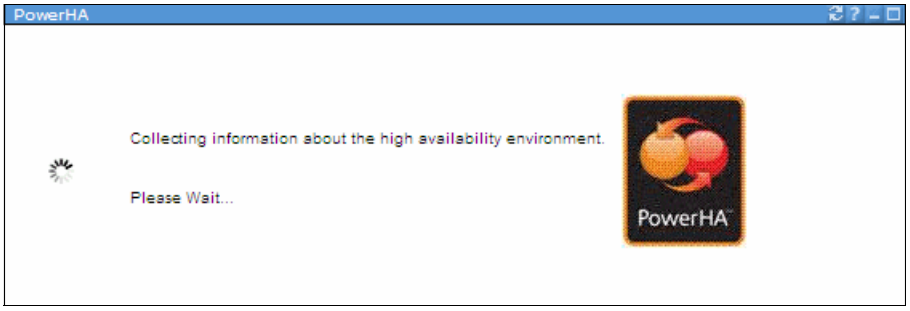


Figure 9-5 PowerHA: Collecting information about the HA environment

See Chapter 11, “Creating a PowerHA base environment” on page 199, for the steps for creating a PowerHA base environment.

9.2.2 Managing the cluster

All of the cluster management functions that are available via options 1 and 2 on **WRKCLU** are also available on the PowerHA GUI. The pull-down menu next to the cluster name allows you

to display and modify cluster properties or to check whether the requirements for the HA environment are satisfied.

Cluster properties

The cluster properties option of PowerHA GUI (Figure 9-6) gives you similar access as the functions on the **CHGCLU** and **CHGCLUVER** command interfaces.

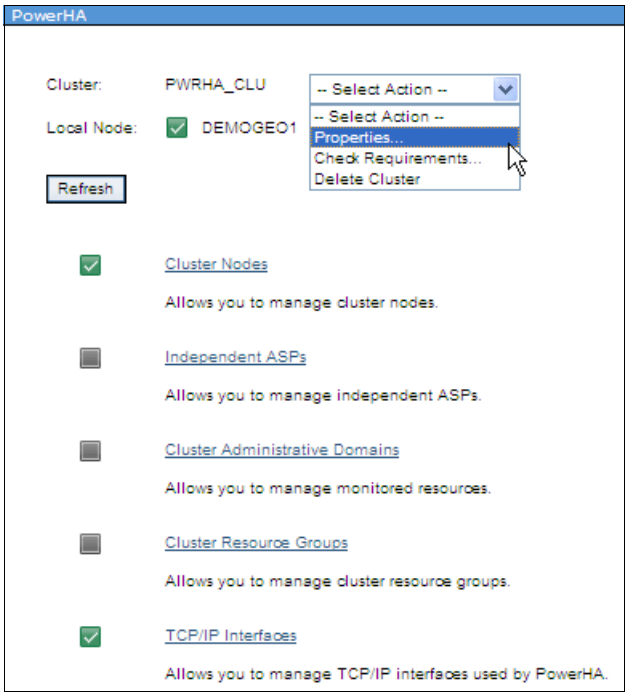


Figure 9-6 Cluster pull-down menu

Figure 9-7 shows the cluster properties of our demo environment.

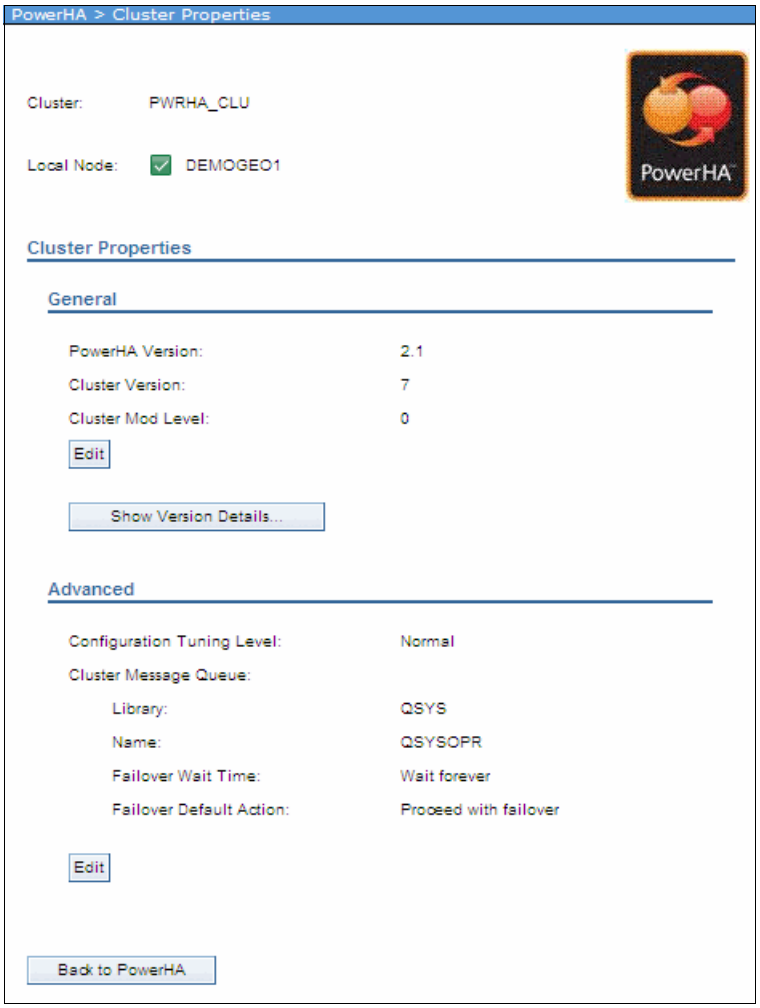


Figure 9-7 Cluster Properties

Check Requirements option

The check requirements option of the cluster level pull-down menu runs through various checks and highlights any condition that might prevent a failover execution.

Note: The Current PowerHA version must be 2.1 for most of the options on PowerHA GUI to work, including the Check Requirements function.

In addition to showing the warnings and suggestions, PowerHA GUI also gives you an option to fix the condition without having to leave the page (Figure 9-8).

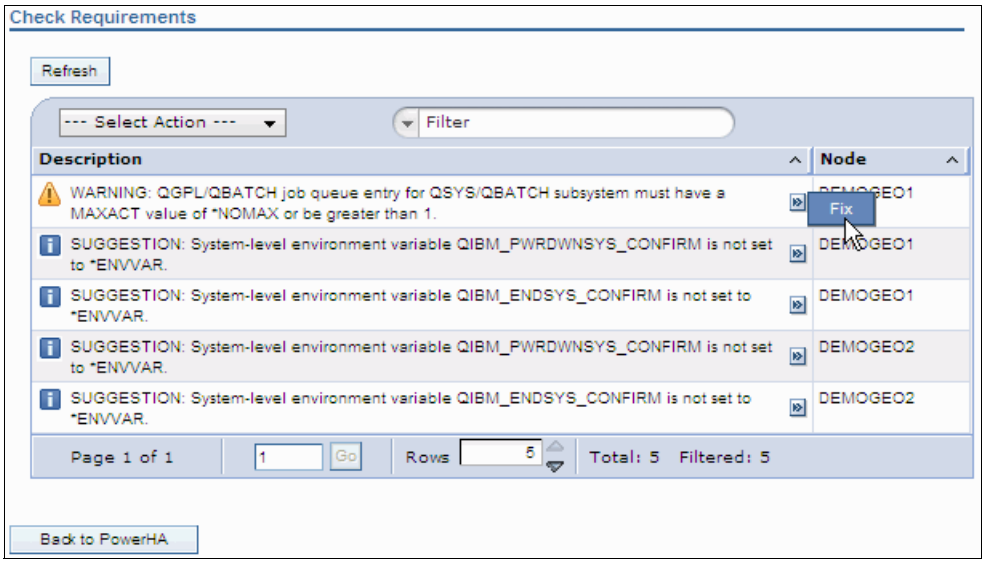


Figure 9-8 Check Requirements: Fix option

Note: Choosing the fix option from the PowerHA warning message panel modifies the QGPL/QBATCH job queue entry for the QSYS/QBATCH subsystem from a shipped default value of 1 to *NOMAX.

Delete Cluster option

When you choose the Delete Cluster menu option, a warning message box (Figure 9-9) displays. If you choose Yes, PowerHA deletes a cluster from all nodes currently in the cluster's membership list.

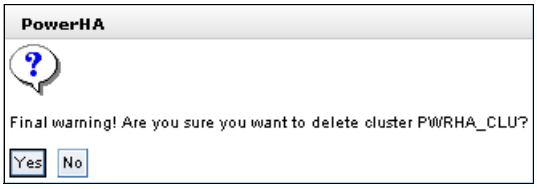


Figure 9-9 Delete cluster warning

9.3 Cluster Resource Services GUI

The Cluster Resource Services graphical user interface (GUI) is provided by IBM Systems Director Navigator for i. This GUI provides you with a task-based approach for setting up and maintaining a high-availability environment that uses PowerHA for i.

The Cluster Resource Services GUI interface (Figure 9-10) allows you to configure and manage clustered resources and environments. Unlike the HASM GUI discussed in 9.4, “High Availability Solution Manager GUI” on page 165, this cluster interface is based on task-oriented goals. This interface allows you to take the following actions:

- Create and manage a cluster.
- Create and manage cluster nodes.
- Create and manage cluster resource groups (CRGs).
- Create and manage cluster administrative domains.
- Create and manage Monitored Resource Entries (MREs).
- Monitor the cluster status for high-availability-related events such as failover, cluster partitions, and so on.
- Perform manual switchovers for planned outages, such as backups and scheduled maintenance of software or hardware.

For more information about Cluster Resource Services GUI, see Chapter 7 of the IBM Redbooks publication *Implementing PowerHA for IBM i*, SG24-7405:

http://www.redbooks.ibm.com/abstracts/sg247405.html

Figure 9-10 Cluster Resource Services GUI

9.4 High Availability Solution Manager GUI

The PowerHA for i High Availability Solutions Manager graphical interface provides a solution-based approach to selecting, configuring, and managing your HA environment.

The High Availability Solutions Manager graphical interface (Figure 9-11 on page 166) provides several predefined solutions. For each of these solutions, the dependent technologies are configured based on your selection. The High Availability Solutions Manager graphical interface provides easy-to-use tools to manage your high-availability solution.

Unlike the CRS GUI discussed in 9.3, “Cluster Resource Services GUI” on page 164, HASM GUI employs a solution-based approach.The GUI operations are deployed via IBM Systems Director Navigator for i, a web-based console that allows you to take the following actions:

- Select a high-availability solution.
- Verify requirements for your high-availability solutions.
- Set up a high availability solution.
- Manage a high availability solution.

For more information about configuring and managing your high-availability solution using the HASM GUI, see Chapter 6 of *Implementing PowerHA for IBM i*, SG24-7405:

http://www.redbooks.ibm.com/abstracts/sg247405.html

Note: High Availability Solution Manager is an all-or-nothing type of tool. That is, to be able to manage a solution, it must have been set up with the HASM GUI. You cannot manage your solution if it was set up outside of this tool.

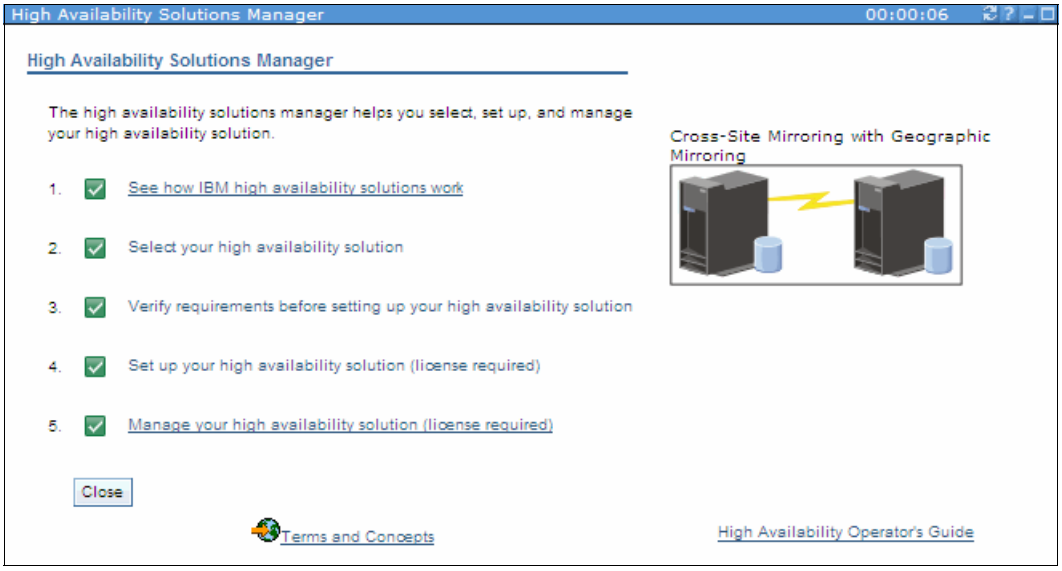


Figure 9-11 High Availability Solutions Manager GUI

10

Advanced Copy Services for PowerHA

Advanced Copy Services (ACS) for PowerHA is an extension to IBM PowerHA SystemMirror for i that provides enhanced functionality and automation through greater integration and management of an IBM System Storage DS6000/DS8000 Copy Services environment.

In this chapter we describe the following major functional enhancements provided by ACS:

- Easy-to-use interface for managing a DS storage and Copy Services environment including externalized DSCLI commands, additional sanity checks, and scripting, which can be used also for non-IBM i workloads (for example, WRKCSE and STRDSMGT)
- Support for customized programming and faster resolution for Copy Services-related issues by using default scripts provided and maintained for the Copy Services environment
- Full automation of the FlashCopy process from the backup node
- Additional safety measures and checks to prevent users from accessing inconsistent IASP data after a FlashCopy relationship ended
- Verification tests for PPRC switch-readiness, which can be user-initiated (CHKPPRC)
- Ability to view and manage DS storage host connections and associated volumes
- Simplified creation of new host connection when replacing a defective Fibre Channel IOA
- Detailed audit trail for debugging issues (VIEWLOG)
- Delivery of DS storage SNMP messages to the IBM i QSYSOPR message queue
- Automated creation of a D volume FlashCopy at a Global Mirror target site
- Support for Metro/Global Mirror (MGM) 3-site disaster recovery solutions
- Support for integration with TPC-R in Metro Mirror or Metro/Global Mirror environments

ACS is implemented as a service offering from IBM STG Lab Services and is supported via the regular IBM i software support structure. For further information about ACS see the IBM STG Lab Services website for Power Systems here:

http://www-03.ibm.com/systems/services/labservices/platforms/labservices_power.html

10.1 Advanced Copy Services interface

The implementation of Advanced Copy Services for PowerHA on top of IBM PowerHA SystemMirror for i requires IBM i clustering to be set up using regular PowerHA commands, with the exception of the cluster resource group for which ACS requires a data CRG to be created for a switchable IASP using **CRTCSECRG**.

ACS requires the user QLPAR on each IBM i cluster node and a user qlpar on the DS6000/DS8000 systems. Instead of the user having to specify a user ID and password for any of the ACS functions, it uses a DSCLI password file, which is set up only once from the ACS storage management interface, as like described in 10.2, “DS storage management” on page 178.

The main entry point to the ACS user interface for Copy Services is the Work with Copy Services Environments accessed by **WRKCSE** (Figure 10-1).

```
                          Copy Services Environments

Type options, press Enter.
  1=Add   2=Change       4=Delete   5=Display         12=Work with
  14=List Stream files   16=Define host connections   18=Make PPRC Paths

Opt   Name         Type    Text

      DEMOGM       GMIR
      FLASH01      FLASH
      FLASH02      FLASH
      TOOLKIT      FLASH
      TOOLKIT      MMIR









                                                                      Bottom
Command
===>
F1=Help   F3=Exit   F4=Prompt   F9=Viewlog   F12=Cancel
```

Figure 10-1 ACS Copy Services Environments panel (WRKCSE)

As the available options from Work with Copy Services Environments shows, the user can create a new Copy Services Environment or change, delete, display, or manage an existing Copy Services Environment for Metro Mirror, Global Mirror, FlashCopy, or LUN-level switching.

When displaying a Copy Services environment with ACS using option 5=Display from the WRKCSSE panel (Figure 10-2), all relevant information is presented on one panel for both source and target relationships (in our case, for the nodes DEMOPROD and DEMOHA). ACS does not use ASP sessions like PowerHA, but under the cover still creates ASP copy descriptions.

```
                          Display a PPRC Environment
 Press Enter to continue.

 Environment  . . . . . :   TOOLKIT
 Type . . . . . :   MMIR
 ASP Device name  . . . . :   TOOLKIT
 Source Copy Description  :   TKMMIRPS
 Target Copy Description  :   TKMMIRPT
 TPC Replication  . . . . :   *NO

 Source node  . . . . . :   DEMOPROD
 Target node  . . . . . :   DEMOHA
 Primary ASP  . . . . . :   222
 Source device  . . . . . :   IBM.2107-75AY031
 Target device  . . . . . :   IBM.2107-75AY032

 Source hmc1  . . . . . :   9.5.168.55
 Target hmc1  . . . . . :   9.5.168.55
 Volume sets  . . . . . :   6
 PPRC Paths . . . . . :   2

                                                                       Bottom
Volume relationships:
  DEMOPROD     DEMOHA
  Volumes      Volumes
  7000-7002    7000-7002
  7100-7102    7100-7102

                                                                       Bottom
 F1=Help   F3=Exit   F8=PPRC Paths   F12=Cancel
```

Figure 10-2 ACS displaying a PPRC environment panel

The ACS has built-in checks to prevent the user from erroneously updating copy descriptions by favoring the remote replication copy descriptions over FlashCopy descriptions and stores all Copy Services environment data in the device domain data to make sure that it is kept consistent within the IBM i cluster. It also provides protection against a user accidentally manually changing the ASP copy descriptions instead of changing them via the ACS Copy Services environment interface by refusing any ASP copy description changes with the following message:

```
Copy Descriptions and ACS data do not match.
```

Using option 12 (Work with Volumes) from the Work with Copy Services environment panel (Figure 10-3) helps the user to manage the DS Copy Services configuration by working with volumes, setting directions for replication after a failover, suspending/resuming PPRC, or viewing out-of-sync tracks.

```
                          Work with MMIR Environment

  Environment  . :   TOOLKIT              Status . . . . :   Running
  Direction  . . :   Normal

Select one of the following:

     2. Pause
     3. Resume

     6. Start Replication after failover

    12. Work with Volumes
    13. Display Out of Sync sectors
    14. List Stream files


Selection


F1=Help   F3=Exit   F5=Refresh Status   F9=Viewlog   F12=Cancel
```

Figure 10-3 ACS Work with MMIR Environment

Using option 12 (Work with Volumes) for a Metro Mirror environment, the user can view the PPRC state and suspend or resume Metro Mirror volume relationships (Figure 10-4).

```
                          Work with MMIR PPRC Volumes

  Environment  . :   TOOLKIT              Direction  . . :   Normal
  Copy Service                           Source device  :   IBM.2107-75AY031
    Type . . . . :   MMIR                Target device  :   IBM.2107-75AY032

Type Volume options; 2=Pause, 3=Resume, press Enter.

  Opt  Src : Tgt  Preferred Source Status        Preferred Target Status
       7000:7000  Full Duplex - Metro Mirror 70  Target Full Duplex - Metro
       7001:7001  Full Duplex - Metro Mirror 70  Target Full Duplex - Metro
       7002:7002  Full Duplex - Metro Mirror 70  Target Full Duplex - Metro
       7100:7100  Full Duplex - Metro Mirror 71  Target Full Duplex - Metro
       7101:7101  Full Duplex - Metro Mirror 71  Target Full Duplex - Metro
       7102:7102  Full Duplex - Metro Mirror 71  Target Full Duplex - Metro


                                                                      Bottom
F1=Help   F3=Exit   F5=Refresh Status   F9=Viewlog   F12=Cancel
```

Figure 10-4 ACS Work with MMIR PPRC Volumes panel

Using option 14 (List Stream files) shows all the stream files for DSCLI profiles and scripts that get automatically created by ACS for a Copy Services environment with the preferred source DS storage system designated as PS and the preferred target designated as PT. These stream files are not supposed to be changed by the user, as ACS changes them any time that the user invokes an option of ACS, but they can be used by the user to easily view information about the DS8000 Copy Services configuration or used for custom programming.

```
                          CS Environment Stream Files
Type options; 2=Change, 4=Delete, 5=Display, 9=Run, press Enter.

Opt   Stream file name                          IFS directory
      pprc_PS.profile                           ...profiles/TOOLKIT_MMIR
      pprc_PT.profile                           ...profiles/TOOLKIT_MMIR
      pprc_9_PS.profile                         ...profiles/TOOLKIT_MMIR
      pprc_9_PT.profile                         ...profiles/TOOLKIT_MMIR
      runds_PS.profile                          ...profiles/TOOLKIT_MMIR
      runds_PT.profile                          ...profiles/TOOLKIT_MMIR
      chhostconn_Add.script                     ...scripts/TOOLKIT_MMIR
      chhostconn_Drop.script                    ...scripts/TOOLKIT_MMIR
      failoverpprc_to_PS.result                 ...scripts/TOOLKIT_MMIR
      failoverpprc_to_PS.script                 ...scripts/TOOLKIT_MMIR
      failoverpprc_to_PT.script                 ...scripts/TOOLKIT_MMIR
      lsavailpprcport_PS.result                 ...scripts/TOOLKIT_MMIR
      lsavailpprcport_PS.script                 ...scripts/TOOLKIT_MMIR
      lsfbvol_PS.result                         ...scripts/TOOLKIT_MMIR
      lsfbvol_PS.script                         ...scripts/TOOLKIT_MMIR
                                                                  More...
Command
===>
F1=Help   F3=Exit   F4=Prompt   F9=Viewlog   F12=Cancel
```

Figure 10-5 ACS Copy Services Environment Stream Files panel

However, while working within option 14 (List Stream files), ACS does not change the stream files, so the user can still make adjustments before invoking a stream file as a DSCLI *script* using option 9 (Run) with its output shown on the panel. This directory also has the `*.result` files from any ACS command function, like CHKPPRC or SWPPRC, which is where a user would find DSCLI errors if the command fails. The user can also run one of the existing stream file *profiles* with option 9 (Run) to quickly get an interactive DSCLI session with the corresponding DS storage system.

Using option 18 (Make PPRC Paths) from the Work with Copy Services Environments (WRKSCE) panel determines the available PPRC paths between the DS storage systems and lets the user create the PPRC paths in both directions for all logical subsystems (LSSs) in the configured environment (Figure 10-6).

```
                              Available PPRC Paths

  Environment  . :   TOOLKIT              Source device  :   IBM.2107-75AY031
  Type . . . . . :   MMIR                 Target device  :   IBM.2107-75AY032

Select all connection pairs to be used, press Enter.
These selections replace all paths currently in use for this environment.
  1=Select

Opt     PPRC Connection Path
          I0040  :  I0240
          I0140  :  I0340

                                                                        Bottom
F1=Help   F3=Exit   F5=Refresh Status   F9=Viewlog   F12=Cancel
```

Figure 10-6 ACS Available PPRC Paths panel

To easily set up Metro Mirror remote replication on the DS storage system with ACS, the user can simply use a combination of the WRKCSE option 18 (Make PPRC Paths) and run the corresponding mkpprc_from_PS.script that ACS does automatically generates based on the DS storage system configuration information that the user has already provided when creating the Copy Services environment.

For setting up Global Mirror remote replication on the DS storage system with ACS, the user can simply invoke the mkpprc_GM_from_PS.script, the mksession scripts for both sides, the chsession_GM_add_PS.script to add the primary volumes to the Global Mirror session, the mkflash_GM_CG_PT.script to create the FlashCopy consistency group C volumes, and finally the mkgmir_PS.script without ever needing to bother with manually entering complex DSCLI commands.

When setting up a Global Mirror environment, the user has the option (Figure 10-7) to specify a FlashCopy D volume on the Global Mirror target site for backup or testing purposes and also to set up an asymmetrical Global Mirror environment (that is, without FlashCopy consistency group C volumes on the preferred primary site). As with native PowerHA, also for ACS the failback with an asymmetrical Global Mirror setup is a manual step, but ACS will tell the user which scripts to run when invoking the failback option.

```
                          Display a PPRC Environment
Press Enter to continue.

Symmetric  . . . . . :   *YES
D-Copy Flash normal  . . :   *YES
D-Copy Flash reversed  . :   *NO
Extra CG Flash . . . . . :   *NO
Override Master LSS  . . :   *YES
Source Mstr LSS  . . . . :   02
                                                                      More...
Volume relationships:
 DEMOPROD     DEMOHA        DEMOHA         DEMOPROD
 PPRC Vols    PPRC Vols   CG Flash Vols  CG Flash Vols
 0200-0201    0200-0201     0250-0251      0250-0251
 0300-0301    0300-0301     0350-0351      0350-0351


                                                                       Bottom
F1=Help   F3=Exit   F8=PPRC Paths   F12=Cancel
```

Figure 10-7 ACS Display a PPRC Environment panel for Global Mirror

The user invokes FlashCopy for the D volumes on the Global Mirror target site either using the Work with Copy Services Environments option 12 (Work with) for the Global Mirror environment (Figure 10-8) or using **STRFLASH** with *GMIRTARGET (Figure 10-9 on page 175) to create a consistent D volume involving a Global Mirror failover and FlashCopy fast reverse restore from C to B and FlashCopy from B to D. The ACS Copy Services environment name for the D copy needs to match its Global Mirror environment name.

```
                          Work with GMIR Environment

  Environment  . :   DEMOGM               GMIR Status  . :   Running
  Direction  . . :   Normal               PPRC Status  . :   Running

Select one of the following:

     2. Pause
     3. Resume
     4. Failover
     5. Symmetrical switchover

     7. Make target D-Copy

    12. Work with Volumes
    13. Display Out of Sync sectors
    14. List Stream files
                                                                        Bottom
Selection


F1=Help   F3=Exit   F5=Refresh Status   F9=Viewlog   F12=Cancel
```

Figure 10-8 ACS Work with GMIR Environment panel

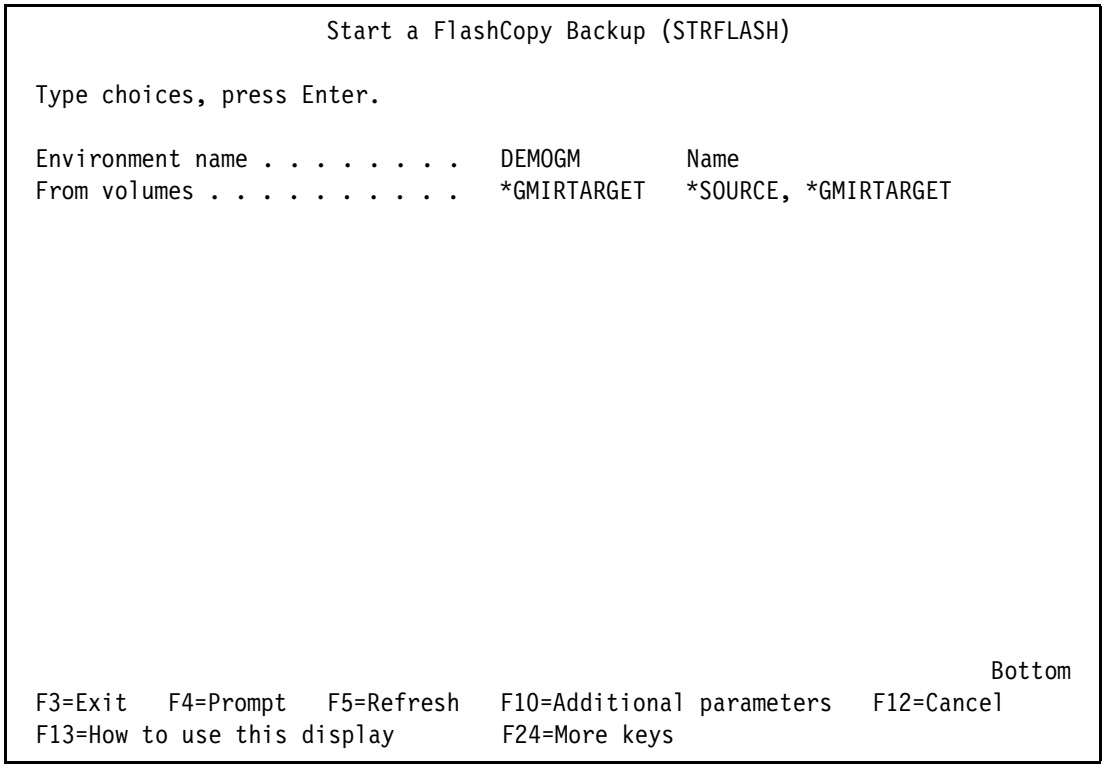


Figure 10-9 ACS Start a FlashCopy Backup panel (STRFLASH)

Creating a data CRG for ACS is done using **CRTCSECRG**, which also allows you to set up full FlashCopy automation, which automatically runs the CHGASPACT to quiesce/resume the IASP before and after taking the FlashCopy, as shown with the Quiesced flash *YES setting in the CHGCSEDTA output shown in Figure 10-10.

```
                              Change CSE CRG Data

Supply all required values, press Enter.

FlashCopy information:
  FlashCopy node . . . . . .   DEMOFC         Name
  Status . . . . . . . . . .   *FLASHED       *NONE, *FLASHED, number
  Warm flash . . . . . . . .   *YES           *YES, *NO
  Incremental flash  . . . .   *NO            *YES, *NO
  Quiesced flash . . . . . .   *YES           *YES, *NO

Second FlashCopy information:
  FlashCopy node . . . . . .                  Name

Third FlashCopy information:
  FlashCopy node . . . . . .                  Name









                                                                        More...
F1=Help   F3=Exit   F12=Cancel
```

Figure 10-10 ACS Change CSE CRG Data panel (CHGCSEDTA)

At any time the user can verify the PPRC switch-readiness of an ACS Copy Services environment by running **CHKPPRC**, which communicates with the DS storage system, similar to displaying an ASP session in PowerHA, to verify that PPRC is running full-duplex and is set up correctly with regard to the ACS PPRC setup.

A switch of PPRC is by design generally not automated by ACS. Rather, it can be explicitly switched using **SWPPRC** (Figure 10-11).

```
                              Switch PPRC (SWPPRC)

Type choices, press Enter.

Environment name . . . . . > TOOLKIT       Name
Switch type  . . . . . > *UNSCHEDULED  *SCHEDULED, *UNSCHEDULED...
Type . . . . . .   *             *, *GMIR, *LUN, *MMIR
Auto Vary On . . . . . .   *YES          *YES, *NO
Auto replicate . . . . . .   *DFT          *DFT, *YES, *NO
Switch paused MMIR . . . . . .   *NO           *YES, *NO

                                                                      Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 10-11 ACS Switch PPRC command panel (SWPPRC)

When using the Auto replicate *DFT option, for a *scheduled* Metro Mirror switch, the command looks at the CSE configuration parameter Automatic PPRC Replicate (whose default value is *YES) to determine whether the PPRC direction will be established in the reverse direction. For an *unscheduled* Metro Mirror switch, Auto replicate *DFT means *NO. That is, no replication in the reverse direction is started. For Global Mirror it always changes the replication direction for a *scheduled* switch, and also for an *unscheduled* switch if the source DS storage system is available. ACS allows the user to also switch a paused PPRC relationship, but only if all PPRC volumes are in the same paused state.

ACS is installed in the QZRDHASM library with its logs written to `/QIBM/Qzrdhasm/qzrdhasm.log`, which can easily be viewed using **VIEWLOG**.

DMPINF allows the user to dump all ACS Copy Services environment information, such as DSPCSEDTA and CSE scripts joblogs, to a stream file like `/QIBM/Qzrdhasm/qzrdhasm_DEMOPROD_110919_1202.txt`.

10.2 DS storage management

ACS provides an integrated DS storage management interface accessed via **STRDSMGT**, which after configuring a DS storage system connection the user can use to easily manage DS host connections, check volumes, Copy Services, or multi-path configurations; configure SNMP trap notifications; or manage DS storage system authentications (Figure 10-12).

```
                          Storage Management Menu
                                                          System:   CTCIHA8Y
 Storage Device  . . . . . . . . . . . . . . . . . . . . :   IBM.2107-75AY031

 SNMP Trap status  . . . . . . . . . . . . . . . . . . . :   Not active

 Select one of the following:

      1. Configure Storage Management
      2. Work with Storage Connections
      3. Save Storage Connections
      4. Delete Saved Connections
      5. Display Saved Connections
      6. Compare Saved Connections

      8. Check Storage Volumes
      9. Check Copy Service Volumes

     11. Check System i Multipath

     14. List Storage Management files

     17. Start Storage SNMP Traps
     18. Stop Storage SNMP Traps

     20. Manage Authority to Storage

Selection


 F3=Exit   F9=Viewlog   F12=Cancel
```

Figure 10-12 ACS Storage Management Menu

After setting up the DS connection information with option 1 (Configure Storage Management) (Figure 10-13), option 20 (Manage Authority to Storage) allows the user to add any user to a DS storage system (option 1), change the qlpar user (option 2) on the DS storage system, or automatically update the DSCLI password file on the IBM i client (option 4) (Figure 10-14 on page 180).

```
                        Configure Storage Management

Press Enter to continue

Storage:
  Device name  . . . . . .   IBM.2107-75AY031     Name
  Primary HMC  . . . . . .   9.5.168.55           IPv4
  Second HMC . . . . . . .                        IPv4

SNMP Trap:
  Message Queue Name . . . . .   *SYSOPR          name, *SYSOPR
  Message Queue Library  . . .                    name
  Issue storage commands . . .   *NO              *YES, *NO

Verbose message logging  . . .   *NO              *YES, *NO

Copy Services installed  . . :   *YES




F1=Help   F3=Exit   F12=Cancel
```

Figure 10-13 ACS Configure Storage Management panel

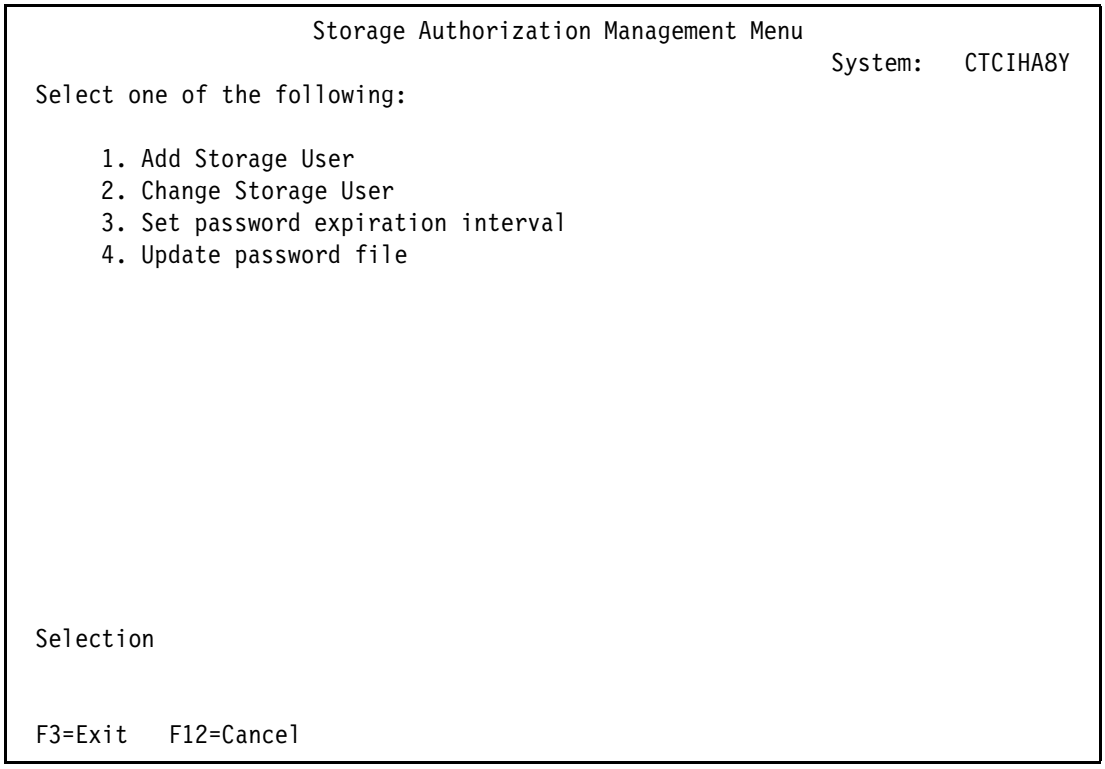


Figure 10-14 ACS Storage Authorization Management Menu

Option 2 (Work with Storage Connections) from the Storage Management Menu builds a map from the IBM i Fibre Channel IOAs to the DS showing all the DS host connection, port login, and volume group information.

```
                          Work with Disk Connections

Storage Device . . . . . . . . . . . . . . . . . :   IBM.2107-75AY031


Type options, press Enter.
  5=Display   10=mkhostconnect   11=Replace IOA

     Resource  IOA   IOA   IOA                          Host  Login  Volume
Opt  Name      Type  Bus   Slot     Host Connect Name   ID    Port   Group
     DC07      280E  768   C2       8YTPCR              0029  I0010   V37
     DC03      280E  775   C5       8yTPCMGM            000F  I0110   V25
     DC01      576B  293   C5    0  HKTEST              0004  I0010   V46
     DC01      576B  293   C5    1                            *****
     DC04      6B25  255   C4                               *****




                                                                      Bottom
Command
===>
F1=Help   F3=Exit   F5=Refresh   F12=Cancel   F17=Save   F18=Display saved
F21=Print report
```

Figure 10-15 ACS Work with Disk Connections panel

Option 5 (Display) on the Work with Disk Connections panel allows the user to display details for an IBM i Fibre Channel connection to the DS showing all the volume and WWPN information (Figure 10-16).

```
                              Connection Details

Press Enter to continue

Resource Name . :   DC03                HostConnect Name:   8yTPCMGM

IO Adapter:                             Host ID . . . . :   000F
  Type  . . . . :   280E                Login Port  . . :   I0110
  Bus . . . . . :   775                 Volume Group  . :   V25
  Slot  . . . . :   C5
                                        Storage Port
System i Tower  :   U5790.001.12940A1     WWPN  . . . . :   50050763070902B8
System i WWPN . :   10000000C967E453
                                        Volumes . . . . :   8

Volume IDs  . . :   EA00 EA01 EA02 EA03 EB00 EB01 EB02 EB03




                                                                        Bottom
F1=Help   F3=Exit   F12=Cancel
```

Figure 10-16 ACS Connection Details panel

Using F17 (Save) on the Work with Disk Connections panel allows the user to save a known good configuration, so whenever a link failure occurs (represented by a **** not logged-in information), the user can easily debug this connection from getting the original port login information back with F18 (Display saved).

Option 11 (Replace IOA) on the Work with Disk Connections panel can be used to have the host connections on the DS automatically updated (deleted and recreated) with the new WWPN information after a replacement of a Fibre Channel IOA.

Option 9 (Check Copy Services Volumes) from the Storage Management Menu allows the user to have ACS verify the ACS Copy Services environment configuration with the IASP configuration. A conflict would be reported as such with the following panel message (Figure 10-17):

```
System i and Copy Services volumes in conflict
```

```
                          Check Copy Services Volumes

Make selections below to subset the information processed.
-
Press Enter to continue

System i:
  ASP list . .   179                                          ASP numbers

Copy Services Environment:
  Name . . . .   TOOLKIT                                      Name
  Type . . . .   MMIR                                         FLASH, GMIR or
                                                               MMIR









F1=Help   F3=Exit   F4=Prompt   F5=Display report   F12=Cancel
F14=Command string
System i and Copy Services volumes in conflict.
```

Figure 10-17 ACS Check Copy Services Volumes panel

With the option F5 (Display report) a detailed log can be displayed (Example 10-1) where the ACS expects IASP volumes in LSS 70/71 while the actual IASP 179 is configured on LSS 02/03.

Example 10-1 /QIBM/Qzrdhasm/Management/CopyServiceVolume.report

```
************Beginning of data**************
#
#  Storage Management - Copy Services Volume Report
#
#  Generated: Mon Sep 19 12:34:12 2011
#
#  Selection input:
#    Storage System . . :  IBM.2107-75AY031
#    System i ASPs . . . . . . . . . :  179
#    Copy Services Environment Name:  TOOLKIT
#                              Type:  MMIR
#
         System i   Volume   Copy Services . . . . . .
           ASP        ID     Environment  Type    Role
         --------    ----    ----------   -----   ----
 Error   ********    7000    TOOLKIT      MMIR    PSrc
 Error   ********    7001    TOOLKIT      MMIR    PSrc
```

```
 Error   ********    7002    TOOLKIT      MMIR    PSrc
 Error   ********    7100    TOOLKIT      MMIR    PSrc
 Error   ********    7101    TOOLKIT      MMIR    PSrc
 Error   ********    7102    TOOLKIT      MMIR    PSrc
 Warning      179    0200    **********   *****   ****
 Warning      179    0201    **********   *****   ****
 Warning      179    0300    **********   *****   ****
 Warning      179    0301    **********   *****   ****
-------------------------------------------------------
   Summary:
     System i volumes:                       4
     Matching Copy Services volumes:         0
     Copy Services volumes not on System i:  6
     System i volumes not in Copy Services:  4
     Copy Services volumes:                  6
 ************End of Data********************
```

Option 11 (Check System i Multipath) from the Storage Management Menu allows the user to check whether the IASP disk units report in as multipath units since the last IPL or IASP switch (that is, multi-path reset) (Example 10-2).

Example 10-2 /QIBM/Qzrdhasm/Management/MultiPath.report

```
************Beginning of data**************
#
#  Storage Management - Multipath Report
#
#  Generated: Mon Sep 19 12:40:42 2011
#
#  Selection input:
#    Storage System . . :  IBM.2107-75AY031
#    System i ASPs:  *ALL
#
#  Non-Multipath volumes are listed below:
#

ASP 179: 0200 0201 0300 0301
ASP 180: 6000 6002 6100 6101 6001 6102
-------------------------------------------
   Summary:
     Non-Multipath System i volumes:  10
 ************End of Data********************
```

The ACS toolkit allows you to receive SNMP messages from the DS storage system for System Reference Code errors or Copy Services events and converts them to QSYSOPR messages (Figure 10-18), which customers usually monitor any time.

```
                        Additional Message Information

Message ID . . . . . :   IAS12O2       Severity . . . . . :   80
Message type . . . . . :   Diagnostic
Date sent  . . . . . :   10/03/11      Time sent  . . . . . :   10:02:42

Message . . . . :   *Primary remote mirror and copy devices on an LSS were
  suspended.
Cause . . . . . : A PPRC volume pair on the storage decice was suspended due
  to an error.  . Use the VIEWLOG command for specific details.
Recovery  . . . : Use the appropriate Copy Services management tools on the
  storage HMC to analyze the problem.



                                                                       Bottom
Press Enter to continue.

F3=Exit   F6=Print   F9=Display message details   F12=Cancel
F21=Select assistance level
```

Figure 10-18 ACS QSYSOPR SNMP message

Corresponding detailed SNMP trap information is logged by ACS in the `/QIBM/Qzrdhasm/qzrdsnmp.log` file (Example 10-3), which can easily be viewed using **VIEWLOG *SNMP**.

Example 10-3 ACS SNMP trap log file

```
2011-10-03 10:02:42 DSM-SNMPTRAP: public, type = 6/202, bindings = 1
                                  2011/10/03 10:10:39 CDT
                                  Primary PPRC Devices on LSS Suspended Due to Error
                                  UNIT: Mnf Type-Mod SerialNm LS LD SR
                                  PRI:  IBM 2107-9B2 75-AY031 EA 00 03
                                  SEC:  IBM 2107-000 75-AY032 EC 00
                                  Start: 2011/10/03 10:10:39 CDT
                                  PRI Dev Flags (1 bit/Dev, 1=Suspended):
                                  8000000000000000000000000000000000000000000000000000000000000000
```

10.3 FlashCopy on Global Mirror target site

Advanced Copy Services for PowerHA extends the Global Mirror base functionality provided by PowerHA SystemMirror for i while also allowing a FlashCopy to be taken from a Global Mirror target site (for example, for a tape backup done at the remote site) (Figure 10-19).

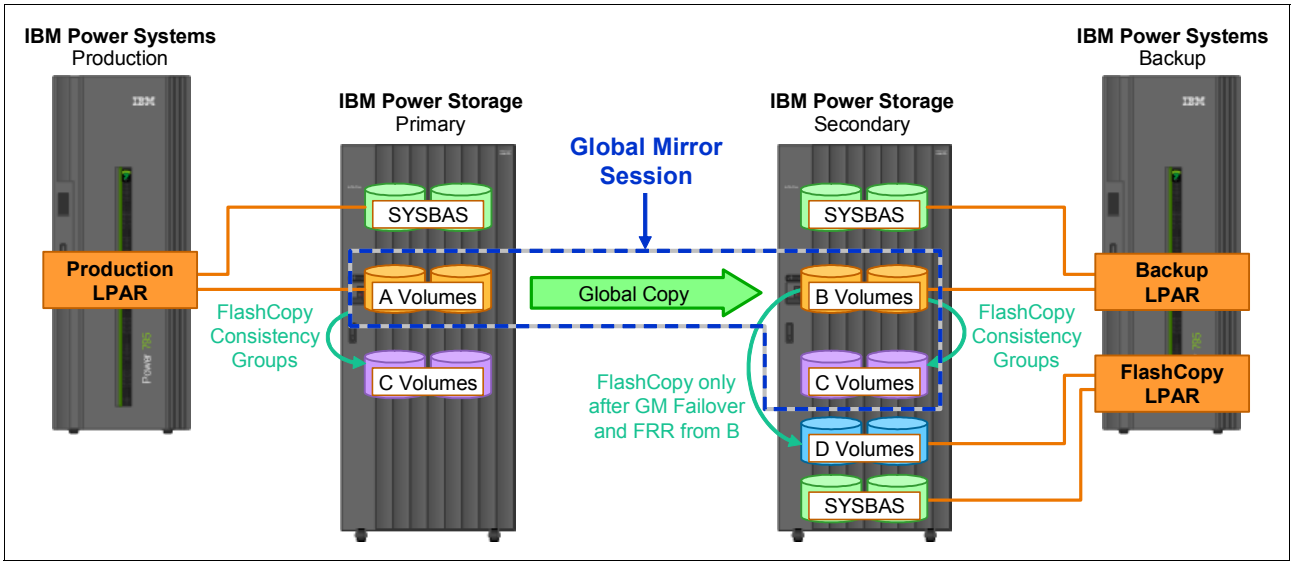


Figure 10-19 ACS FlashCopy on Global Mirror target site

Configuration of a FashCopy D volume with ACS is done when creating a Global Mirror environment (Figure 10-7 on page 173) by specifying **`D-Copy Flash normal *YES`** to set up a FlashCopy D volume configuration on the Global Mirror target site, and by specifying **`D-Copy Flash reversed *YES`** to allow for a FlashCopy on the target site also when the Global Mirror direction is reversed (that is, the DS storage system at the target or secondary site acts as the Global Mirror source).

Creating a FlashCopy to the D volumes, which involves an automatic prior Global Mirror failover and FlashCopy fast-reverse-restore from the C volumes to the B volumes before consistent data is on the B volumes, which can finally be flashed to the D volumes, can be initiated by the user either with an option from the Global Mirror environment (Figure 10-8 on page 174) or by using **`STRFLASH`** (Figure 10-9 on page 175).

10.4 Metro/Global Mirror and TPC-R support

When an IBM System Storage DS8000 3-site disaster recovery solution with Metro/Global Mirror (MGM) support is desired (Figure 10-20), it is supported with ACS only with the IBM Tivoli® Storage Productivity Center for Replication (TPC-R) software product.

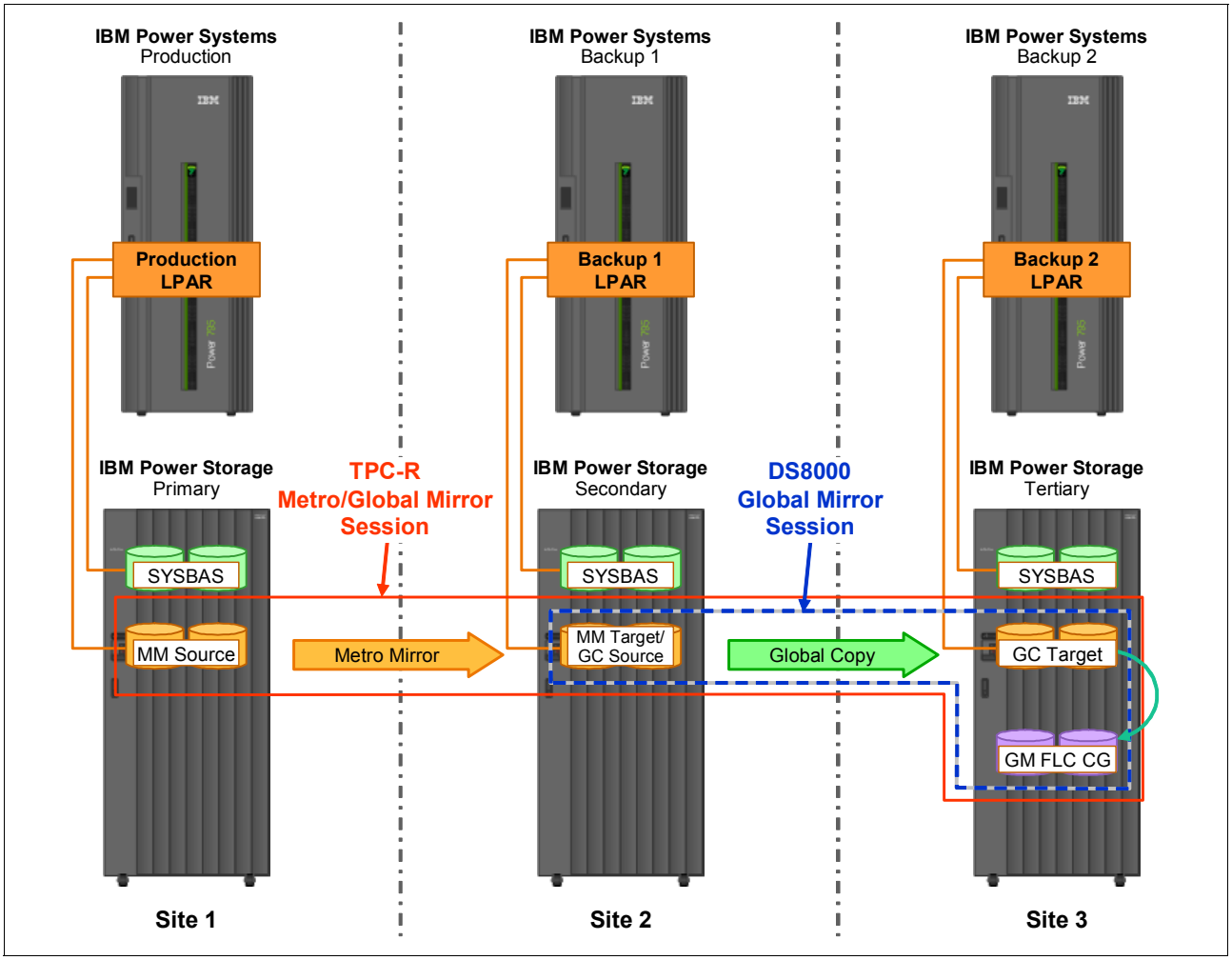


Figure 10-20 DS8000 3-site disaster recovery solution with Metro/Global Mirror

Due to its *session* concept, TPC-R makes managing consistency across multiple Copy Services volume *sets* and failover or corrective actions for a session in a failed state easy (Figure 10-21).

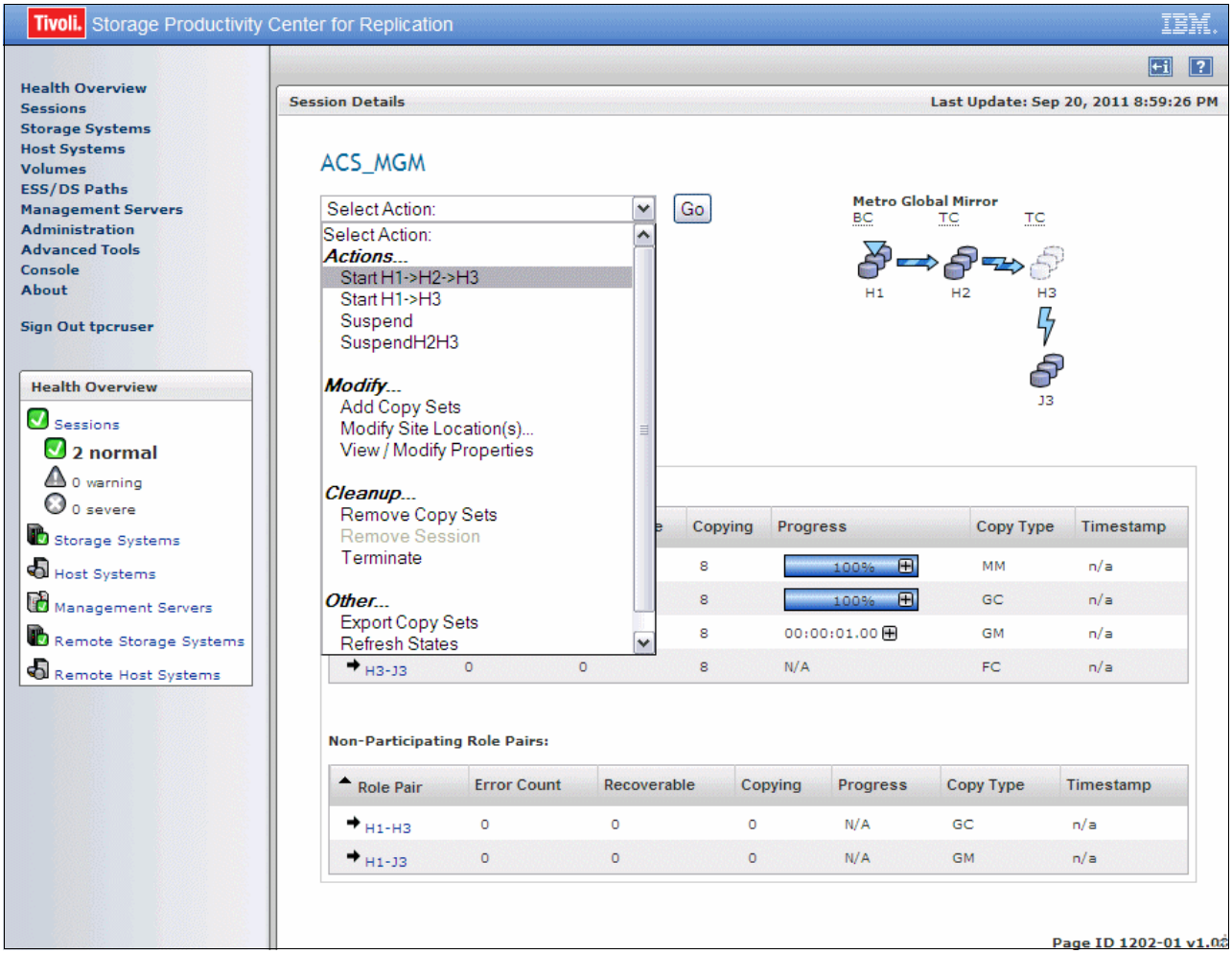


Figure 10-21 TPC-R Metro/Global Mirror session

Other reasons for using TPC-R with ACS are if a customer likes to use Metro Mirror consistency groups or uses TPC-R already with another environment and wants the IBM i server platform included. Even with configured TPC-R support, for FlashCopy ACS always uses its DSCLI interface. Similarly, ACS also does not make use of TPC-R's Metro/Global Mirror with Practice session, as it always fails over to the practice volume instead of the GlobalCopy target volumes.

ACS interaction with TPC-R is supported for either Metro Mirror or Metro/Global Mirror such that the user sets up communication with the TPC-R server for the ACS Copy Services environment (Figure 10-22).

```
                          Change a GMIR Environment
Type choices, press Enter.

Environment  . . . . . :   IASP01

Global Mirroring Power HA, ASP information:
  Device name  . . . . . .   IASP01                  Name
  GMIR Src is MMIR Src . :   *YES
  Source Copy Description    PROD                    Name
  Target Copy Description    DR                      Name

TPC information:
  TPC Replication  . . . .   *YES                    *YES, *NO
  Metro-Global Mirroring     *YES                    *YES, *NO
  Primary server . . . . .   9.5.167.82              IPv4
  Secondary server . . . .   9.5.167.57              IPv4
  Session name . . . . . .   MGM                     Name
  User . . . . . . . . . .   tpcruser                Profile name
  Password . . . . . . . .

                                                                        More...
F1=Help   F3=Exit   F12=Cancel
```

Figure 10-22 ACS Change a GMIR Environment panel

ACS will then stop using the DSCLI and instead use its Java API interface into the TPC-R server for managing the Copy Services functions. The CSE stream file scripts for DSCLI still get created but without accompanying result files. Instead, there are Tpcr*Rtn.xml result files showing the return status of the ACS Java API communication with the TPC-R server.

Note: Before planning to implement a Metro/Global Mirror environment with ACS for IBM i, contact IBM STG Lab Services, as it requires careful planning to ensure that customer requirements are met:

http://www-03.ibm.com/systems/services/labservices/contact.html

10.5 Custom programming

ACS supports custom programming for a Copy Services environment by allowing a user to retrieve information from the DS storage system or TPC-R server that can be further processed by a user-written program and used to invoke desired actions either on the storage system or the TPC-R server.

Using **RUNDSCMD**, a user can run a DSCLI script for retrieving the output in a comma-separated text file that can be validated using a CL program, which in turn can make use again of the existing ACS Copy Services environment stream files.

For example, the verification of an existing FlashCopy relationship can be done by querying for the DSCLI message code CMUC00234I, which is returned for a non-existing FlashCopy relationship, as shown by the validation routine example in Figure 10-23.

```
                    Run DS Scripted Command (RUNDSCMD)

Type choices, press Enter.

Result validation list:
  Column position  . . . . . > 1             1-20
  Expected value . . . . . . > 'CMUC00234I'
  Logic to next in list  . . .                *AND, *OR
             + for more values
Result file rows . . . . . .   *ALL          *ONE, *ALL
Summation column . . . . . .   *NONE         *NONE, 1-20
CL variable for returned total                TYPE(*DEC) LEN(9 0)
Return column  . . . . . . .   *NONE         *NONE, 1-20
Return key value . . . . . .   *NONE
CL variable for returned value                TYPE(*CHAR) LEN(80)
Comment  . . . . . . . . . . .


                                                                      Bottom
F3=Exit   F4=Prompt   F5=Refresh   F10=Additional parameters   F12=Cancel
F13=How to use this display        F24=More keys
```

Figure 10-23 ACS Run DS Scripted Command panel

Another example of using validation with **RUNDSCMD** is using its `Summation column` and `CL variable for returned total` parameters to query the sum of certain values like the out-of-sync tracks from all listed PPRC relationships (for example, to write a customized program to predict the remaining synchronization time).

By using the Return column, Return key value, and CL variable for returned value in the validation program, the user can search for the specified key value in the DSCLI output and get the value at the specified return column returned in a CL variable for further processing in customized programming.

RUNTPCACT of ACS allows a user to run any TPC-R session commands, such as **'Start H1->H2->H3'** or **'Suspend'** (Figure 10-24).

```
                          Run TPC Action (RUNTPCACT)

Type choices, press Enter.

Environment name . . . . . > IASP01        Name
Type . . . . . .   *MMIR         *GMIR, *MMIR
Action . . . . . > *RUN          *CHK, *RTV, *RUN, *ACS, *RCS
TPC command  . . . . . .   'Suspend'




                                                                      Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 10-24 ACS Run TPC Action command panel

RTVTPCCMD can be used to list the available TPC session actions as output in a CL variable, which can be processed/verified in a CL program before running an available action to take control of TPC-R from an IBM i client.

10.6 IBM i full-system FlashCopy replication

STG Lab Services also offers a *Full-System FlashCopy Toolkit (FSFC)* for IBM i 5.4 and later that is installed on a production partition to be cloned via FlashCopy and a managing partition that controls the entire FlashCopy process via IP communication to the production and FlashCopy backup partition including SSH communication through the HMC (Figure 10-25).

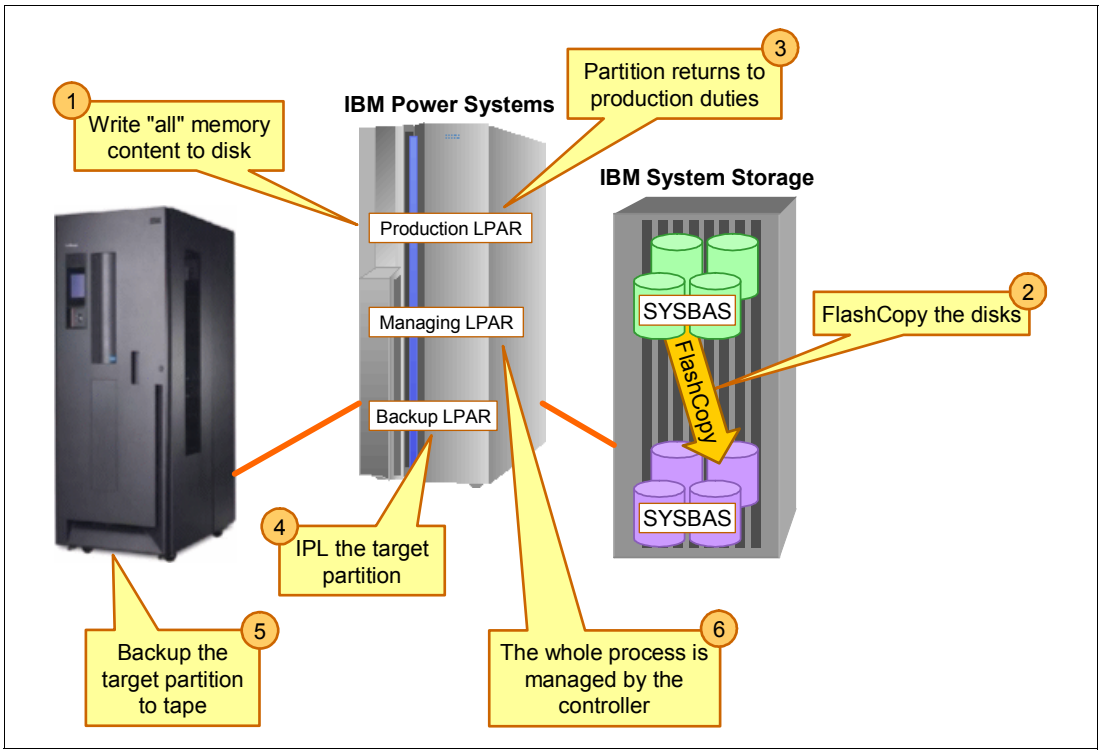


Figure 10-25 Full-System FlashCopy Toolkit

FSFC is licensed separately from Advanced Copy Services for PowerHA but uses the same Copy Services Environment menus for setting up the profiles and scripts to communicate with the DS storage system. It needs to be installed on the managing and production partition in the QZRDIASH5 library.

Note: While taking a warm FlashCopy with the full-system or IASP online, even when quiesced, any object that is resident in memory and not journaled, such as files, data queues, data areas, and so on, is subject to potential damage due to object inconsistencies from parts of the object not on disk. Journaling allows the target partition to recover from possible damaged journaled objects by applying journal entries to journaled files during the IPL.

It is only with a full-system FlashCopy that there is a chance that on taking a warm FlashCopy some IBM i internal objects can be found corrupted at IPL time. You might need a SLIP install, or another FlashCopy might be needed for a recovery.

In addition to controlling the IBM i full-system FlashCopy process itself, FSFC also provides a high level of integration with IBM i Backup Recovery and Media Services (BRMS) to make sure that the backup by BRMS run from the FlashCopy backup partition appears in the BRMS history containing all the backup and media information, as was done from the production

partition with the BRMS history (QUSRBRM) automatically transferred to the production partition after the backup has finished (Figure 10-26).

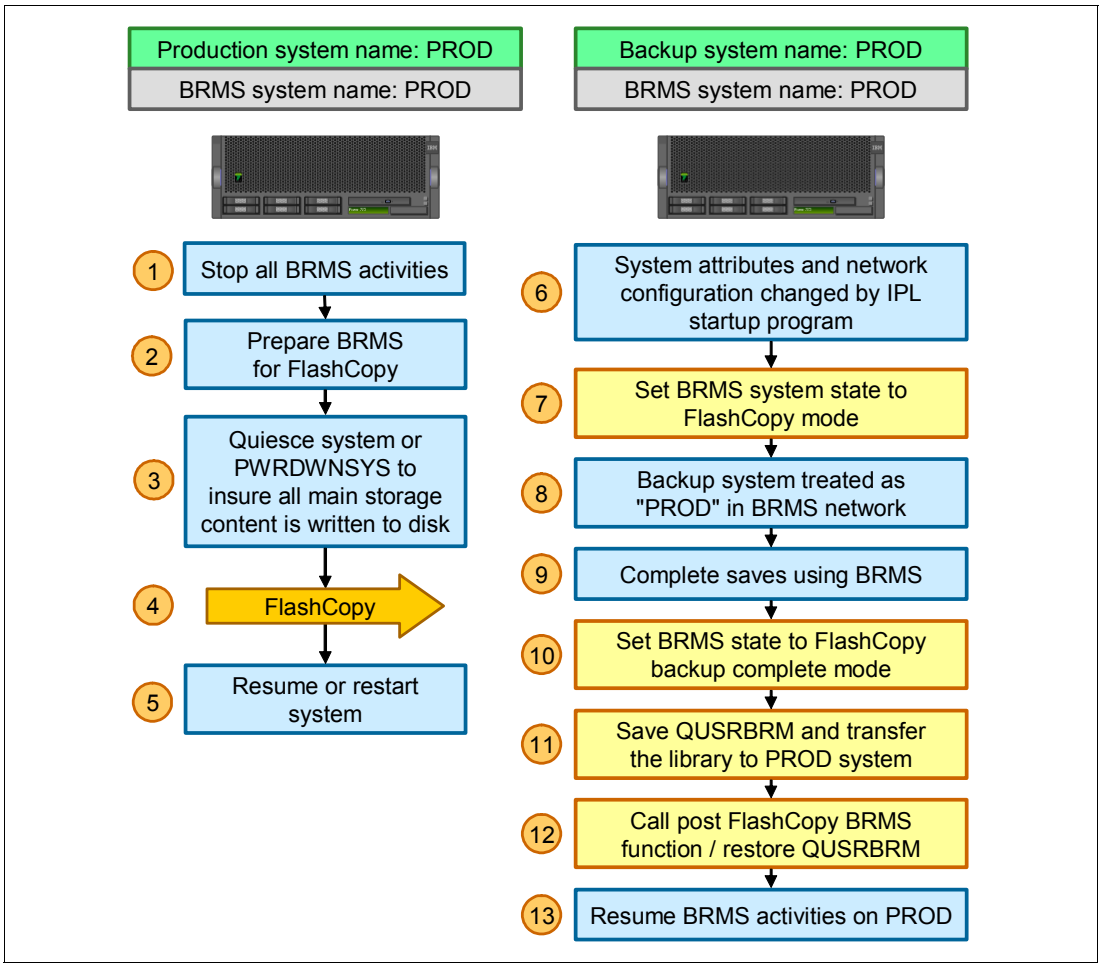


Figure 10-26 Full-system FlashCopy toolkit integration with BRMS

The FSFC full-system FlashCopy functions are configured on the managing partition using **CRTSYSCPY** (Figure 10-27 on page 194). In our example we show a FlashCopy configuration where CTCIHA7A is the production partition to be flashed via the managing partition to the target partition CTCIHA7C with quiescing the production partition (OPTYPE *QUIESCE) before taking the FlashCopy, resuming the production partition, and activating the backup partition by starting its specified IP interface and BRMS backup job.

```
                    Create Full Sys Flash Copy (CRTSYSCPY)

 Type choices, press Enter.

 Configuration Name . . . . . .   FLC_HA7A      Character value
 Environment name . . . . . .     HA7A           F4 to prompt
 Source partition host name . . .  CTCIHA7A.RCHLAND.IBM.COM

 Source HMC partition name  . . .  CTCIHA7A
 Source partition profile . . . .  DEFAULT
 Source managing system . . . . .  CTCIHA7
 Source HMC1 address  . . . . . .  9.5.168.169
 Source HMC2 address  . . . . . .  *NONE
 Shutdown target  . . . . . .      *YES          *YES, *NO
 Restart target partition . . . .  *YES          *YES, *NO, *INQ, *OFF
 Target LPAR IPL source . . . . .  *PANEL        *PANEL, A, B, D
 Target LPAR keylock position . .  *PANEL        *PANEL, *AUTO, *MANUAL
 Target LPAR pre-activation pgm    *NONE         *NONE, name
   Library  . . . . . .            *LIBL         *LIBL, library name

                                                                    More...
 Target partition host name . . .  CTCIHA7C.RCHLAND.IBM.COM

 Target HMC partition name  . . .  CTCIHA7C
 Target partition profile . . . .  DEFAULT
 Target managing system . . . . .  *SOURCE
 Target HMC1 address  . . . . . .  *SOURCE
 Target HMC2 address  . . . . . .  *NONE
 Backup Application . . . . . .    *BRMS         *NATIVE, *BRMS
 Lock BRMS  . . . . . .            *BOTH         *BOTH, *NO, *SRCONLY...
 Restricted BRMS media classes  .  *NONE           F4 to prompt
               + for more values

                                                                    More...
Target LPAR Device Config:
   Backup device name . . . . . .  TS3400        *NONE, device name
   Backup device serial number  .  78-1011003    *NONE, serial number
   Robot host . . . . . .          *NOCHANGE
   Local internet address . . . .  *NOCHANGE
               + for more values
 Program to move QUSRBRM  . . . .  *SAVF_SOCK    *SAVF_SOCK,
*VRTTAP,*NONE,name
   Library  . . . . . .            *LIBL         *LIBL, library name
 Save compression for QUSRBRM . .  *DEV          *DEV, *YES, *NO, *LOW...
 Operation type . . . . . .        *QUIESCE      *QUIESCE, *IPL, *NOIPL
 Storage Type . . . . . .          *DS8K         *DS5K, *DS8K, *SVC
 Issue IPL confirmation message    *NO           *YES, *NO
 Source LPAR shutdown command . .

More...
 F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
 F24=More keys
```

Figure 10-27 Full-system FlashCopy toolkit CRTSYSCPY command panel

Figure 10-28 shows the remaining pages from the CRTSYSCPY command panel.

```
Minutes to wait for power down     60            minutes
Force *SYSBAS resume . . . . . .   60            seconds
Force flash copy . . . . . . . .   *NO           *YES, *NO
FlashCopy exit program . . . . .   *NONE         *NONE, name
  Library  . . . . . . . . . . .     *LIBL       *LIBL, library name
Minutes to wait for power up . .   600           minutes
Target LPAR startup program  . .   *NONE         *ORIG, *NONE, name
  Library  . . . . . . . . . . .     *LIBL       *LIBL, library name
Startup program job queue  . . .   QCTL          *NONE, name
  Library  . . . . . . . . . . .     QSYS        *LIBL, library name
Hold scheduled jobs on target  .   *YES          *YES, *NO
Target LPAR default route:
  Binding interface  . . . . . .   *NOCHANGE
  Next hop . . . . . . . . . . .

                                                                      More...
Target Comm Interfaces:
  IO card serial number  . . . .   00-EC756OA    *NONE, Serial number
  Line Description . . . . . . .   FSFCL         *NEW, line name
  IO card port number  . . . . .   0             0-32
  IO card IP address . . . . . .   9.5.168.117
  Network Mask . . . . . . . . .   '255.255.255.0'
              + for more values
Target partition backup cmd  . .   STRBKUBRM USERDATA SBMJOB(*NO)

              + for more values

Wait for final notification  . .   *NO           *YES, *NO
                                                                      More...
Stop target after process  . . .   *RMV          *YES, *NO, *RMV
```

Figure 10-28 Full-system FlashCopy toolkit CRTSYSCPY command panel (continued)

After the Copy Services environment has been created by using **WRKCSE** and the configuration using **CRTSYSCPY**, a full-system FlashCopy can be started by using **MKSYSCPY** (Figure 10-29).

```
                      Make Full System Copy (MKSYSCPY)

Type choices, press Enter.

Configuration Name . . . . . .   FLC_HA7A        F4 to prompt
User Profile . . . . . . . . .   POWERHA        *DDM, User Profile
Password . . . . . . . . . . .                  Character value



                                                                     Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 10-29 Full-system FlashCopy toolkit STRFLASH command panel

For further information about the full-system FlashCopy toolkit service offering contact STG Lab Services here:

http://www-03.ibm.com/systems/services/labservices/contact.html

Part 3

Implementation examples and best practices

In this part, the final part of the book, we show you practical scenarios and implementation examples with step-by-step instructions for how to configure and manage your high-availability environment using IBM PowerHA SystemMirror for i.

This part has the following chapters:

- Chapter 11, “Creating a PowerHA base environment” on page 199
- Chapter 12, “Configuring and managing Geographic Mirroring” on page 235
- Chapter 13, “Configuring and managing DS8000 Copy Services” on page 263
- Chapter 14, “Configuring and managing CSVC/V7000 Copy Services” on page 371

We also discuss best practices to consider and follow in setting up and managing your high-availability environment in Chapter 15, “Best practices” on page 397.

11

Creating a PowerHA base environment

In this chapter, we show you how to set up the basic building blocks of a highly available environment using PowerHA. This chapter provides a step-by-step approach for the following tasks:

- Creating a cluster
- Setting up cluster monitors and advanced node failure detection
- Creating an IASP
- Setting up an administrative domain

In this section, we set up a basic *cluster* environment with two nodes in it. We then create an IASP on the production site. We set up advanced failure detection by adding a cluster monitor and registering the cluster nodes with the HMC CIM server. In addition, we create an administrative domain and add monitored resource entries to it. The entire setup is done using the new PowerHA GUI. In addition, we provide you with the CL commands that you can use alternatively to do this setup.

11.1 Creating a cluster

To create your basic cluster setup using the new PowerHA GUI, take the following steps:

1. Access and log in to IBM Systems Director Navigator for i from the following URL:

   `http://<your_server_ip_address>:2001`

   Figure 11-1 shows you the list of possible tasks.

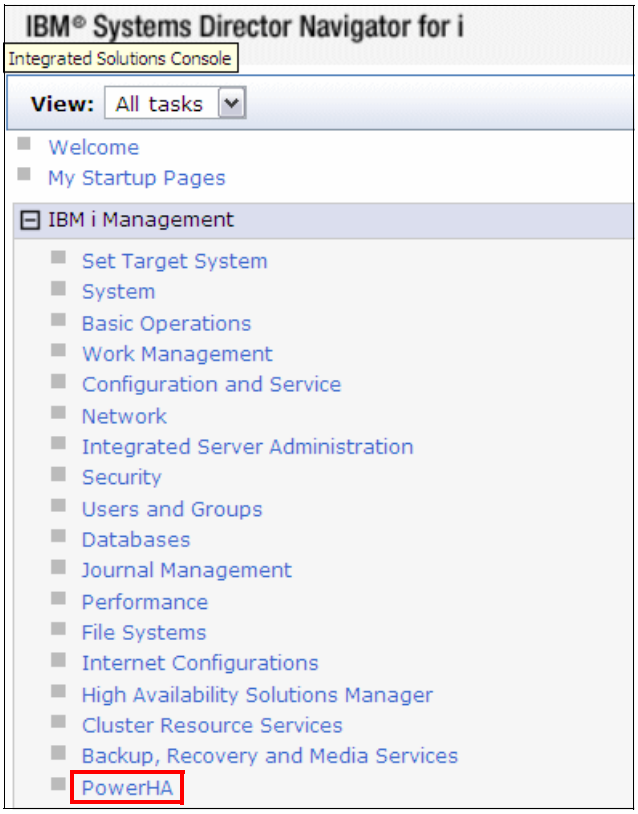


Figure 11-1 IBM Systems Director Navigator for i: Welcome panel

2. Choosing **PowerHA** leads you to the new PowerHA GUI. The GUI connects to your system and checks whether any cluster configuration has already been done. This is not the case in our environment (Figure 11-2). Click **Create a new Cluster** to start the wizard.

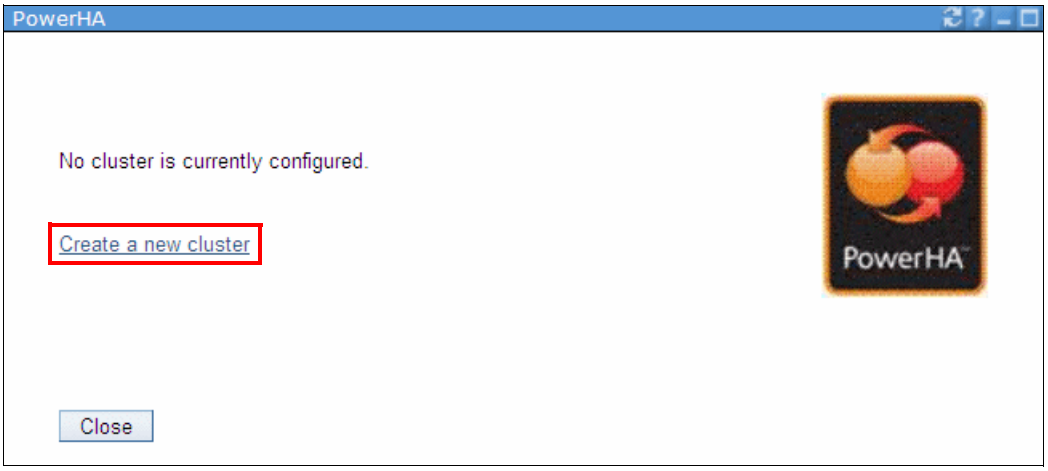


Figure 11-2 PowerHA GUI: Create new cluster

3. The wizard guides you through the steps necessary to create the cluster and asks you for all information required for the setup. Figure 11-3 shows the steps.

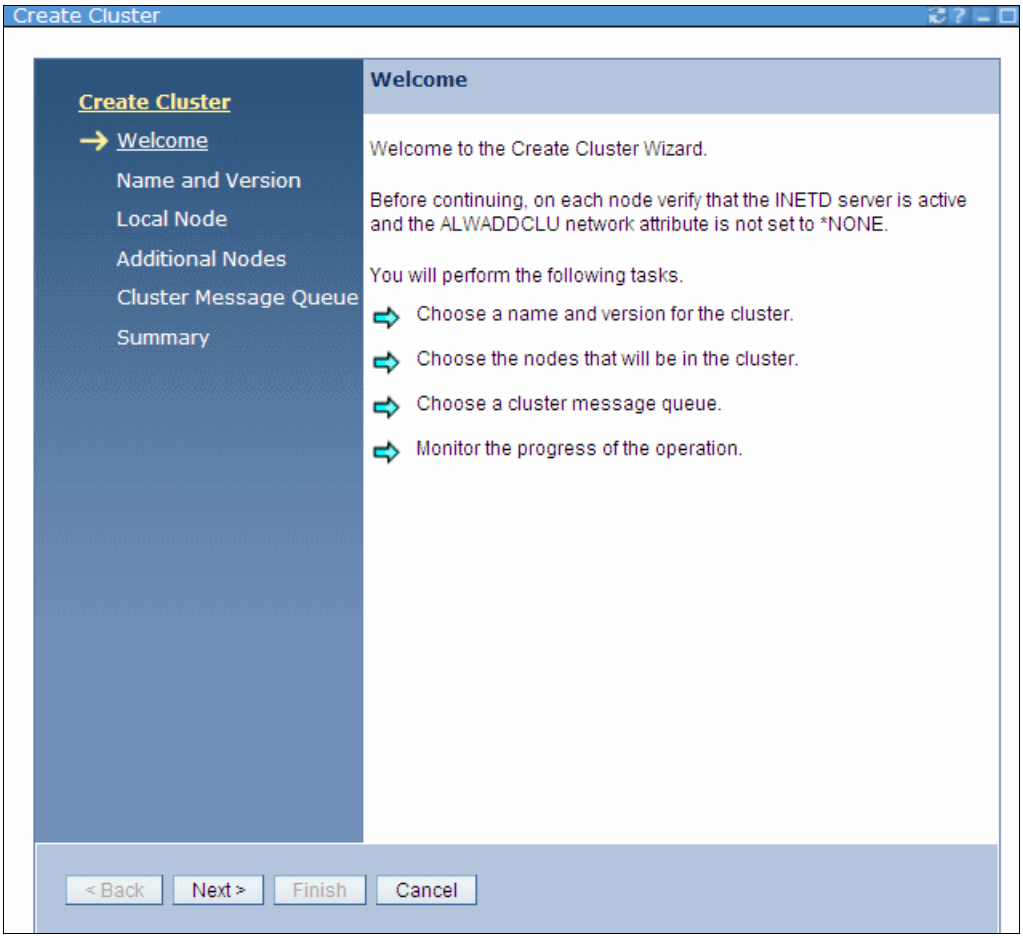


Figure 11-3 PowerHA GUI: Welcome screen

4. Skipping over the welcome page takes you to the page shown in Figure 11-4. Here, you enter the name of your cluster (PWRHA_CLU in our case). PowerHA version, cluster version, and cluster modification default to the most current levels the system that you are connected to supports. These values can be changed. However, PowerHA Version 2.1 with cluster Version 7 is the minimum level required for the Power HA GUI to work properly.

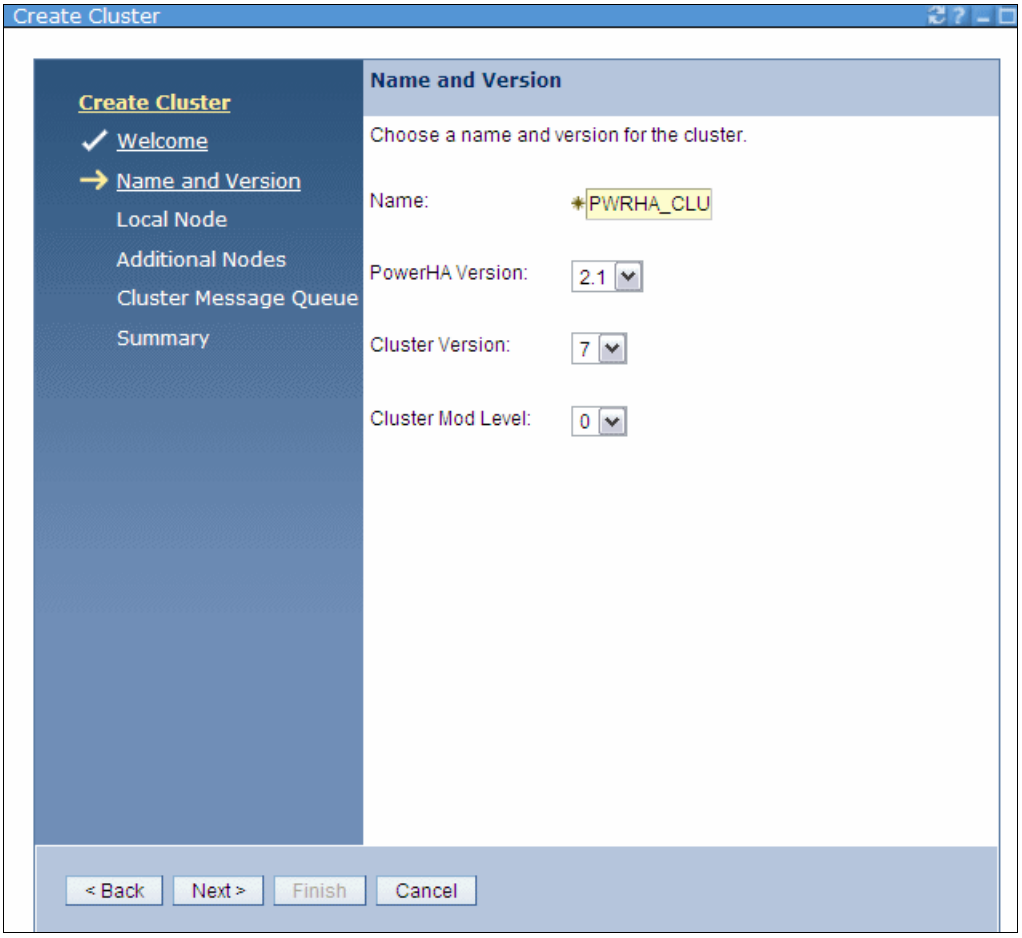


Figure 11-4 PowerHA GUI: Cluster name and version

5. On the next page you enter cluster node information pertaining to the system that you are connected to. As can be seen in Figure 11-5, the node name defaults to the current system name value from the network attributes of the system, but it can also be changed. The cluster IP addresses you provide here are used for cluster heartbeat and internal cluster communication. For a production environment it is a good practice to use two dedicated, redundant Ethernet ports with cluster heartbeat addresses, especially when not using IBM i 7.1 advanced node failure detection to help prevent cluster partition conditions because of single-points-of-failure in the cluster communication setup.

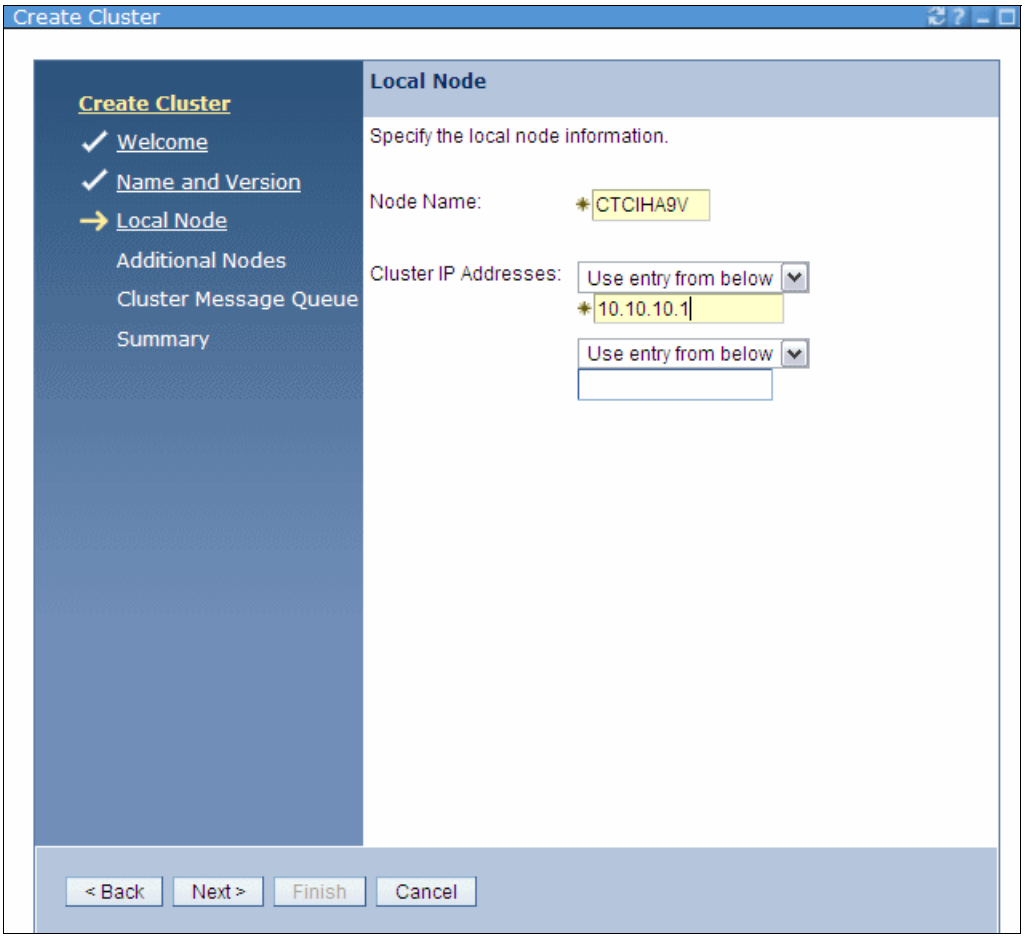


Figure 11-5 PowerHA GUI: Local node information

6. In the next step, you additional nodes into the cluster. You have to provide the node name and the cluster IP addresses used for heartbeat (Figure 11-6). You can also specify whether that cluster node should be started when the creation of the cluster is finished.

> **Note:** To have clustering on a node started automatically after an IPL, change the system's startup program with adding a STRCLUNOD entry.

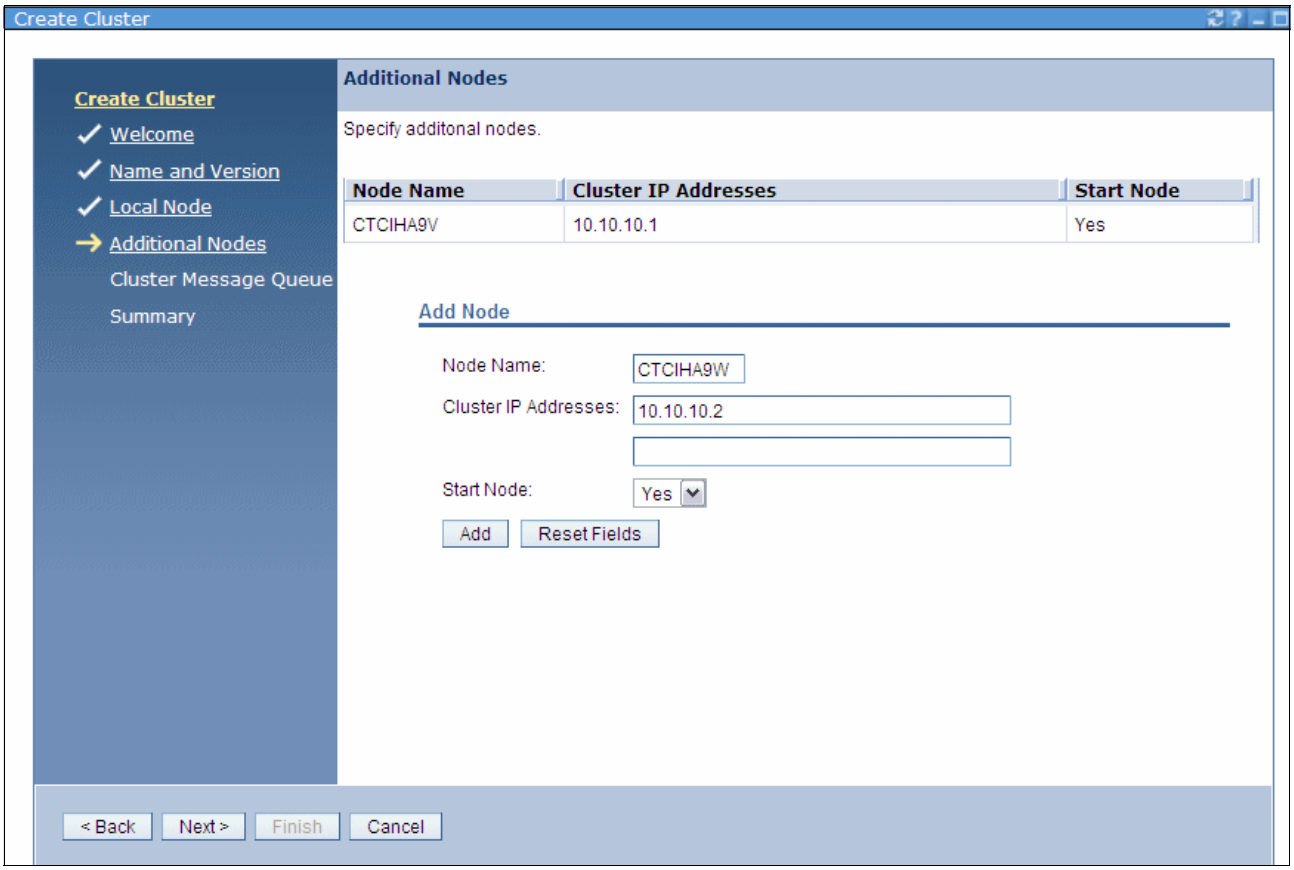


Figure 11-6 PowerHA GUI: Add additional nodes

7. Specify whether you want to define a cluster-wide failover message queue. Should the primary node fail, a message is sent to this message queue on the node that the cluster would fail to. You can also specify how long this message should wait for an answer and what the default action should be if there is no answer within that time frame (Figure 11-7). This setting is then valid for all cluster resource groups (CRGs) within this cluster. Alternatively, you can define failover message queues for each CRG individually.

   Note: If you do not specify any failover message queue, then failover happens immediately without any message being sent in advance.

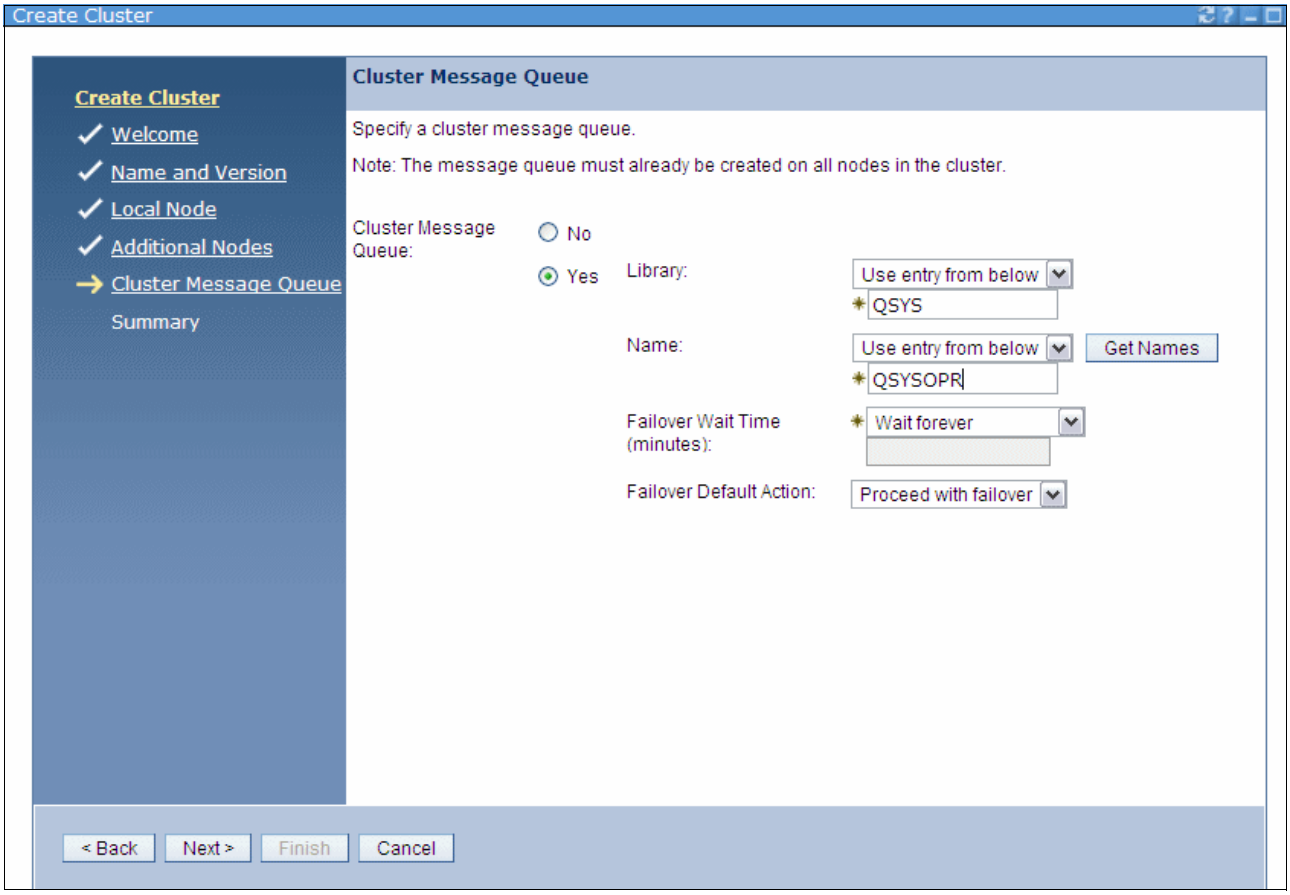


Figure 11-7 PowerHA GUI: Specify cluster message queue

8. As shown in Figure 11-8, the summary page provides you with an overview of the data that you entered in the wizard. Click **Finish** to create the cluster.

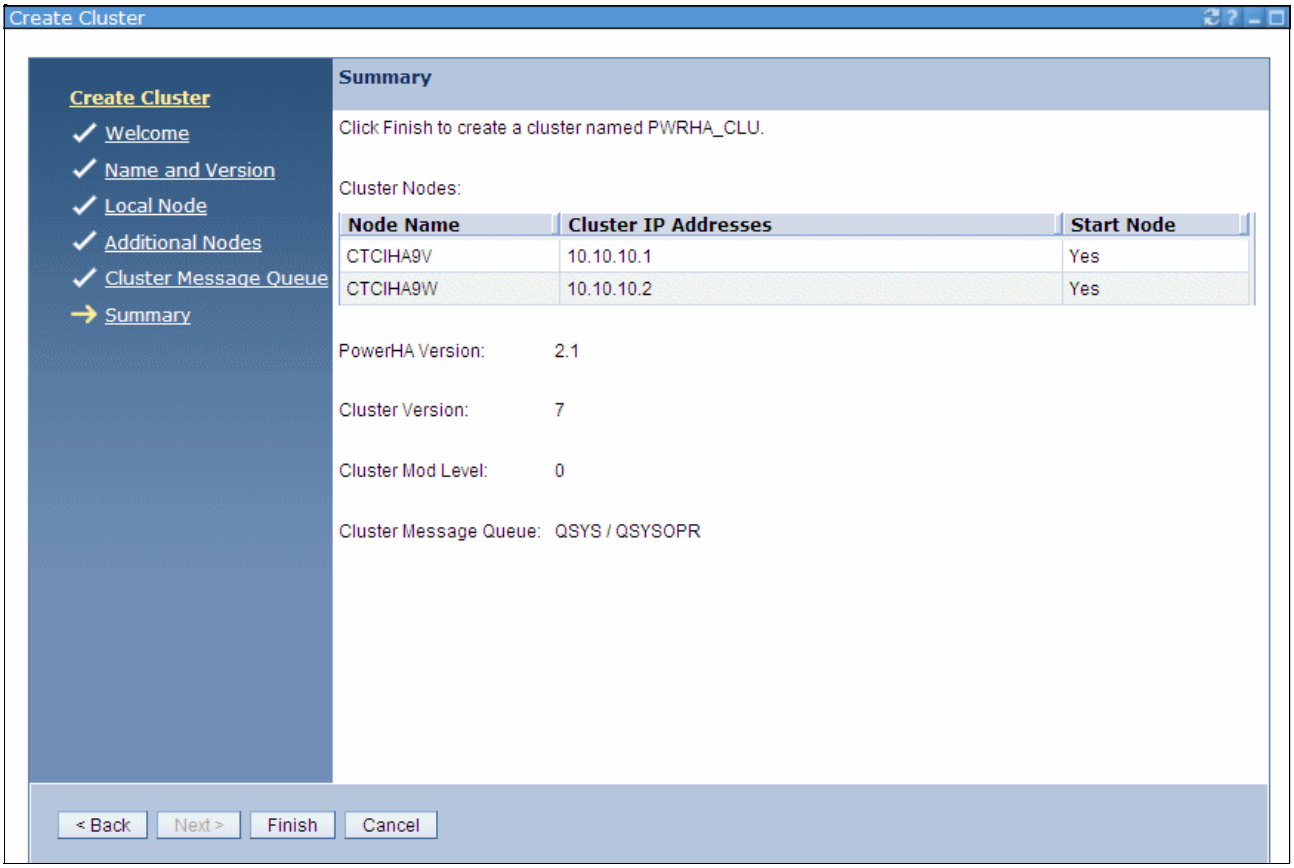


Figure 11-8 PowerHA GUI: Create cluster summary

9. You will see the page shown in Figure 11-9 after the cluster is created and all nodes are started. The PowerHA GUI can then help you determine whether all requirements for successfully running a cluster environment are met. Click **Check Requirements** to do so.

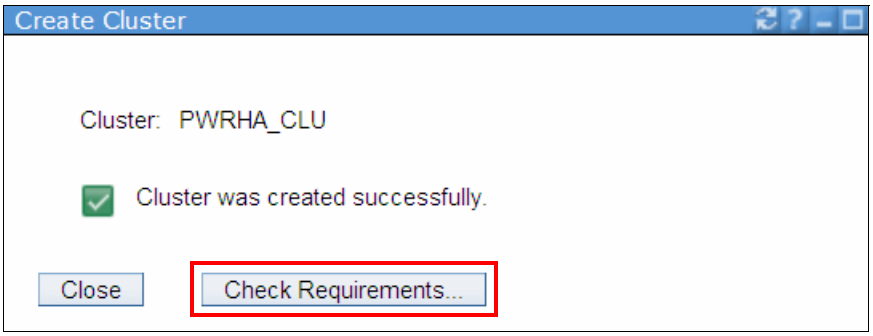


Figure 11-9 PowerHA GUI: Cluster successfully created

10. In our example, we receive a list of warnings and suggestions (Figure 11-10). For example, the system value QRETSVRSEC has to be set to 1 if you want to use an administrative domain to synchronize user passwords between the nodes of your cluster.

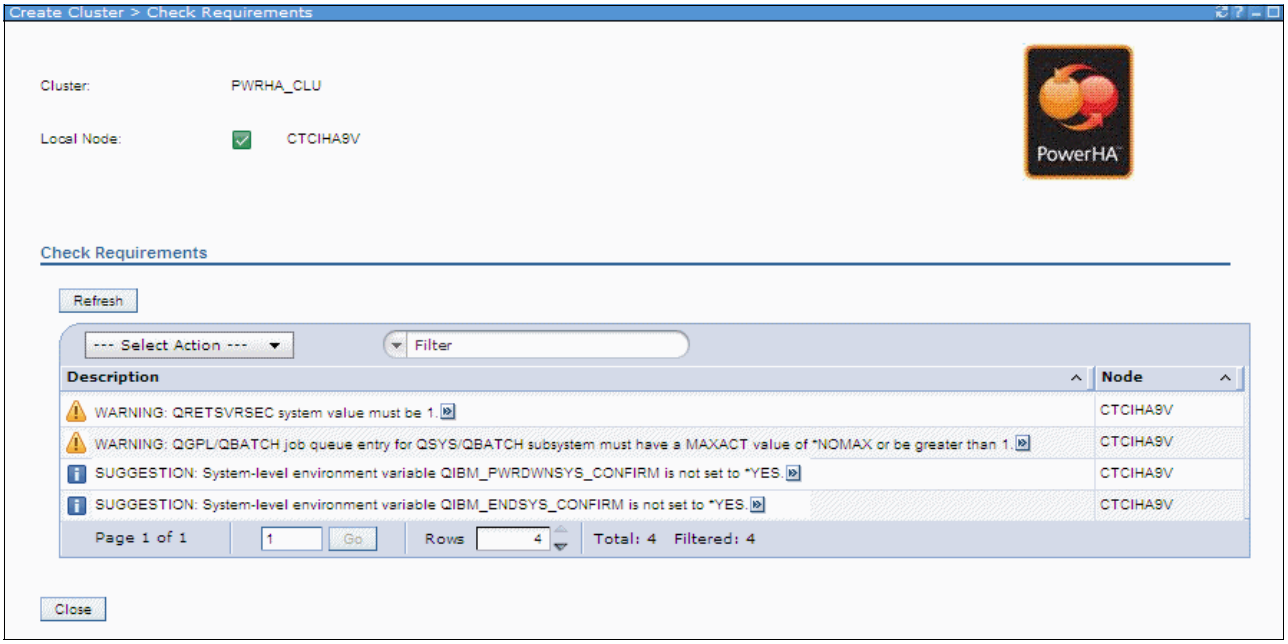


Figure 11-10 PowerHA GUI: Check requirements

11. Clicking the toggle beside the message opens a Fix icon (Figure 11-11). If you click this button corrective action is taken on the respective cluster node.

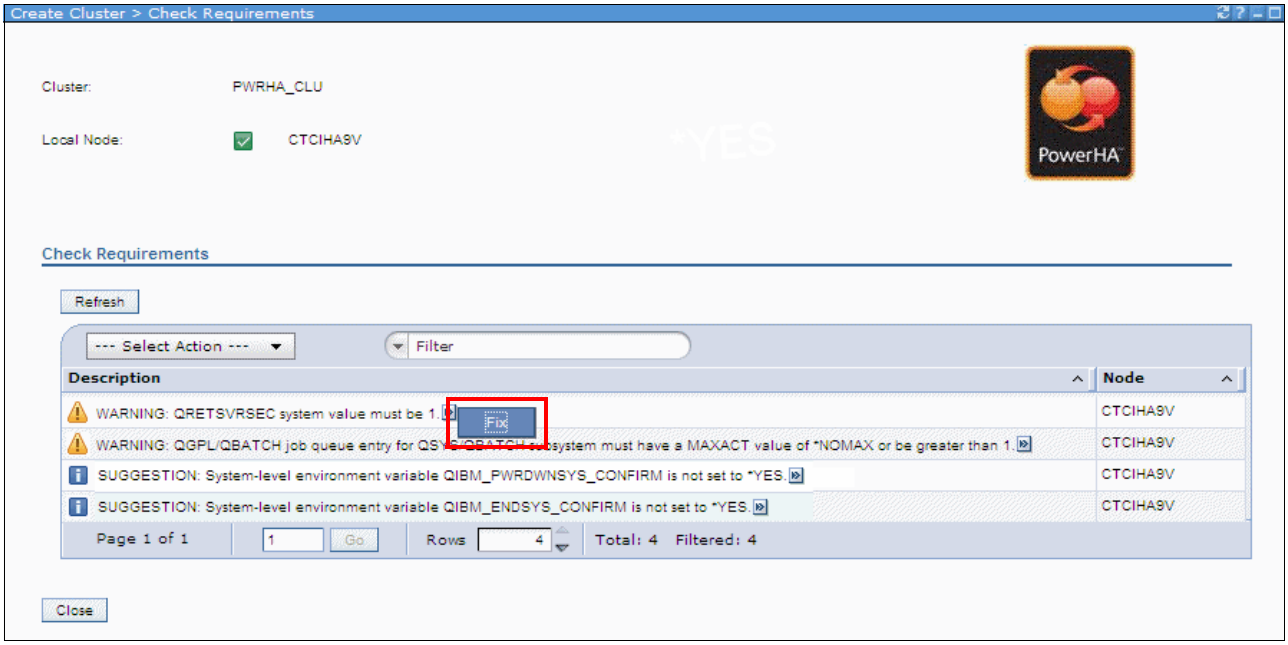


Figure 11-11 PowerHA GUI: Fix requirements

12. After all the requirements are met, you will see a page like that show in Figure 11-12. Close this information page to continue with your cluster setup.

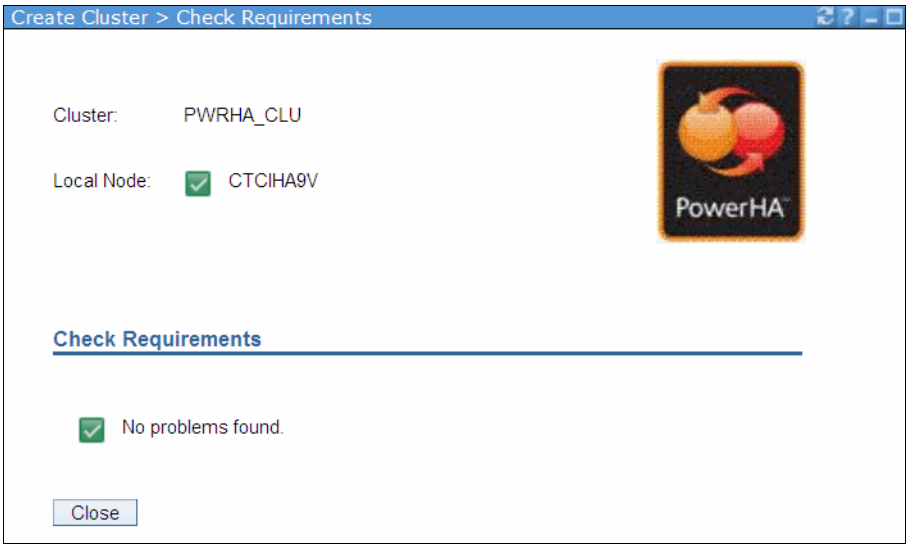


Figure 11-12 PowerHA GUI: Cluster requirements met

11.2 Setting up cluster monitors

Cluster monitors were introduced with IBM i 7.1. Setting them up can help to avoid *cluster partition* situations that occur because a cluster node was not able to send out a panic message before going down. To allow the HMC CIM server to notify registered IBM i cluster nodes of sudden partition or system failures, you need to set up SSH communication between your cluster nodes and the HMC.

Steps to set up SSH

Take these steps to set up SSH:

1. License program 5733-SC1 (IBM Portable Utilities for i) with options base and 1 have to be installed on your cluster nodes.
2. License program 5770-UME (IBM Universal Manageability Enablement for i) has to be installed on your cluster nodes.
3. TCP/IP server *CIMOM has to be started.
4. The *CIMOM TCP server must be configured and started on each cluster node that has a cluster monitor configured on it. The default configuration of the *CIMOM server that is provided by the installation of the 5770-UME license program must be changed so that the IBM i system can communicate with the CIM server. To do that, two configuration attributes that control security aspects need to be changed by running **`cimconfig`** within a PASE shell:
   a. With the *CIMOM server running, start a PASE shell from the command line with **`call qp2term`**.
   b. Enter:

```
/QOpenSys/QIBM/ProdData/UME/Pegasus/bin/cimconfig -s
enableAuthentication=false -p
```

c. Enter:

```
/QOpenSys/QIBM/ProdData/UME/Pegasus/bin/cimconfig -s
sslClientVerificationMode=optional -p
```

d. End the PASE shell and restart the CIMOM server using **ENDTCPSVR *CIMOM** and **STRTCPSVR *CIMOM**.

5. Add a digital certificate from your HMC to the certificate truststore using these steps:
   a. TCP server *SSHD must be running. If this is not the case start it with **STRTCPSVR *SSHD**.
   b. The IBM i user profile used to copy the certificate has to have a home directory associated with the profile, and that directory must exist.
   c. You must use the physical monitor and keyboard attached to your HMC. You cannot use Telnet or a web interface to the HMC.
   d. Open a restricted shell on the HMC.
   e. Use the secure copy command to copy a file to your IBM i cluster node:

      ```
      scp /etc/Pegasus/server.pem QSECOFR@CTCIHA9V:/server_name.pem
      ```

      In the above command, change QSECOFR to the IBM i profile that you want to use, change CTCIHA9V to your IBM i IP name, and change `server_name.pem` to the file name that you want to use for the certificate file, for example, my_hmc.pem.
6. Sign off the HMC.
7. On your IBM i system, start the PASE shell environment using **call qp2term**:
   a. Move the HMC digital certificate:

      ```
      mv /myhmc.pem /QOpenSys/QIBM/UserData/UME/Pegasus/ssl/truststore/myhmc.pem)
      ```

      Replace the name, myhmc.pem, with your specific file name.
   b. Add the digital certificate to the truststore:

      ```
      /QOpenSys/QIBM/ProdData/UME/Pegasus/bin/cimtrust -a -U QSECOFR -f
      /QOpenSys/QIBM/UserData/UME/Pegasus/ssl/truststore/myhmc.pem -T s
      ```

      Replace the name, myhmc.pem, with your specific file name.
   c. Exit the PASE shell by pressing F3.
8. Restart the CIM server to pick up the new certificate by doing an **ENDTCPSVR *CIMOM** and a **STRTCPSVR *CIMOM**.

9. When all these requirements are met you can add a cluster monitor to each of your cluster nodes. From the PowerHA GUI main menu, choose **Cluster Nodes** and then open the properties information of your first cluster node (Figure 11-13).

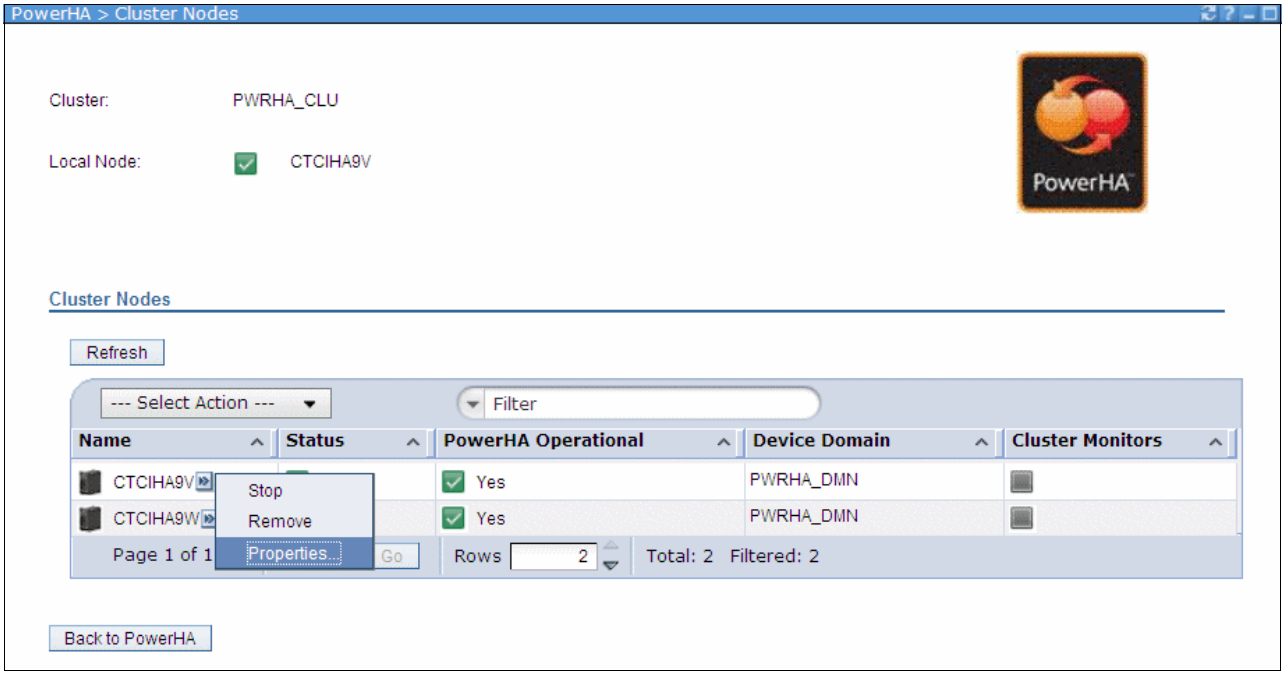


Figure 11-13 Cluster Nodes: Show properties

As can be seen in Figure 11-14, there is currently no cluster monitor defined for node CTCIHA9V.

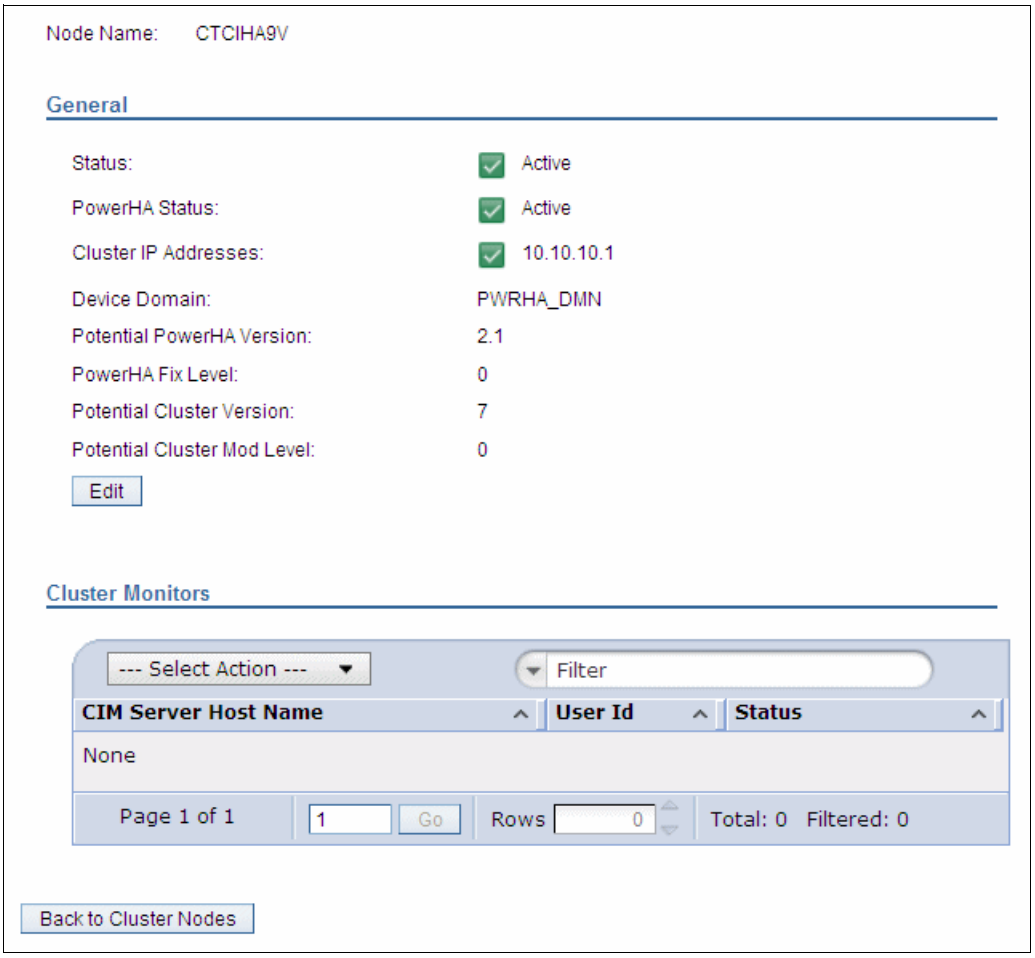


Figure 11-14 Cluster monitor not defined

10. From the Select Action pull-down menu, choose **Add Cluster Monitor** (Figure 11-15).

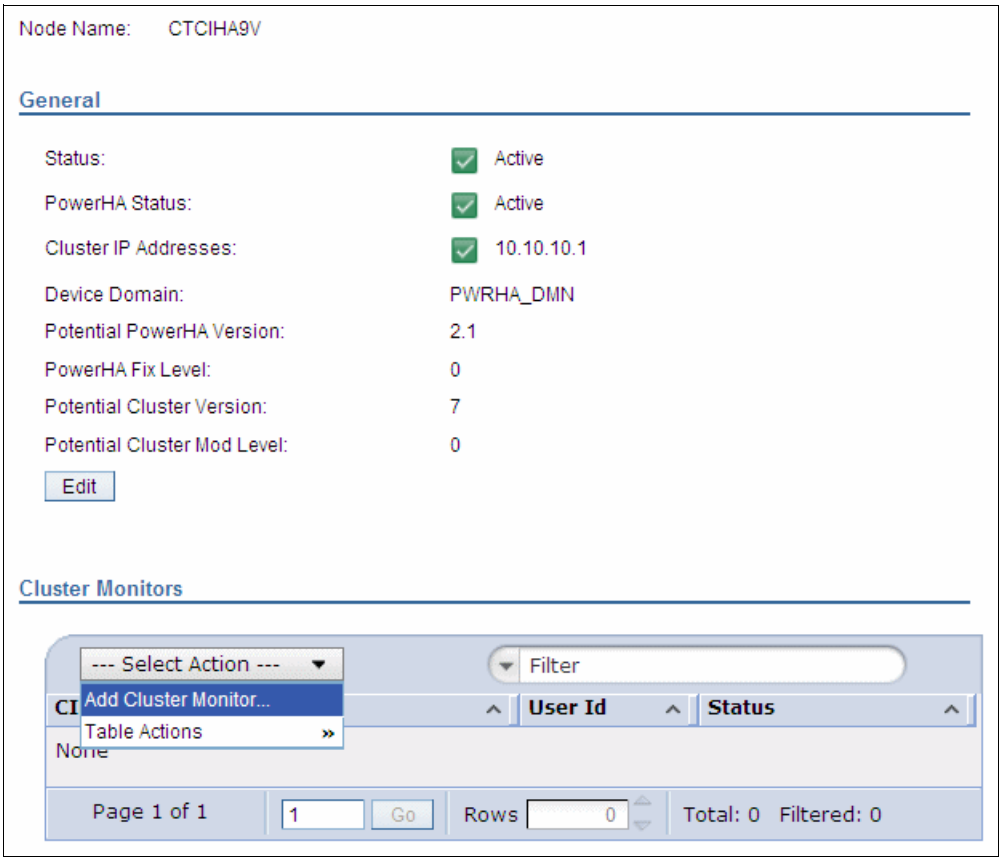


Figure 11-15 Add Cluster Monitor

11. Provide the information shown in Figure 11-16. The CIM server host name is the IP name of your HMC. The user ID and password are also those used to access the HMC.

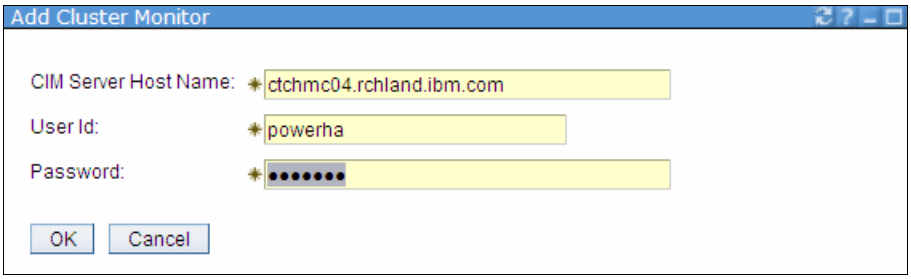


Figure 11-16 Provide cluster monitor information

In Figure 11-17 you can see that the cluster monitor was successfully added for node CTCIHA9V.

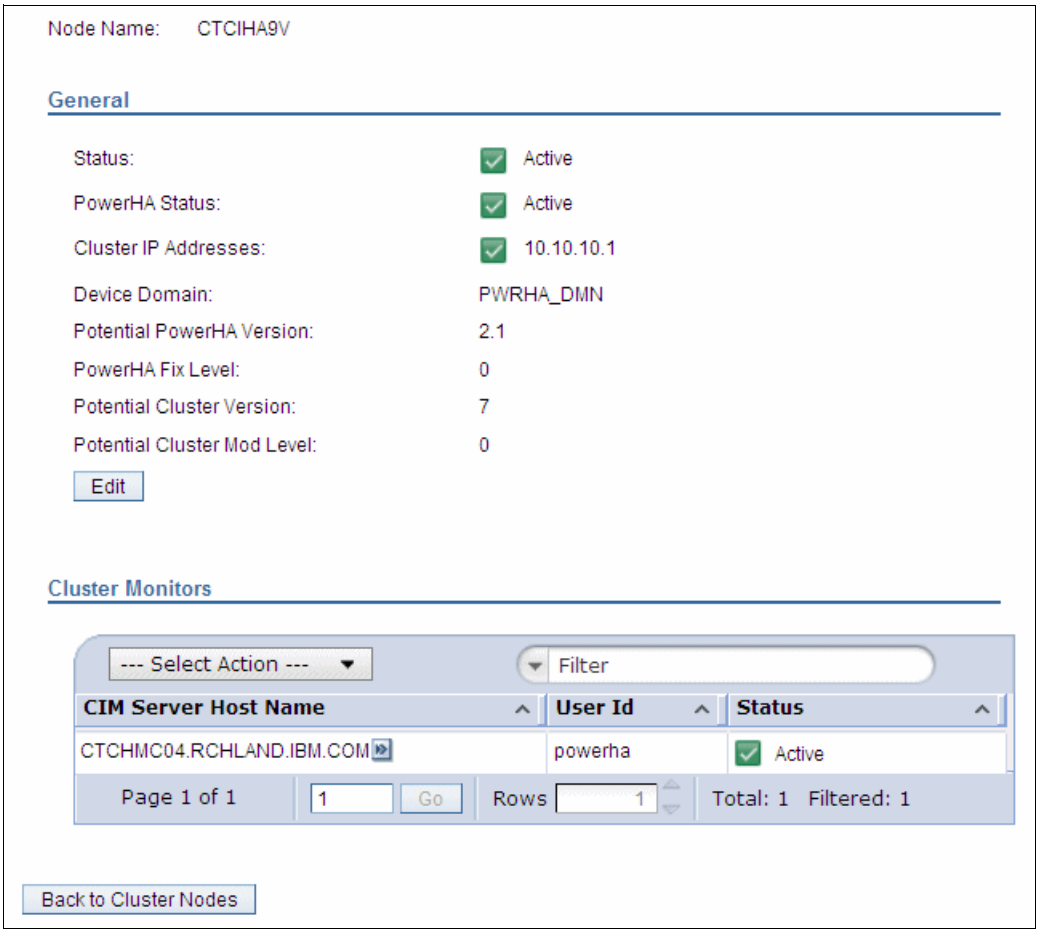


Figure 11-17 Cluster monitor added

12. Perform the same steps for all nodes in your cluster that should have a cluster monitor defined. In our example, we set up cluster monitors for both nodes (Figure 11-18).

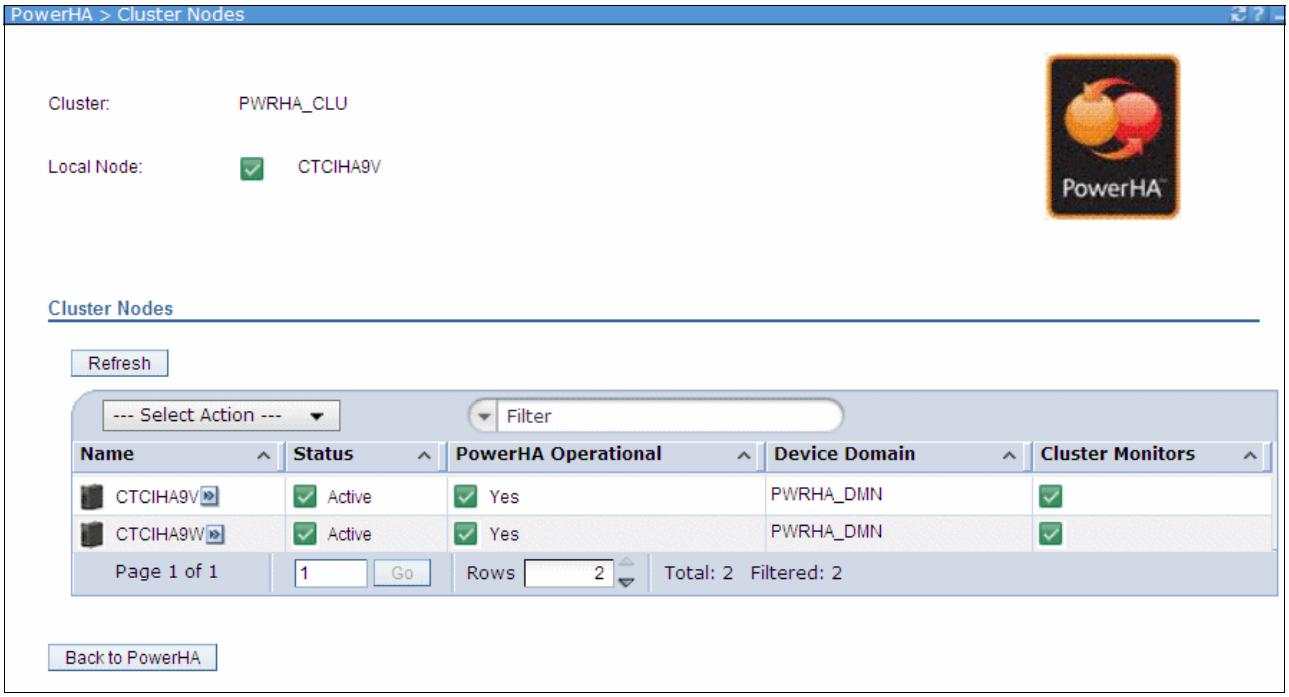


Figure 11-18 Cluster monitor added and active

The setup that we have done so far can also be done using the commands shown in Example 11-1.

Example 11-1 Command to create a 2-node cluster with cluster monitors

```
CRTCLU CLUSTER(PWRHA_CLU) NODE((CTCIHA9V ('10.10.10.1'))) CLUMSGQ(QSYS/QSYSOPR) FLVWAITTIM(*NOMAX)
ADDCLUNODE  CLUSTER(PWRHA_CLU) NODE(CTCIHA9W ('10.10.10.2'))
CHGCLUVER CLUSTER(PWRHA_CLU) HAVER(*UP1MOD)
ADDCLUMON CLUSTER(PWRHA_CLU) NODE(CTCIHA9V) CIMSVR(CTCHMC04.RCHLAND.IBM.COM hmc_user (hmc_password))
ADDCLUMON CLUSTER(PWRHA_CLU) NODE(CTCIHA9W) CIMSVR(CTCHMC04.RCHLAND.IBM.COM mhc_user (hmc_password))
```

11.3 Creating an IASP

From the page shown in Figure 11-19 you can proceed to create an independent ASP (IASP).

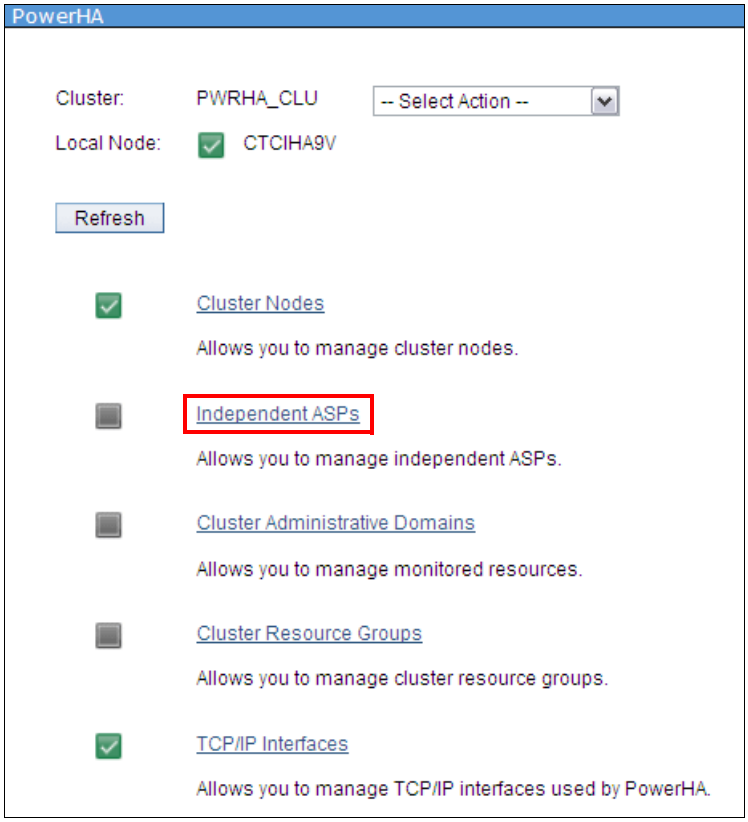


Figure 11-19 PowerHA GUI: Cluster overview

To proceed:

1. Figure 11-20 shows you a list of highly available IASPs. If you want to create a new IASP, you first have to click **Show all others**.

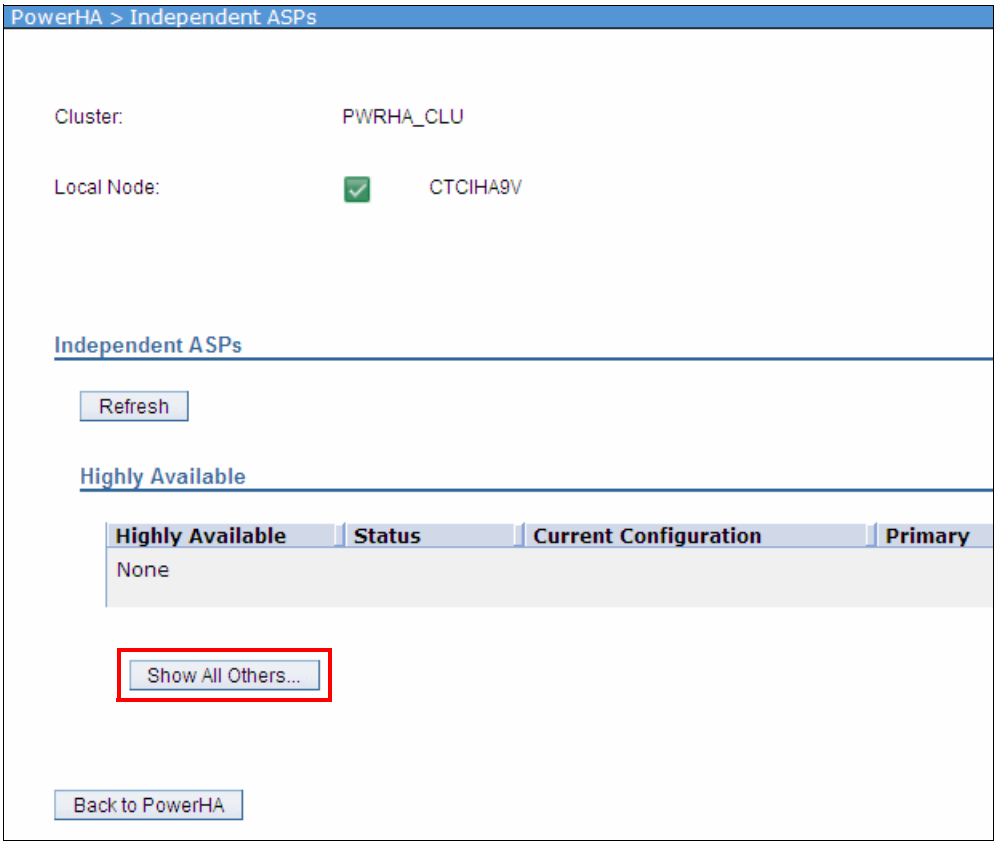


Figure 11-20 PowerHA GUI: IASP overview

2. From the Select Action pull-down menu you can select **Create Independent ASP** (Figure 11-21).

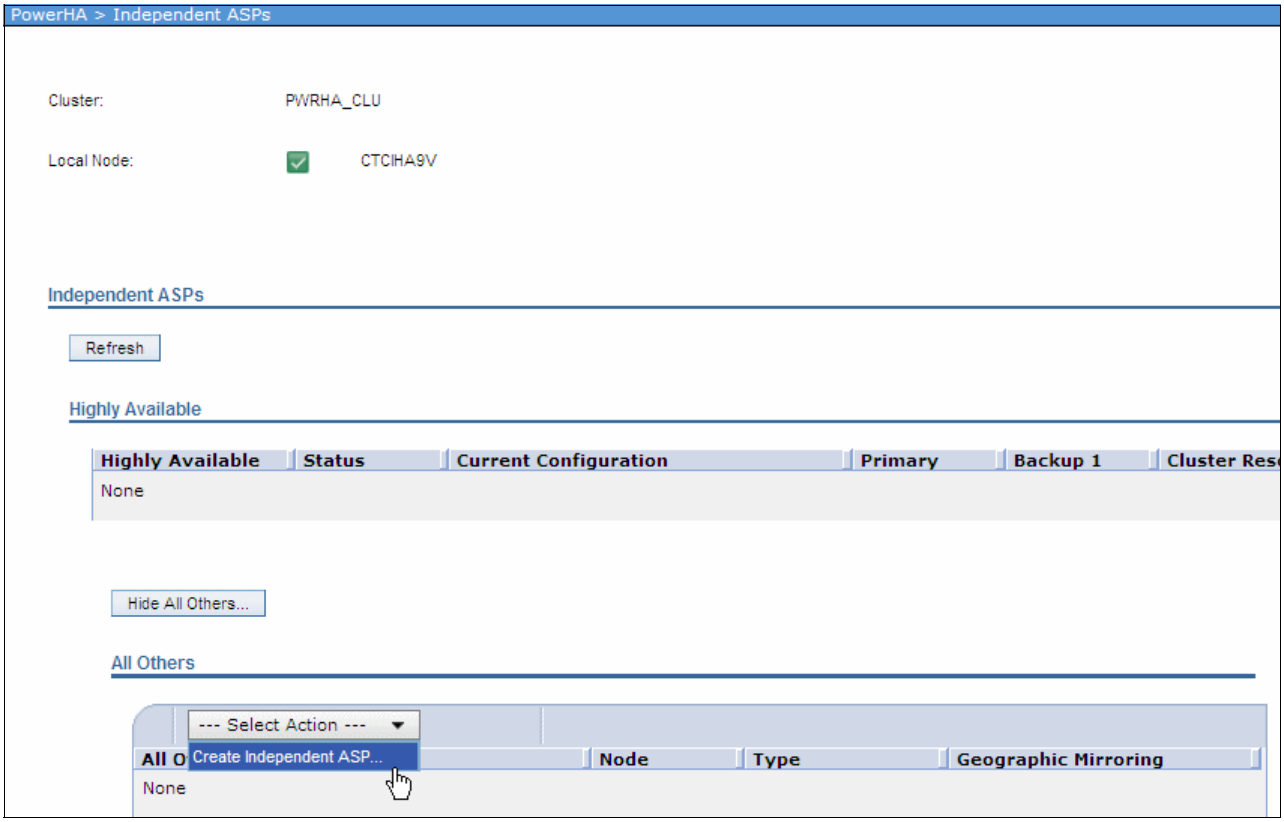


Figure 11-21 PowerHA GUI: Create independent ASP

3. A wizard guides you through the steps required to create an IASP. The Welcome page (Figure 11-22) gives you an overview of the information that you need to provide.

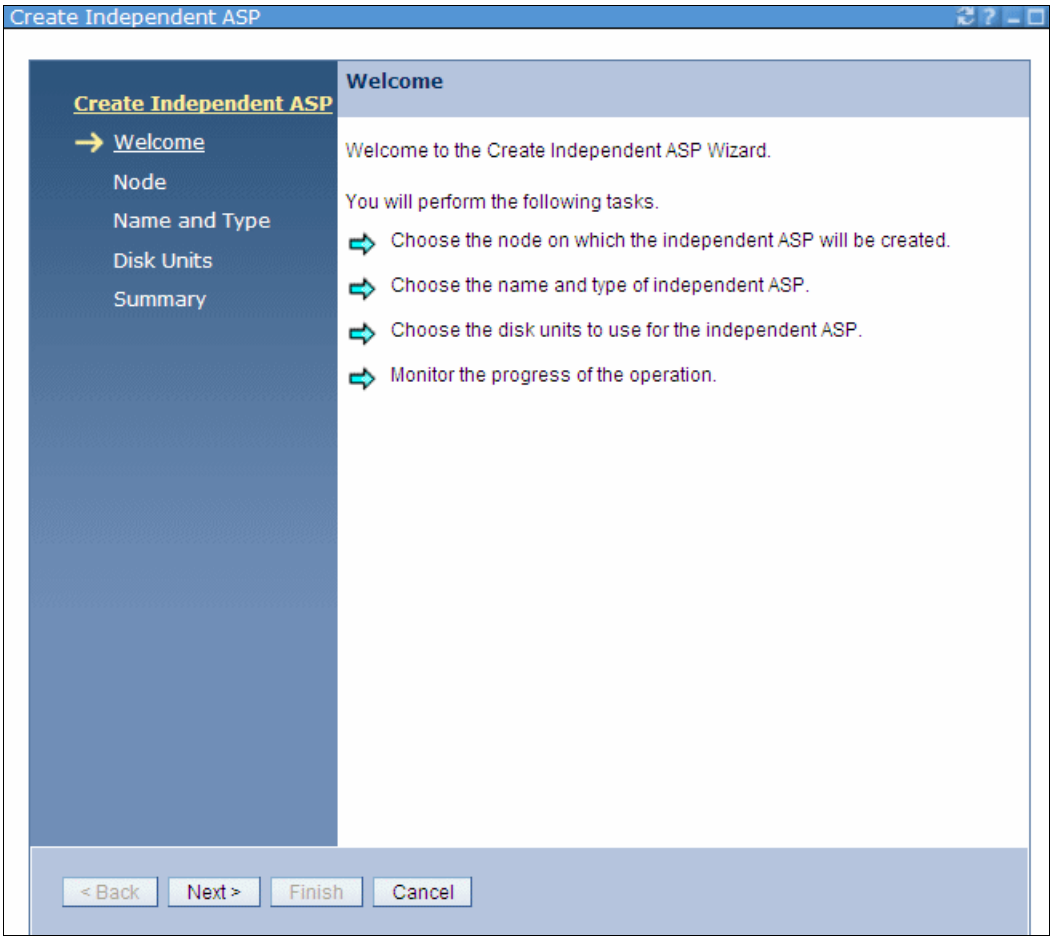


Figure 11-22 PowerHA GUI: IASP Welcome panel

4. In the first step you have to decide which node in the cluster you want to create the IASP. By default, this is the node on the system that you are connected to (Figure 11-23).

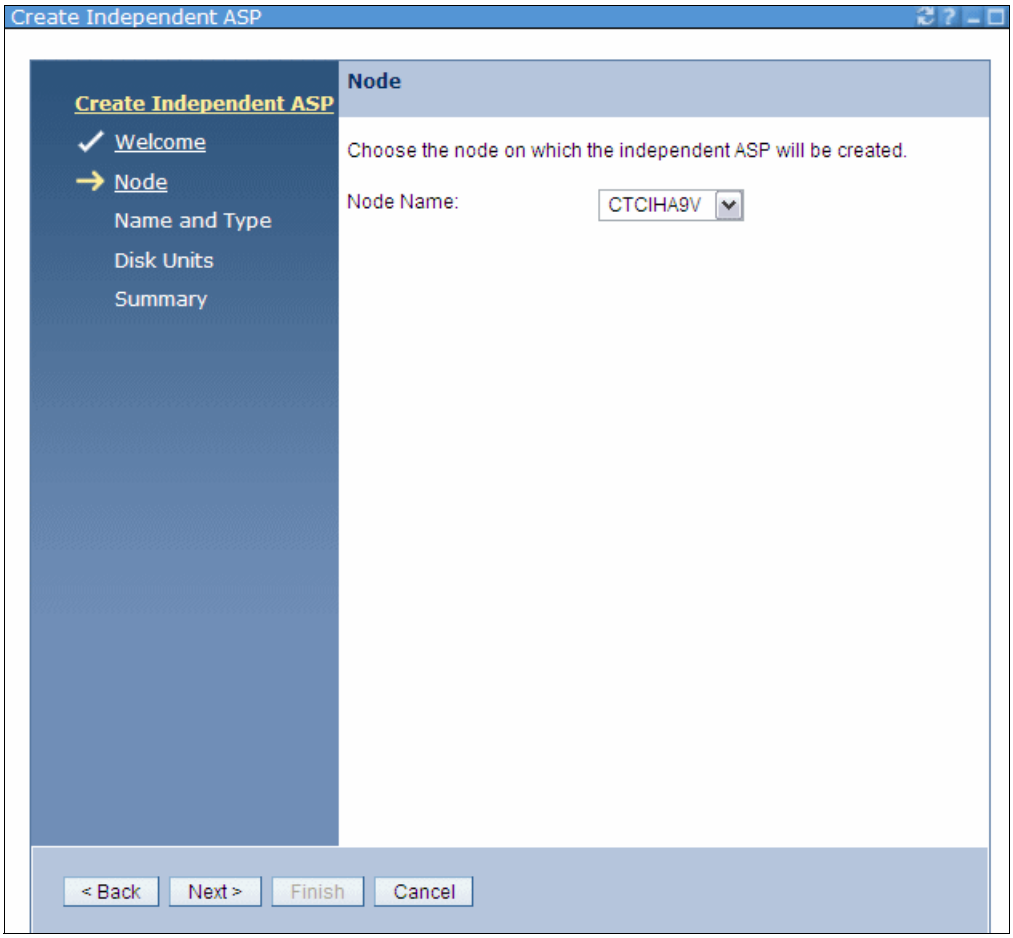


Figure 11-23 PowerHA GUI: Node selection

5. Provide a name for the IASP and decide whether it is a primary, secondary, or UDFS IASP. You can also encrypt the IASP (Figure 11-24). The Protected field allows you to specify whether you want to have the new IASP protected by either parity protection or IBM i mirroring. Depending on your selection for this field, only corresponding disk units with the required protection level will be shown. SVC/V7000 LUNs will always show up as unprotected disk units regardless of whether they are RAID protected by the storage system.

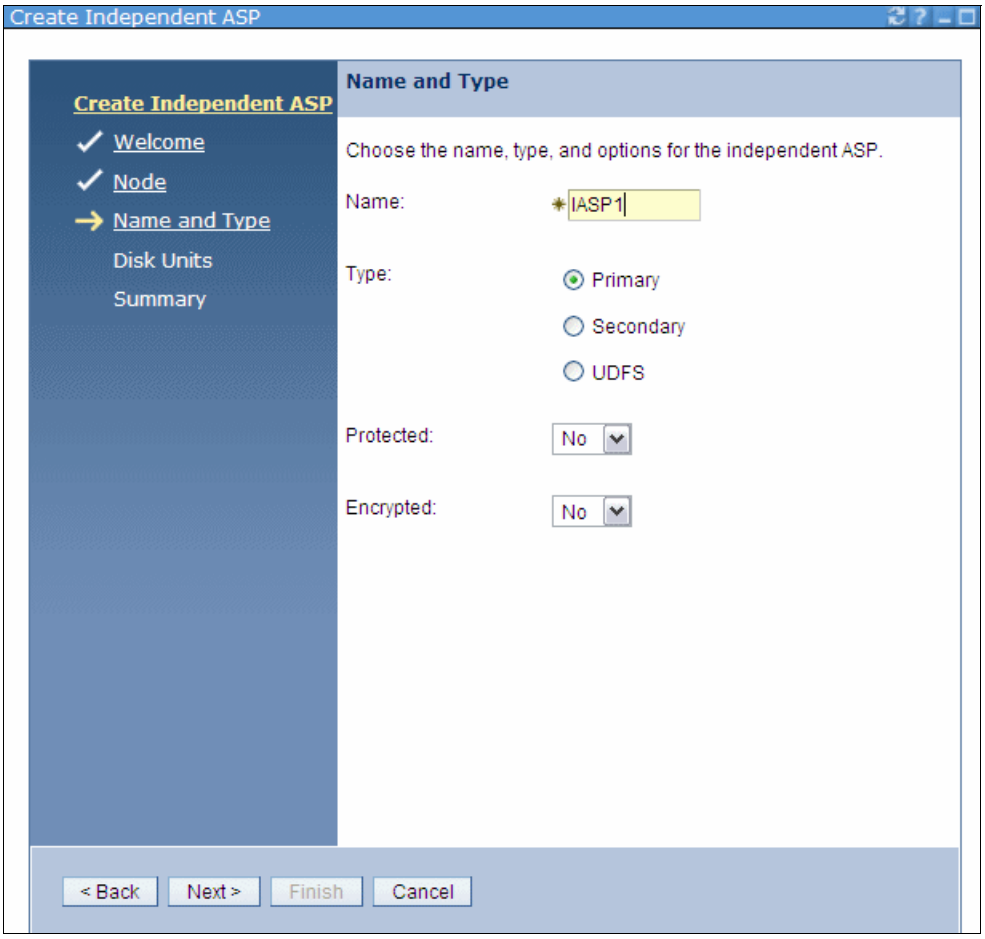


Figure 11-24 PowerHA GUI: IASP name and type

6. The wizard then shows you a list of unconfigured disks on your system. Notice that to create an IASP with the new PowerHA GUI you do not have to provide an SST password or have an SST user ID with the same name and password as your normal IBM i user ID anymore. Choose the disks that you want to add into your IASP and click **Add** (Figure 11-25).

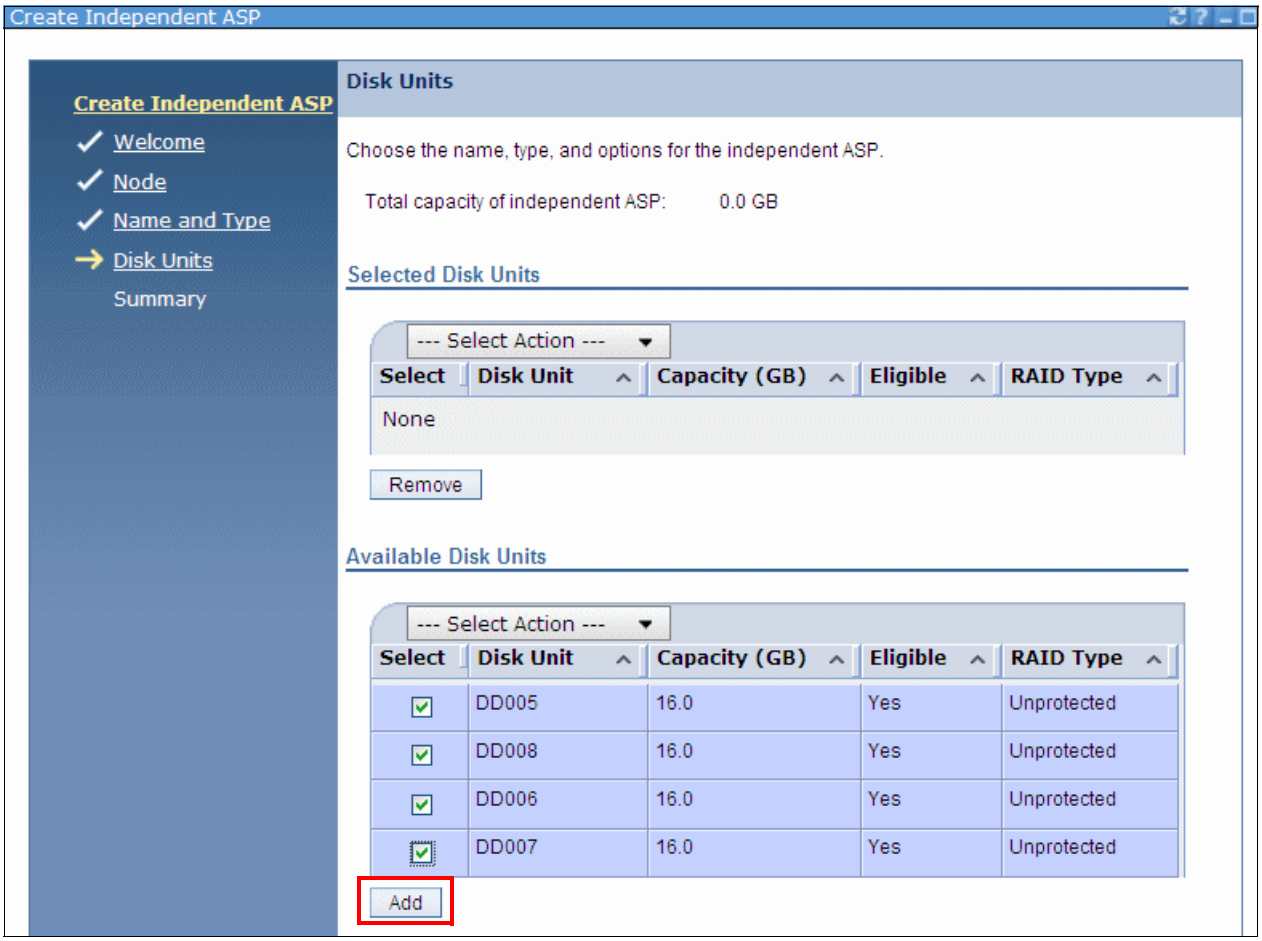


Figure 11-25 PowerHA GUI: Select disks for IASP

7. The wizard updates the information shown with your choice (Figure 11-26). You can still remove individual disks from your selection.

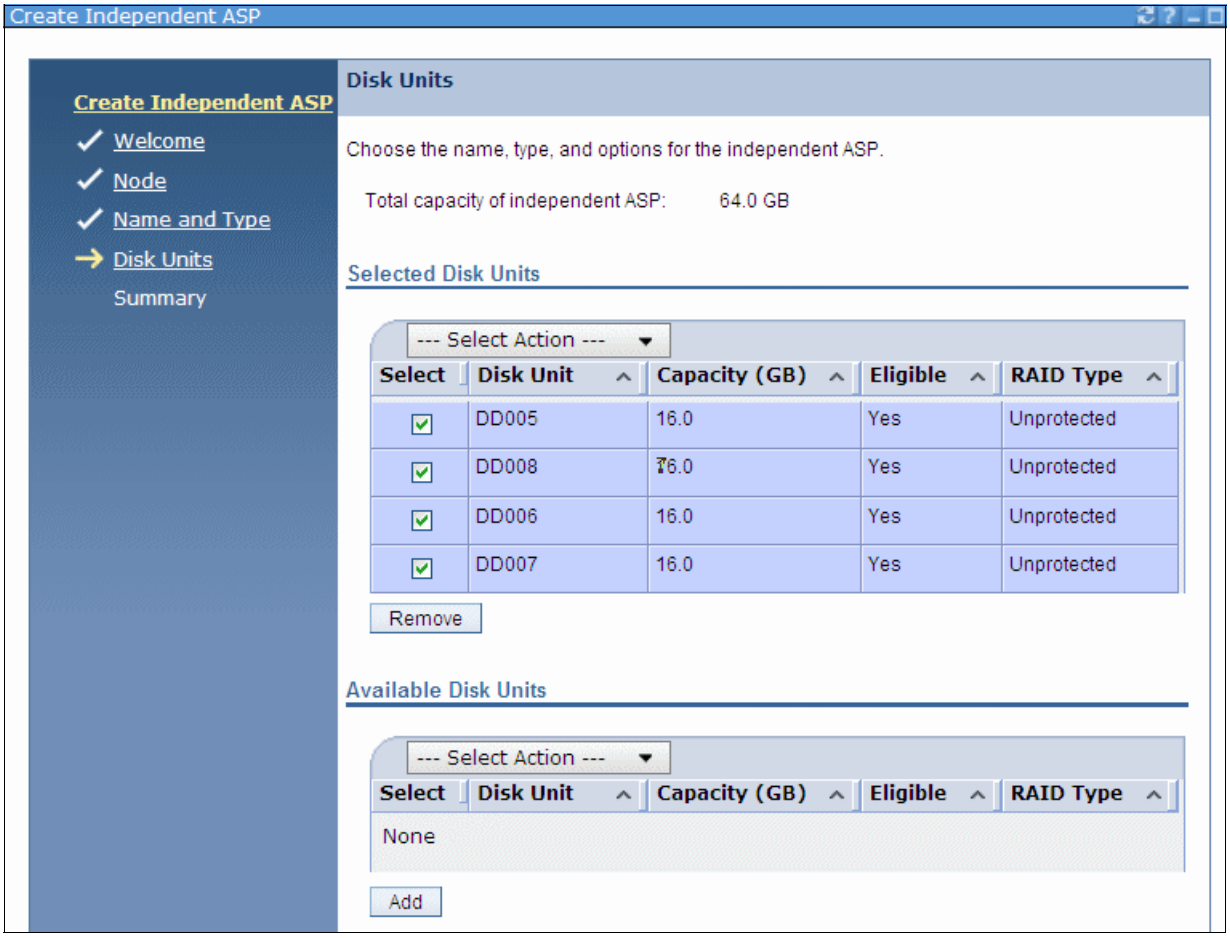


Figure 11-26 PowerHA GUI: Disks selected for IASP

8. The Summary page (Figure 11-27) provides an overview of the configuration settings that you have chosen. Click **Finish** to create the IASP.

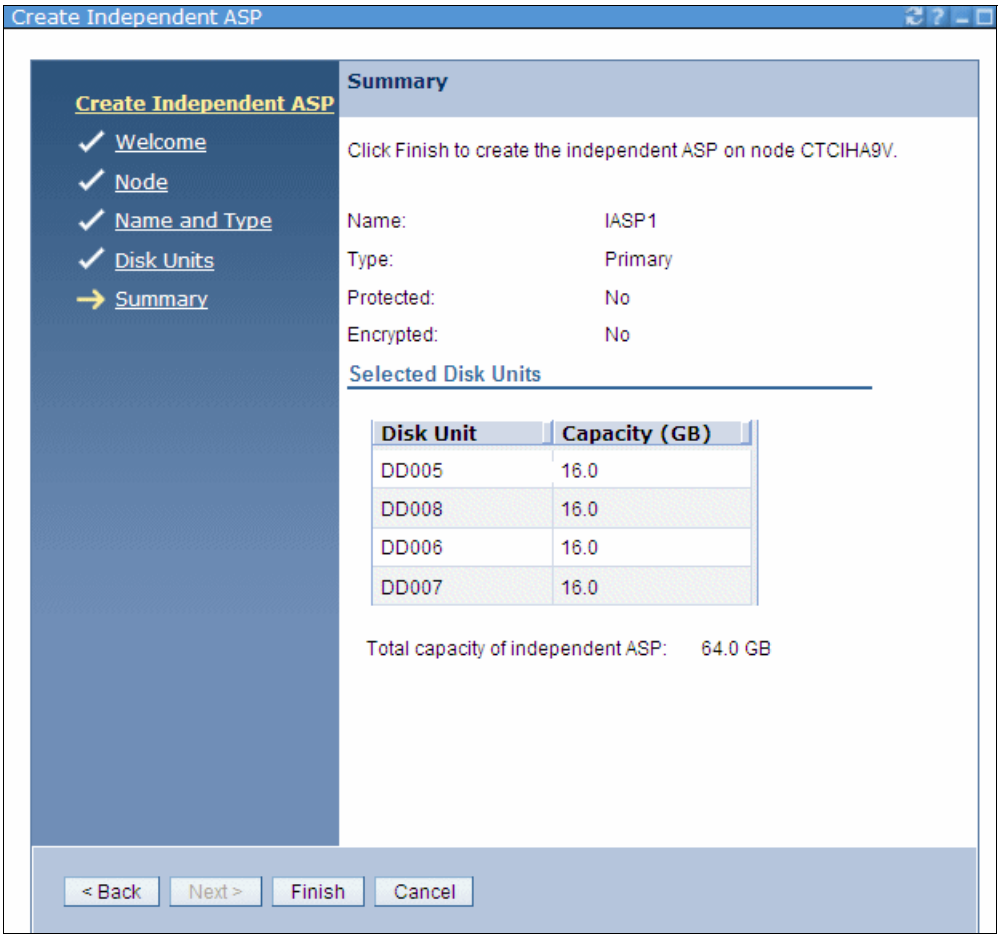


Figure 11-27 PowerHA GUI: IASP creation summary

9. The IASP is created. The wizard regularly updates the completion status (Figure 11-28).

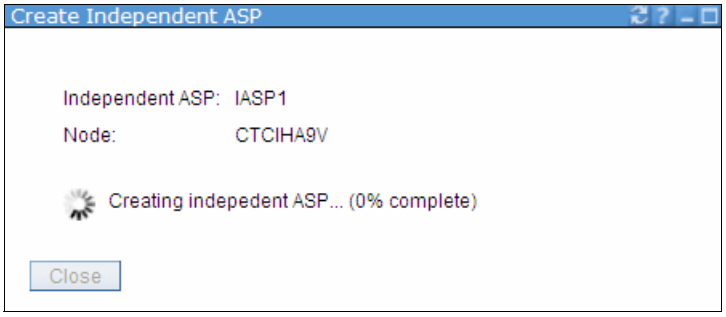


Figure 11-28 PowerHA GUI: IASP is created

Alternatively, you can use a CL command to create your iASP. Figure 11-29 shows the required parameters for **CFGDEVASP**.

```
                    Configure Device ASP (CFGDEVASP)

Type choices, press Enter.

ASP device . . . . . > IASP1          Name
Action . . . . . > *CREATE        *CREATE, *DELETE
ASP type . . . . . .   *PRIMARY       *PRIMARY, *SECONDARY, *UDFS
Protection . . . . . .   *NO           *NO, *YES
Encryption . . . . . .   *NO           *NO, *YES
Disk units . . . . . .   *SELECT       Name, *SELECT
                                                  + for more values




                                                                  Bottom
F1=Help   F9=Calculate Selection   F11=View 2   F12=Cancel
```

Figure 11-29 CFGDEVASP to create IASP

Specifying *SELECT for the disk unit parameter shows the panel shown in Figure 11-30. It provides you with a list of unconfigured disks on your system. Choose which ones you want to add to your IASP and press Enter to create the IASP.

```
                    Select Non-Configured Disk Units

ASP device . . . . . :   IASP1
Selected capacity  . . . . . :   0
Selected disk units  . . . . . :   0

Type options, press Enter.
  1=Select

     Resource
Opt  Name        Serial Number    Type Model  Capacity  Rank  Eligible
 1   DD007       YQKJGD54BUK6     6B22 0050      19088  002      Yes
 1   DD006       YDP4V2FVUK63     6B22 0050      19088  002      Yes
 1   DD008       YUNHA7W9URJL     6B22 0050      19088  002      Yes
 1   DD005       YWPZGH6N8LA9     6B22 0050      19088  002      Yes



                                                                  Bottom
F1=Help   F9=Calculate Selection   F11=View 2   F12=Cancel
Configuration of ASP device IASP1 is 8% complete.
```

Figure 11-30 CFGDEVASP: Select disks to put into IASP

A message on the bottom of the page shows the progress of the IASP creation. Your age is locked as long as the creation of the IASP is running.

11.4 Setting up an administrative domain

An administrative domain can be used to synchronize a number of object types that cannot reside in an IASP between cluster nodes. The following steps can be used to set up an administrative domain and add monitored resource entries into it:

1. Start from the main menu of the PowerHA GUI by choosing **Cluster Administrative Domain** (Figure 11-31).

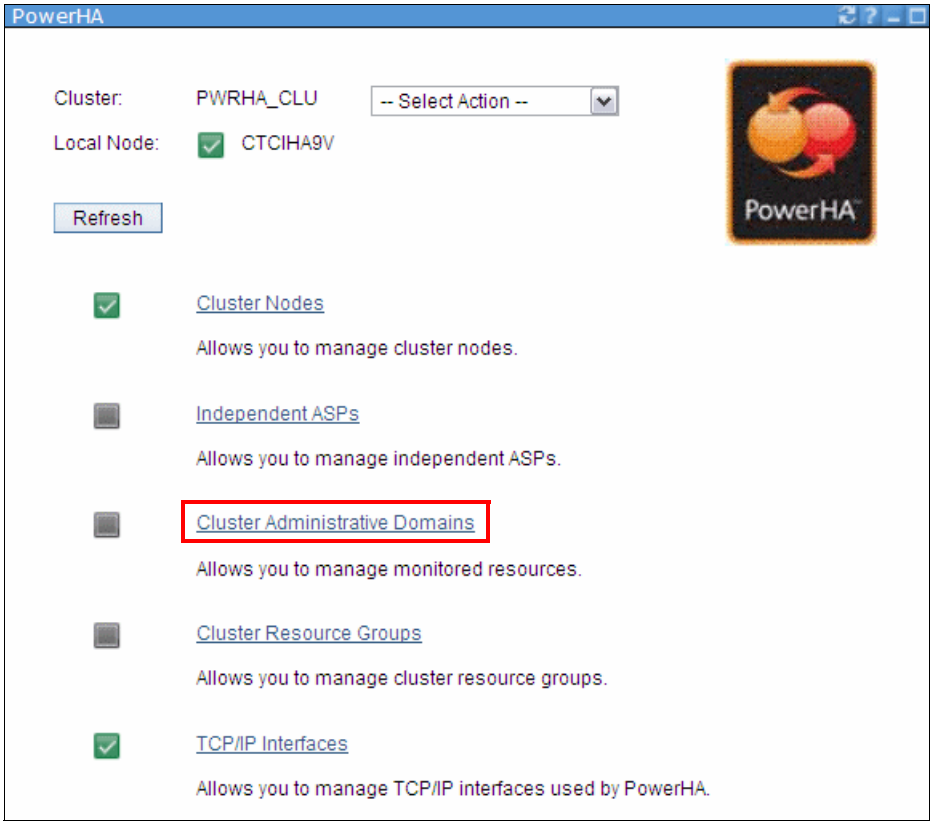


Figure 11-31 Cluster administrative domain setup

As you can see in Figure 11-32, we currently have not set up an administrative domain.

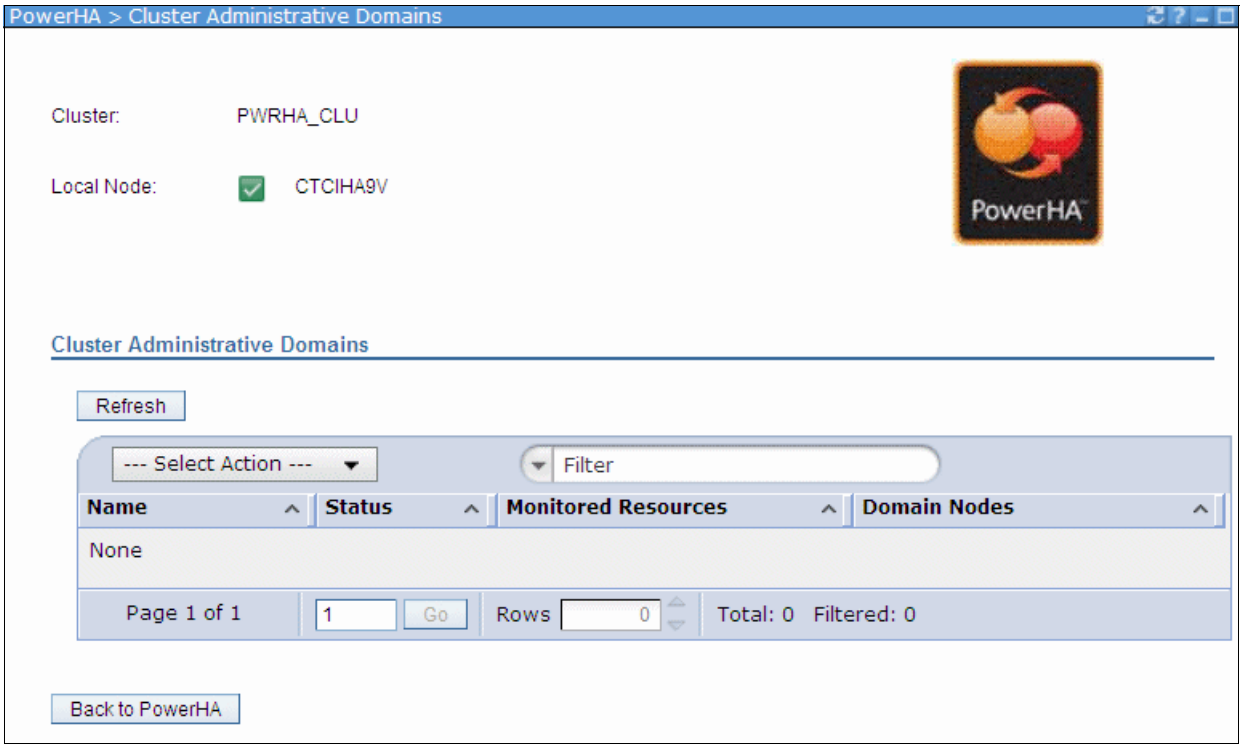


Figure 11-32 Cluster administrative domain not configured

2. To start the setup process choose **Create Cluster Administrative Domain** from the pull-down menu (Figure 11-33).

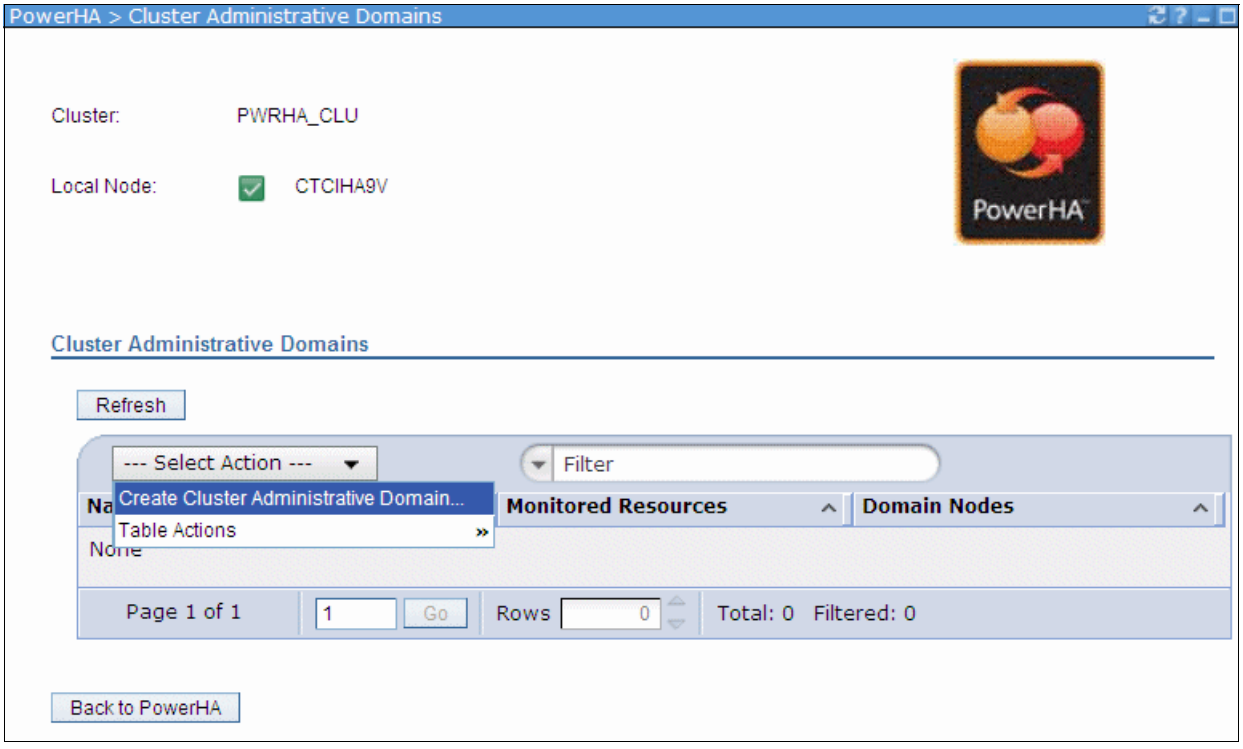


Figure 11-33 Create Cluster Administrative Domain

3. Provide a name for the administrative domain and decide on the synchronization option. This can be either *last change* or *active domain*. For a detailed description of these options see “Monitored resources” on page 56. Add the nodes that you want to be part of the administrative domain by choosing them from the list of available nodes and adding them to the list of selected nodes (Figure 11-34).

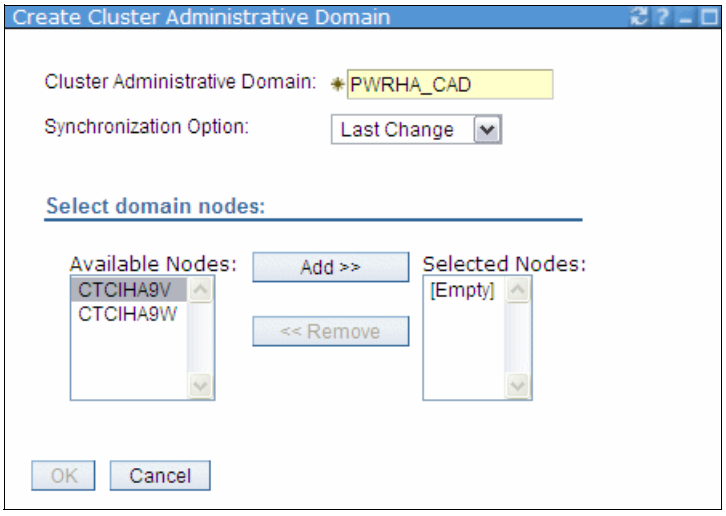


Figure 11-34 Cluster Administrative Domain: Add nodes

4. After you have selected all nodes to be part of the administrative domain click **OK** (Figure 11-35).

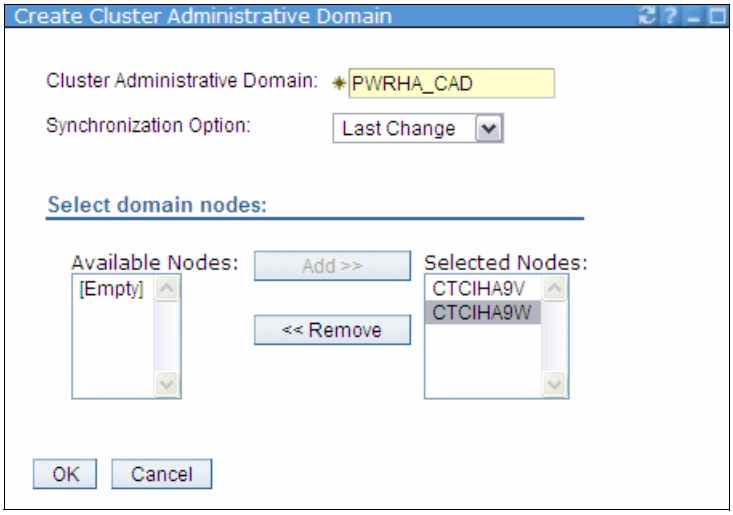


Figure 11-35 Cluster Administrative Domain: All nodes selected

The administrative domain is created but not yet started (Figure 11-36).

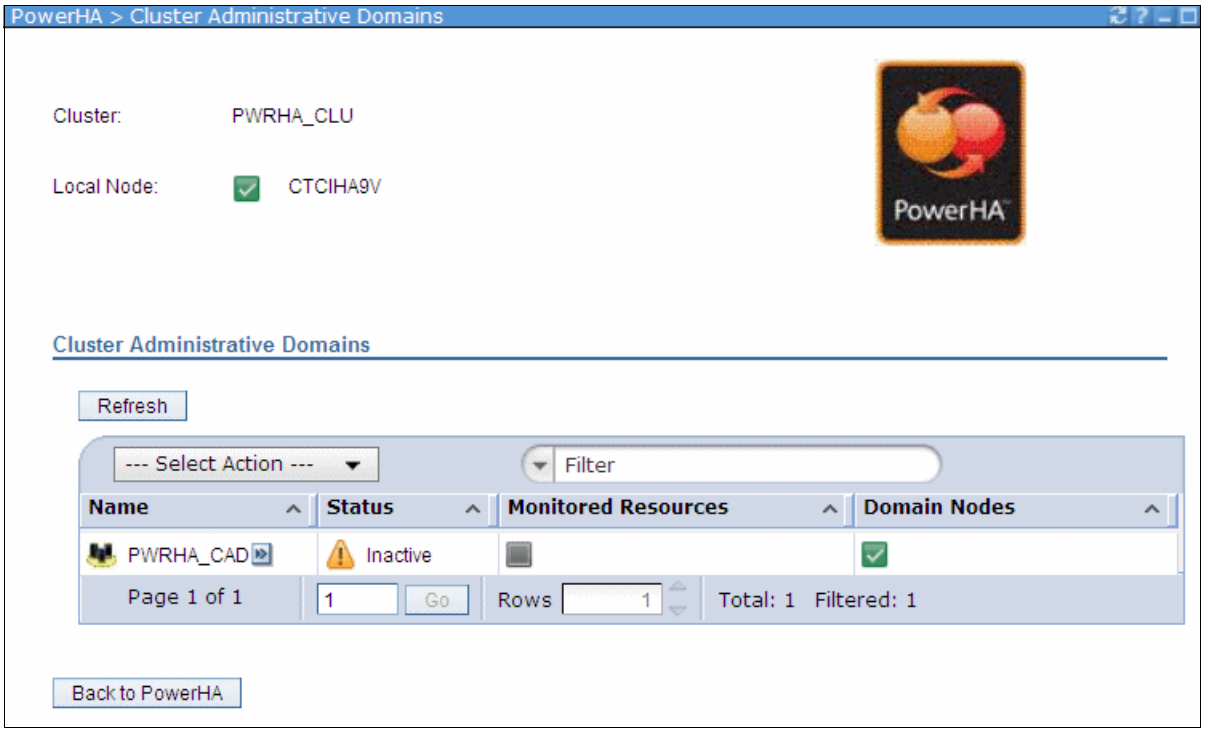


Figure 11-36 Cluster administrative domain created

5. Choose **Start** (Figure 11-37).

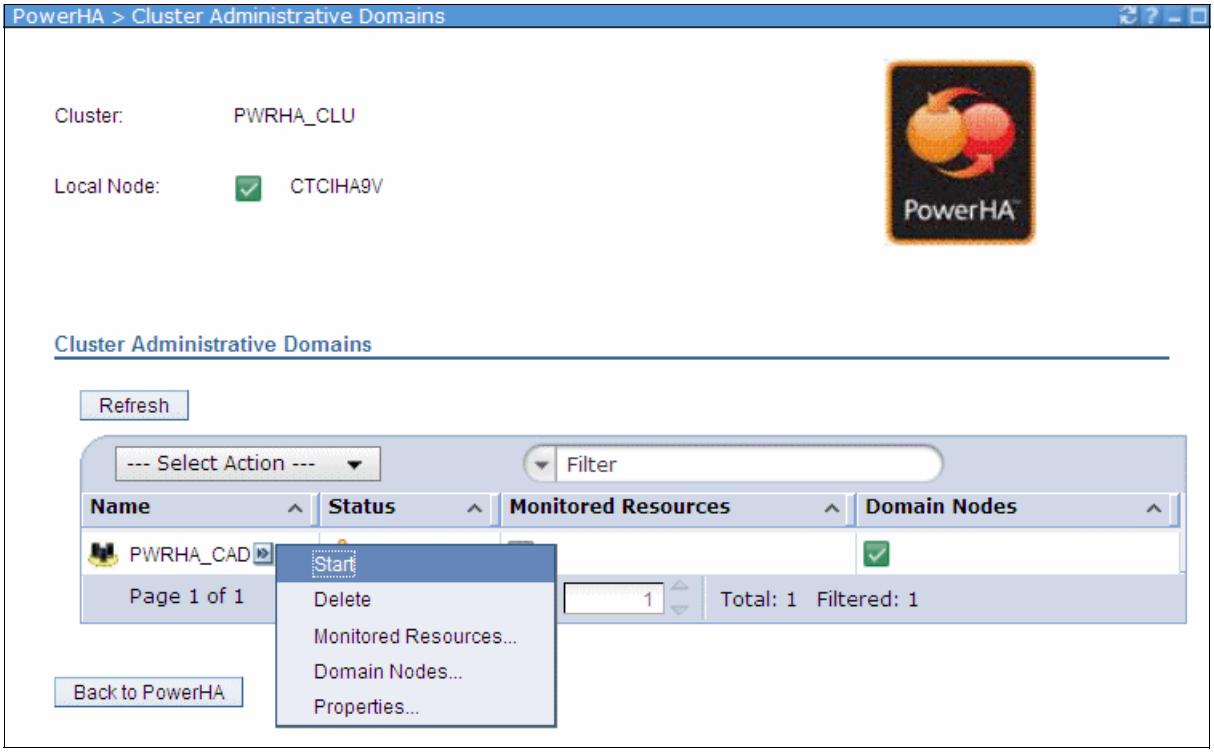


Figure 11-37 Start cluster administrative domain

The administrative domain becomes active (Figure 11-38). There are currently no monitored resource entries defined.

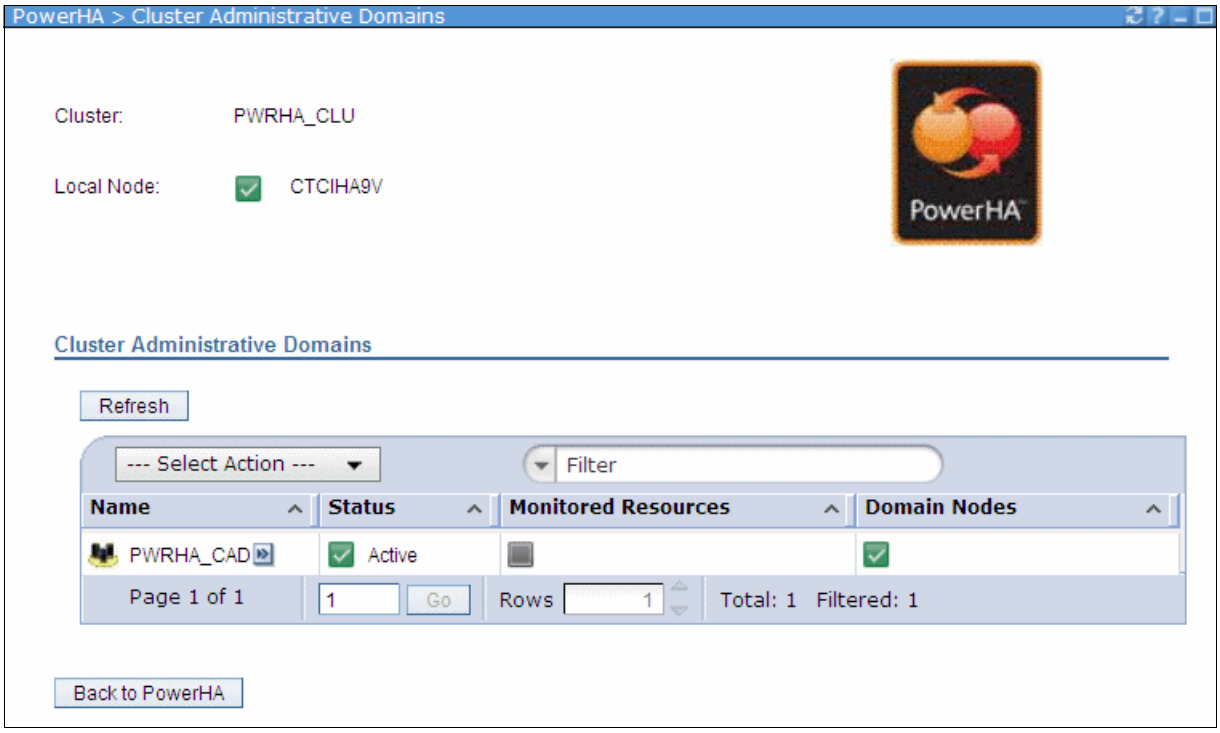


Figure 11-38 Cluster administrative domain active

6. Choose **Monitored Resources** (Figure 11-39).

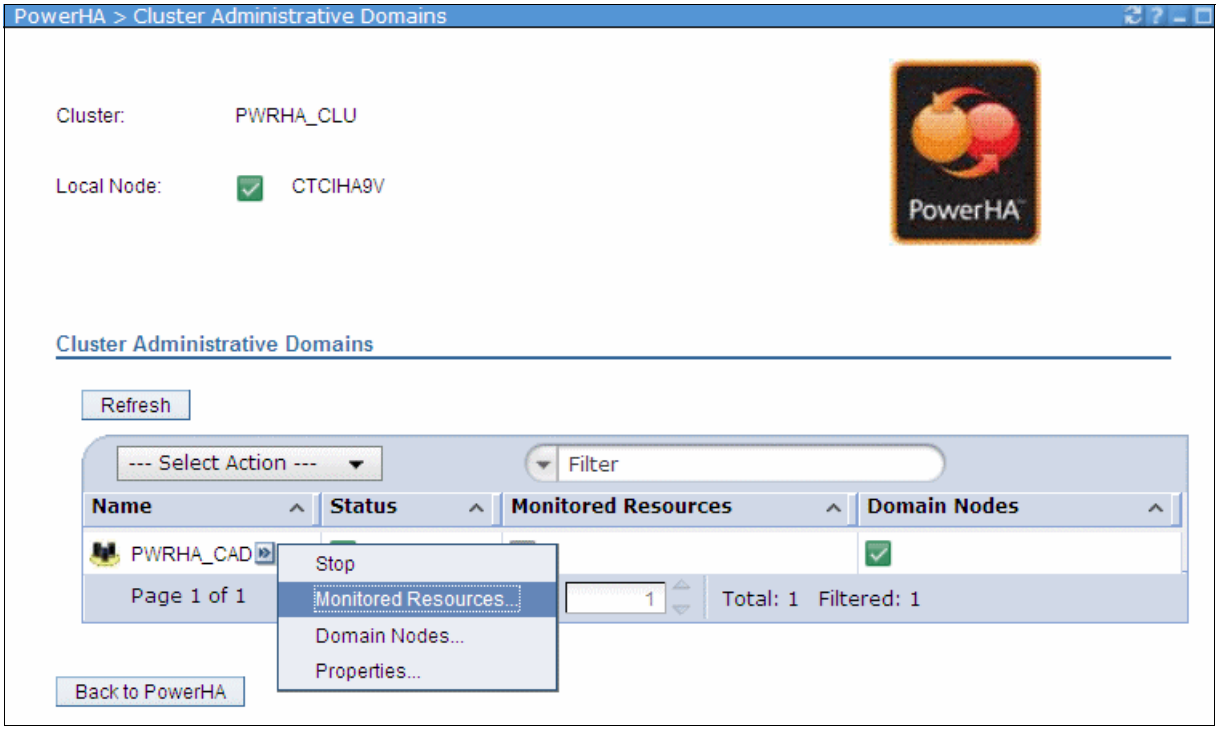


Figure 11-39 Cluster administrative domain: Check monitored resources

7. Choose **Add Monitored Resources** from the pull-down menu (Figure 11-40).

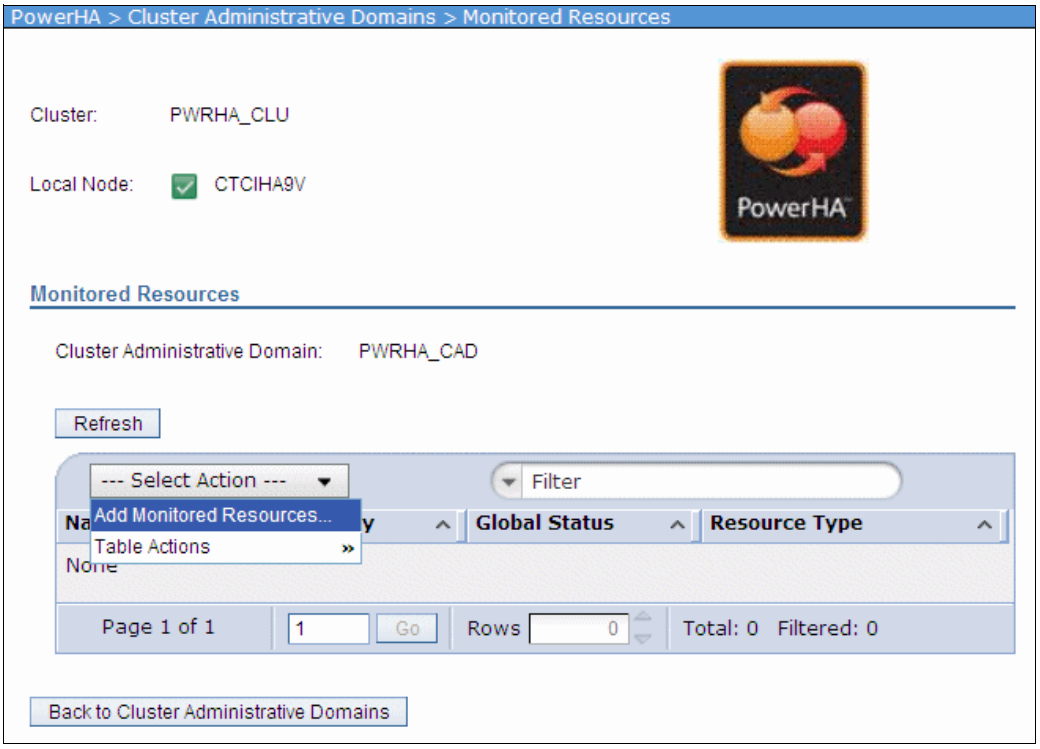


Figure 11-40 Cluster administrative domain: Add MRE

8. On the page shown in Figure 11-41 we add the subsystem QBATCH as a monitored resource to the administrative domain. CTCIHA9V is the system that is used as a source for the first synchronization. Be aware that, in general, there is no leading system in an administrative domain. Changes on any node in the administrative domain are propagated to all nodes in the domain. To add resources, you can either type their name or select them from the corresponding list. You can also decide whether you want to synchronize all attributes of a specific MRE or just subsets of attributes.

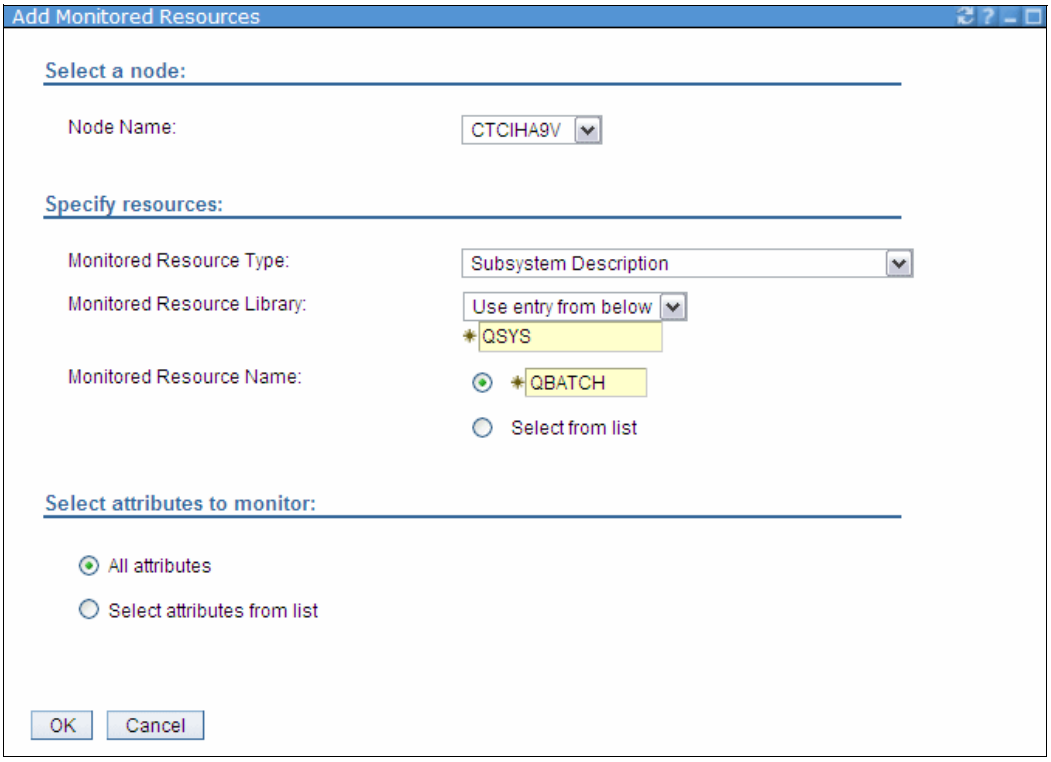


Figure 11-41 Cluster administrative domain: Add MRE for subsystem description

9. When using the list functions, you can also add several MREs of one object type into the administrative domain with one configuration step (Figure 11-42).

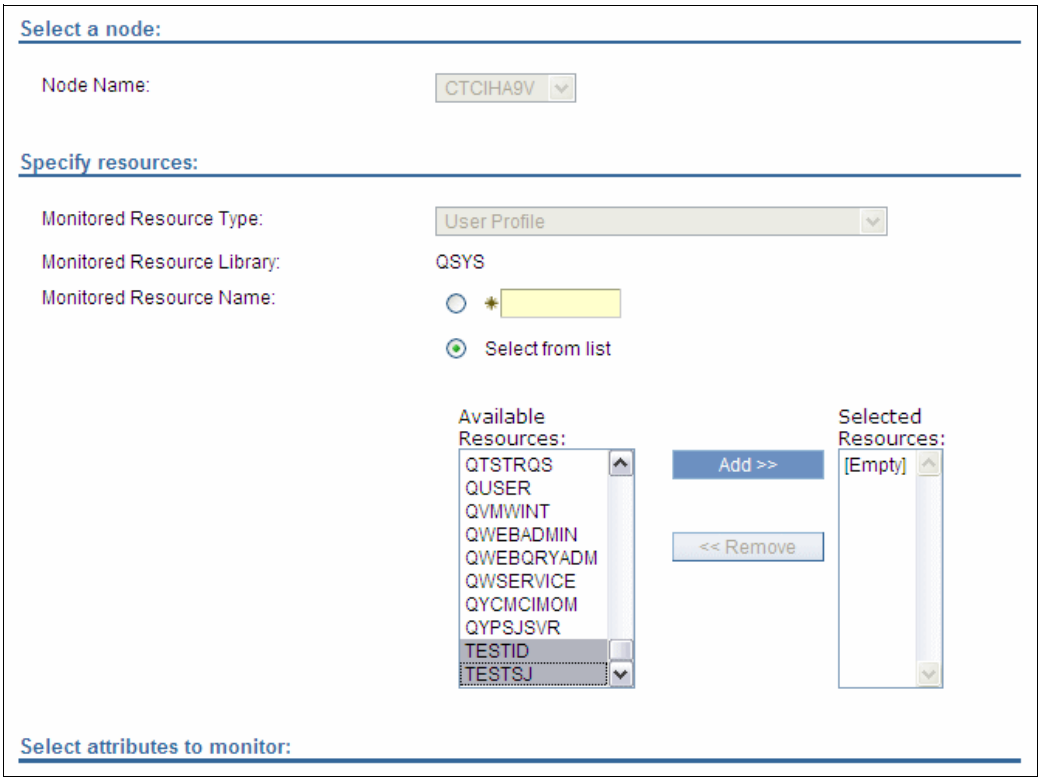


Figure 11-42 Cluster administrative domain: Add MRE for user profile

10. The monitored resource entries are added to the administrative domain and a first synchronization occurs (Figure 11-43). If the MREs do not exist on any of the other nodes in the administrative domain, they are automatically created.

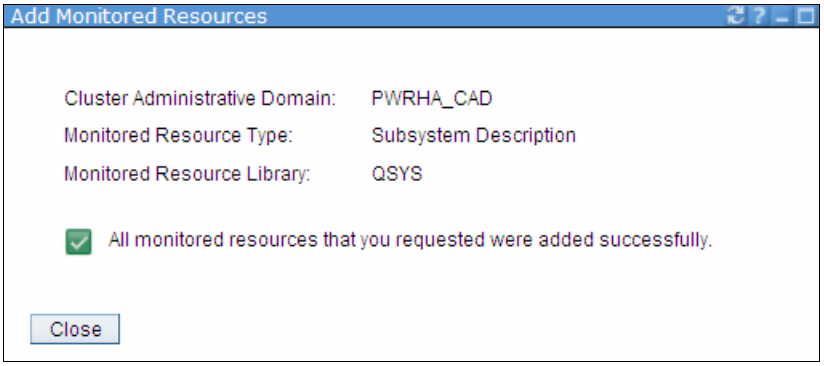


Figure 11-43 Cluster administrative domain: MRE successfully added

11. Figure 11-44 shows that the added MREs are in a consistent status within the administrative domain.

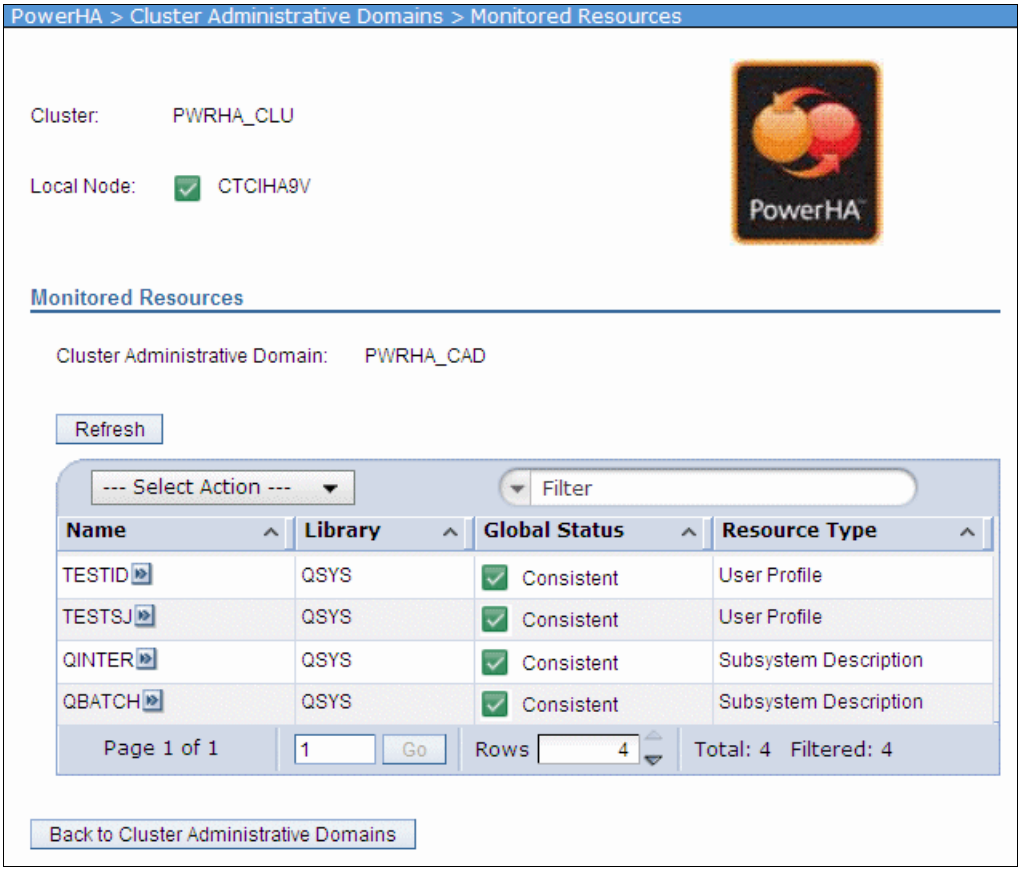


Figure 11-44 Cluster administrative domain: MRE overview

The setup that we have done so far can also be done using the commands shown in Example 11-2. Notice that you have to use individual commands for each monitored resource that you want to add to the administrative domain.

Example 11-2 Command to create an administrative domain and add monitored resource entries

```
CRTCAD CLUSTER(PWRHA_CLU) ADMDMN(PWRHA_CAD) DMNNODL(CTCIHA9V CTCIHA9W)
SYNCOPT(*LASTCHG)

ADDCADMRE CLUSTER(PWRHA_CLU) ADMDMN(PWRHA_CAD) RESOURCE(QBATCH) RSCTYPE(*SBSD)
RSCLIB(QSYS)
ADDCADMRE CLUSTER(PWRHA_CLU) ADMDMN(PWRHA_CAD) RESOURCE(TESTSJ) RSCTYPE(*USRPRF)
```

12

Configuring and managing Geographic Mirroring

This chapter describes a scenario that uses the new PowerHA GUI for setting up and managing your geographic mirroring high-availability environment.

A geographic mirroring solution requires a cluster with a minimum of two cluster nodes, an IASP, and optionally cluster monitors and an administrative domain. These components make up the base environment for a geographic mirroring solution. The steps are discussed in Chapter 11, “Creating a PowerHA base environment” on page 199.

12.1 Setting up geographic mirroring

We set up geographic mirroring for an existing independent ASP that was created using the steps in 11.3, “Creating an IASP” on page 215.

Geographic mirroring is set up using the Make Highly Available wizard, which is found on PowerHA GUI → Independent ASPs → Show All Others → pop menu of the iASP → Make Highly Available (Figure 12-1).

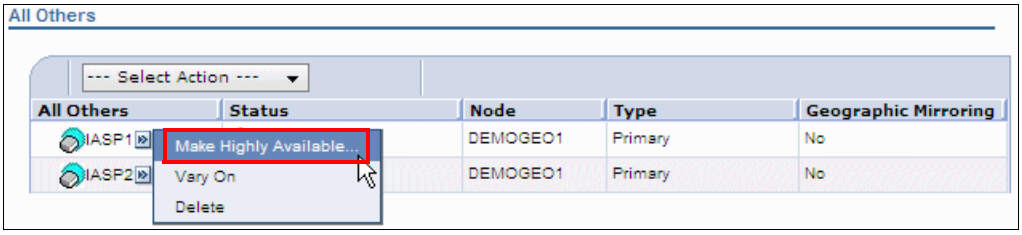


Figure 12-1 Make Highly Available

Follow the steps below to set up geographic mirroring:

1. When you choose the **Make Highly Available** option, the Welcome screen of the wizard displays (Figure 12-2). This screen page you an overview of the steps involved in setting up the environment. Click **Next**.

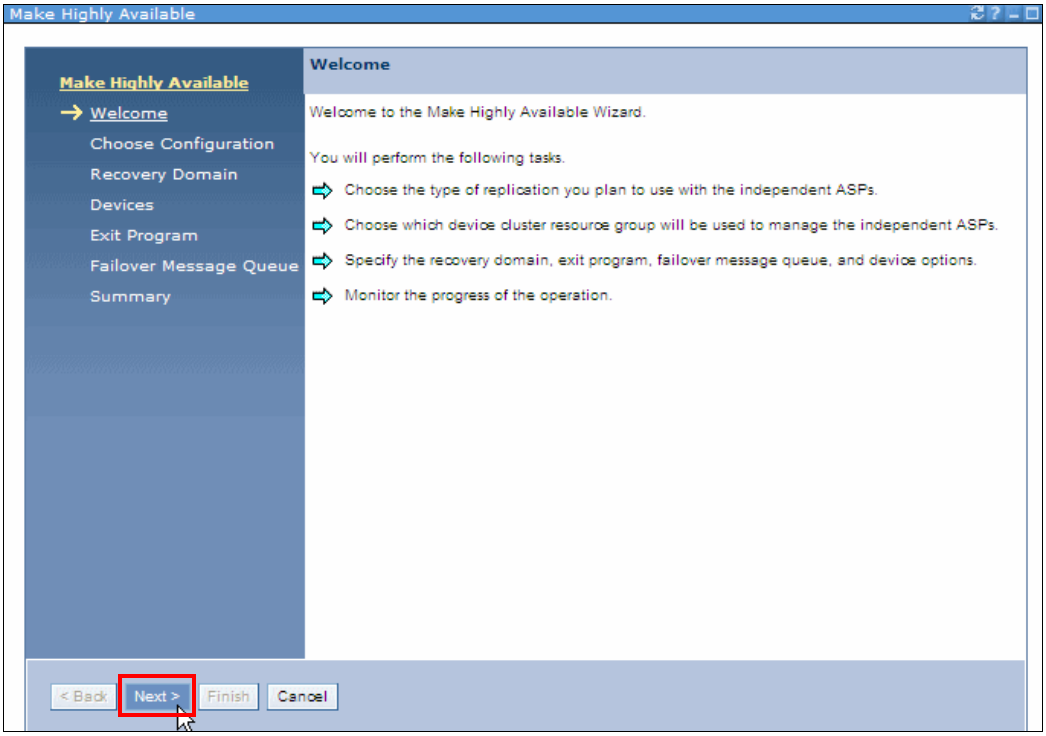


Figure 12-2 Make Highly Available Wizard’s Welcome panel

2. From the Choose Configuration page, click the **Geographic Mirroring** radio button (Figure 12-3).

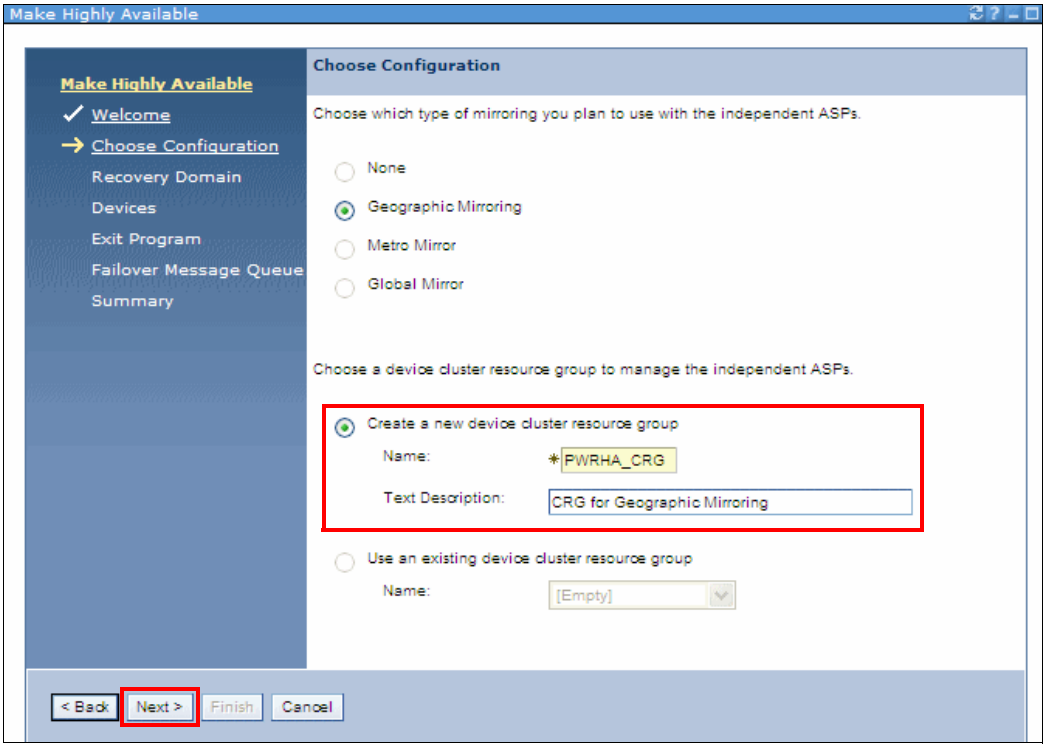


Figure 12-3 Choose Configuration

3. The IASP must be part of a device cluster resource group (CRG) for setting up geographic mirroring. On the Configuration page, you can either create a new device CRG or use an existing one. For our scenario, we choose to create a new device CRG named PWRHA_CRG. Click **Next** to work with the recovery domain.

4. From the Recovery Domain page, you can add the cluster nodes and specify their roles and the data port IP addresses (Figure 12-4). First we must select a node for the primary role and assign a site name for identification. PowerHA allows you to choose up to four data port IP addresses to ensure communication redundancy.

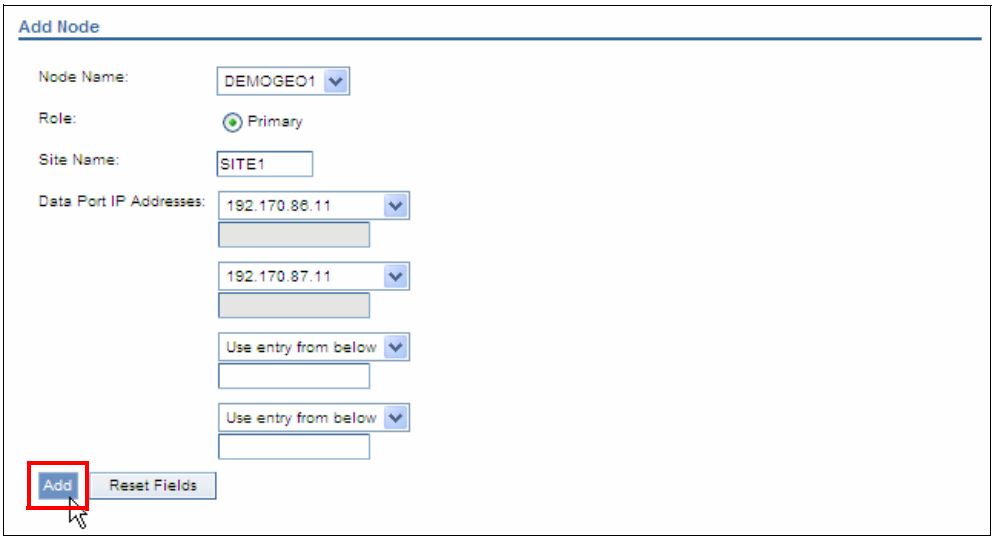


Figure 12-4 Add a primary node in the recovery domain

5. Add backup node 1 and specify a name for the site and up to four data port IP addresses (Figure 12-5).

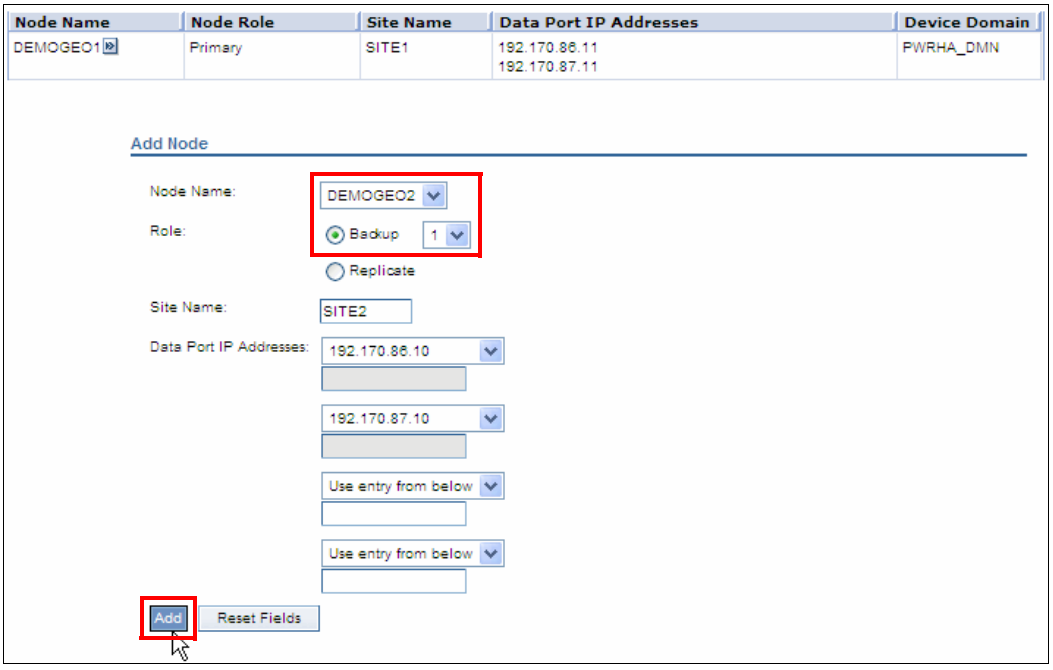


Figure 12-5 Add backup node to the recovery domain

6. The Recovery Domain page shows a summary of the node configuration options. From this page, you can either click a node to modify the properties or click **Next** to continue with the setup wizard (Figure 12-6).

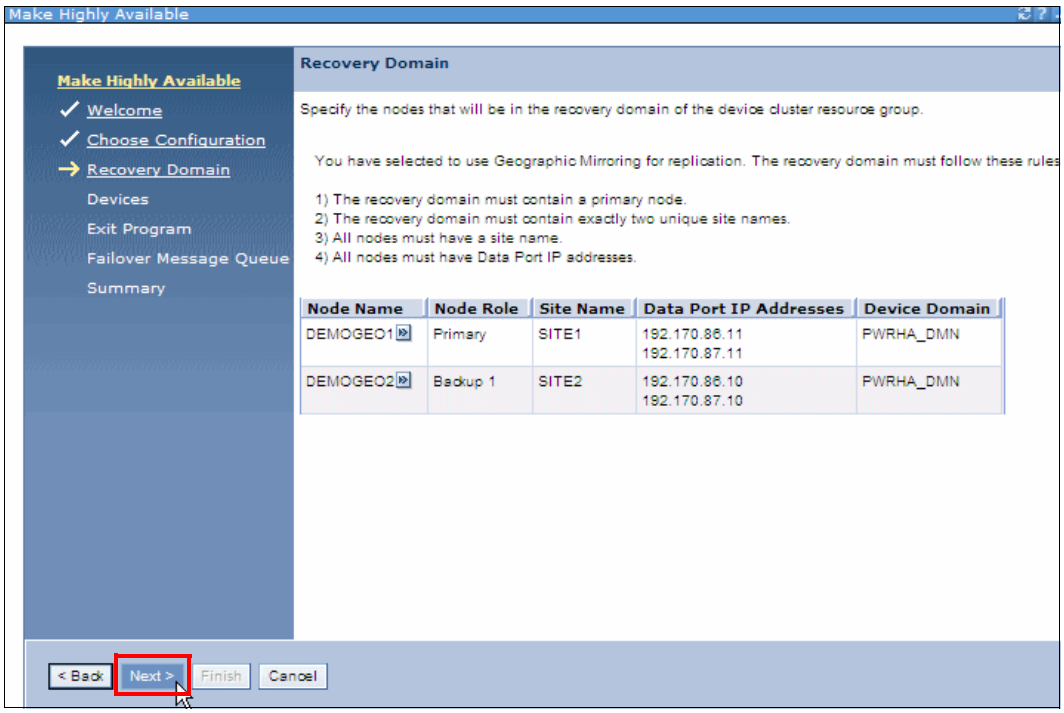


Figure 12-6 Recovery Domain summary

7. If a device domain does not already exist, the wizard prompts you to and add the nodes to a domain (Figure 12-7). Click **OK** to add them.

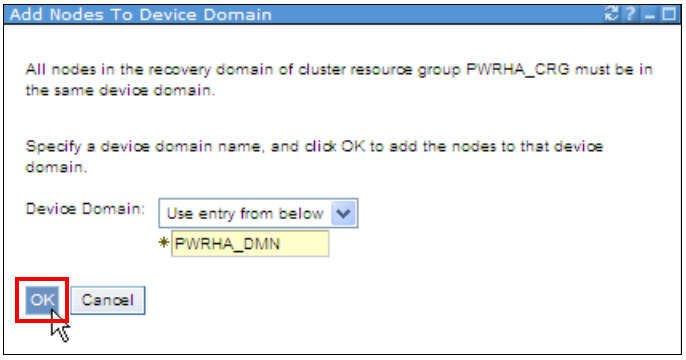


Figure 12-7 Add Nodes to Device Domain

The wizard shows progress pop-ups when adding the nodes to a device domain (Figure 12-8).

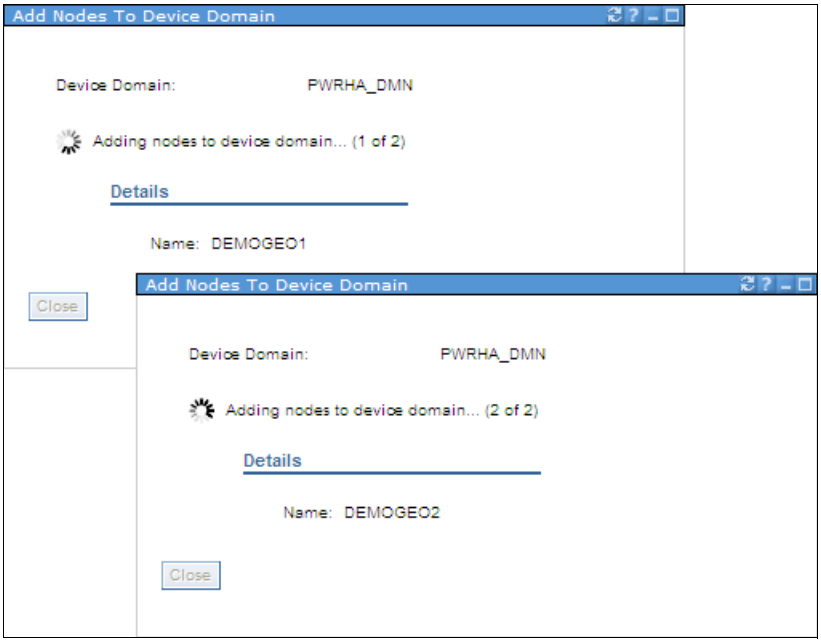


Figure 12-8 Adding nodes to device domain progress pages

8. On the Devices panel, choose **Modify** from the IASP's popup menu (Figure 12-9) to specify a *server takeover IP address* and vary on options for the IASP.

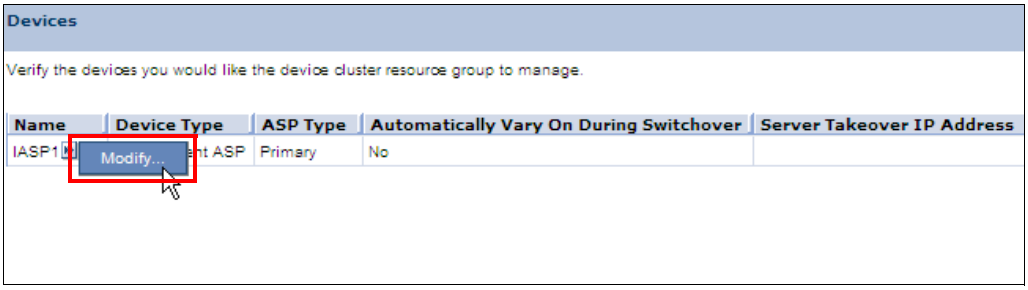


Figure 12-9 Server takeover IP address

9. For our environment, we choose to automatically vary on the device on the backup node and specify a server takeover IP address for the users to connect to, which will be switched/failed over together with the device CRG. Click **Save** to go back to the Devices page (Figure 12-10).

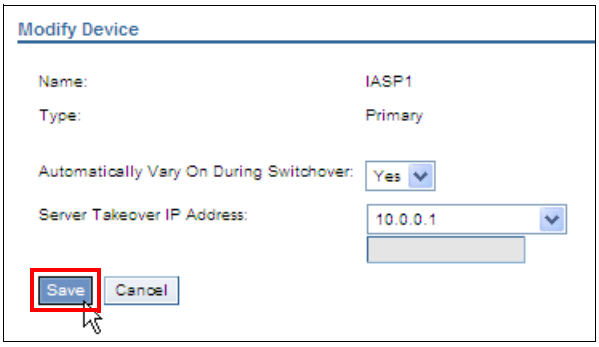


Figure 12-10 Modify IASP properties

10. You can specify an exit program to be executed after a switchover. This is optional, but if specified, it must exist on all nodes in the recovery domain of the CRG. The exit program can be used to set up an application environment and any other priming steps that are necessary before users are allowed on the new production node after the switchover. For our scenario, we leave the default Exit Program option as No and click **Next** (Figure 12-11).

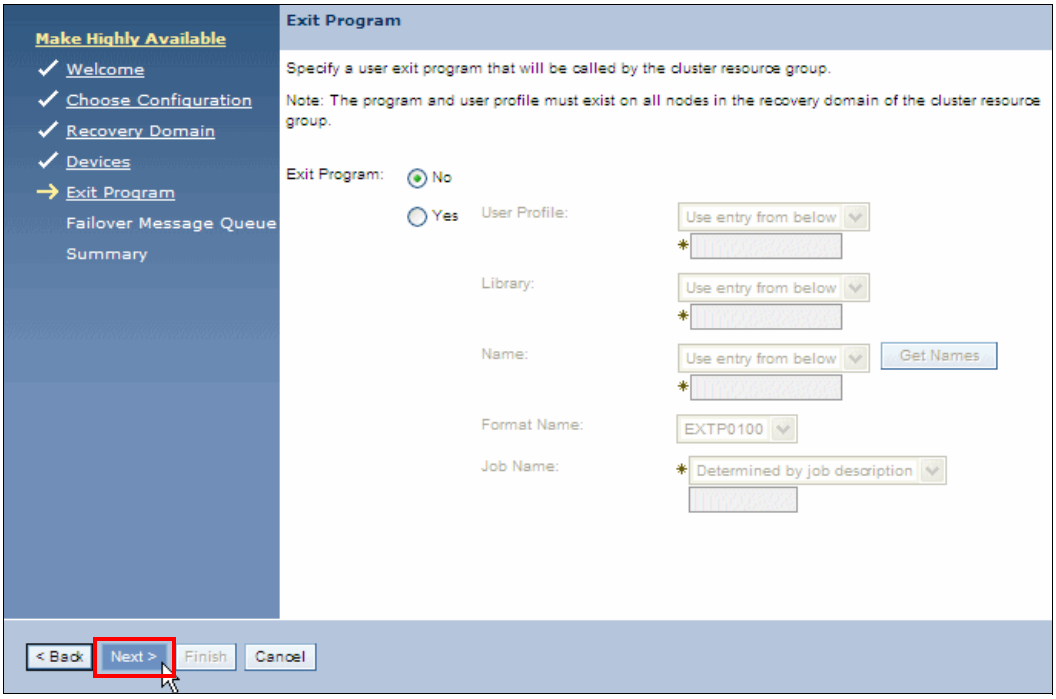


Figure 12-11 Specify Exit Program

11. Specify an optional *failover message queue* to prevent an automatic failover in case of any unplanned node or server failures. For our environment, we use the QSYSOPR message queue with an indefinite wait time parameter so that we can respond to the CPABB02 inquiry message to either proceed with or cancel the failover. We choose the default message action as *Cancel failover* (Figure 12-12).

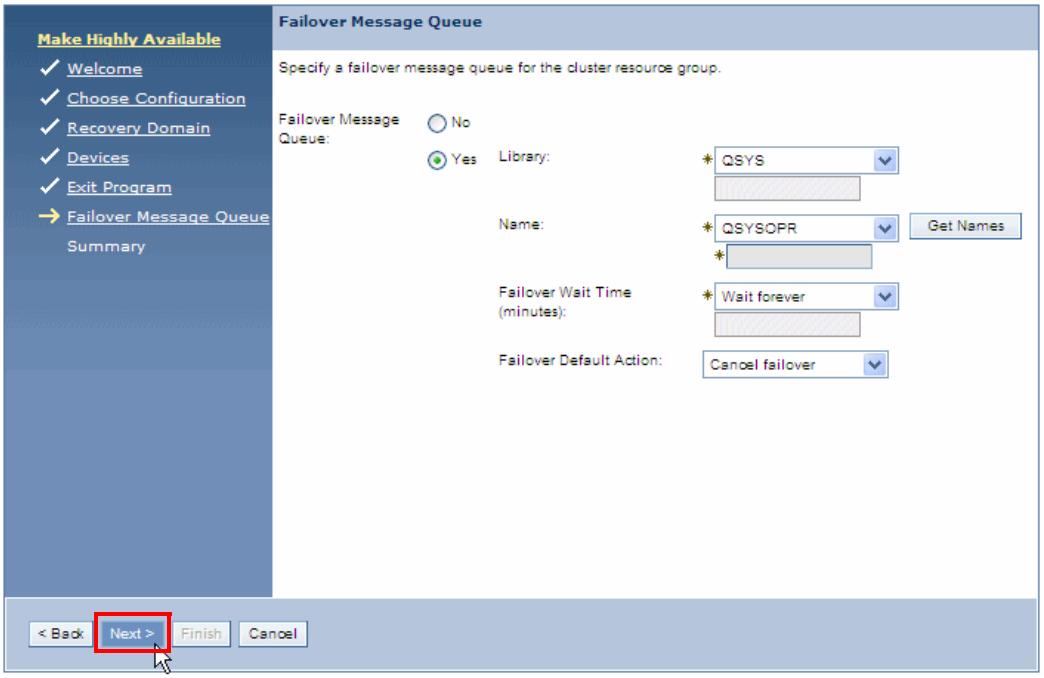


Figure 12-12 Specify Failover Message Queue

12. The summary page (Figure 12-13) lets you view the configuration options chosen up until this point and gives you an opportunity to go back and modify any of the options. If you are satisfied with the summary, click **Finish** to make the IASP highly available.

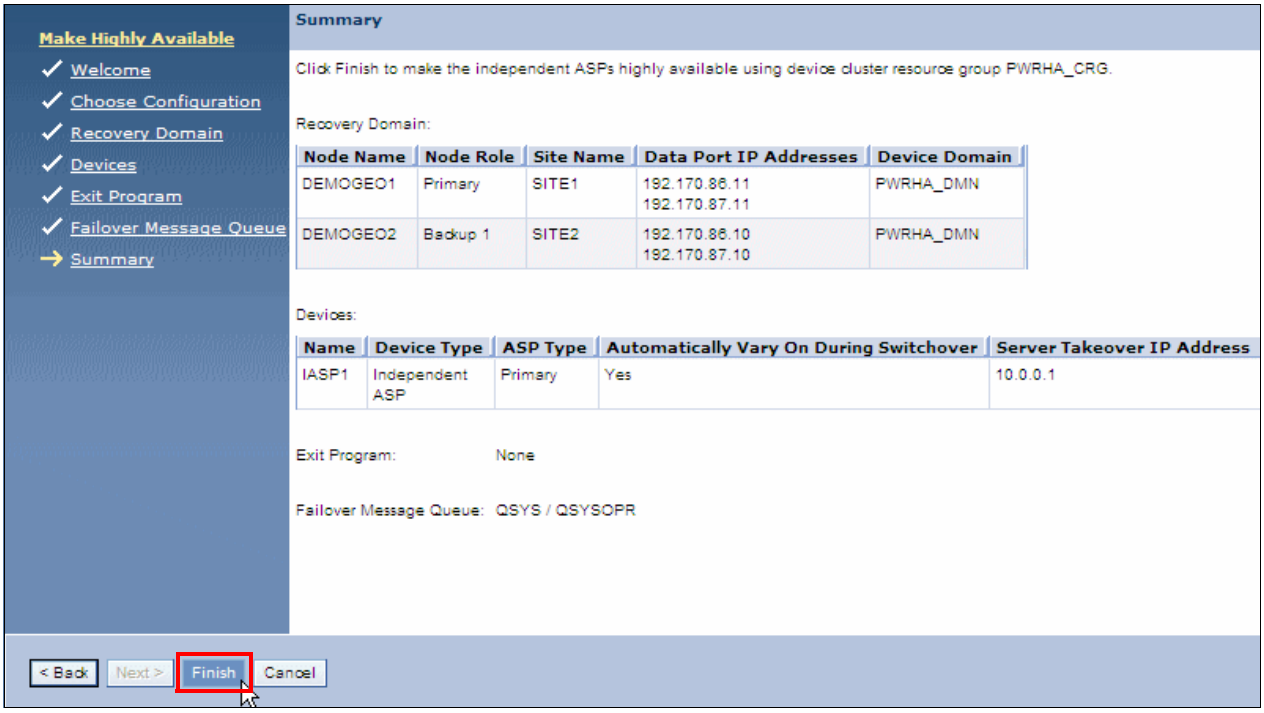


Figure 12-13 Summary page of configuration options

13. If there are no issues with the configuration, you will see a success dialog box (Figure 12-14). PowerHA will add the nodes to the recovery domain, set up ASP copy descriptions, and set up an ASP session for geographic mirroring.

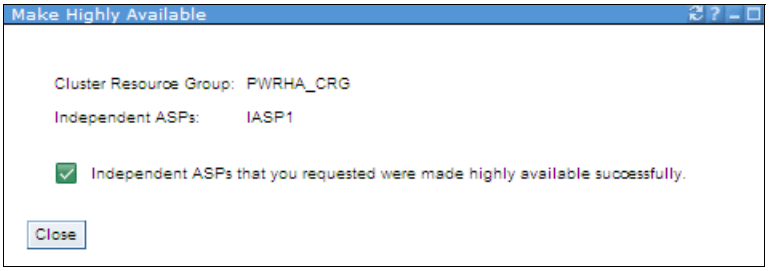


Figure 12-14 Make Highly Available success dialog box

14. If the IASP was never set up for geographic mirroring, the wizard asks you whether you want to configure mirroring (Figure 12-15). Choose **Yes** to start the Configure Mirroring wizard in PowerHA GUI.

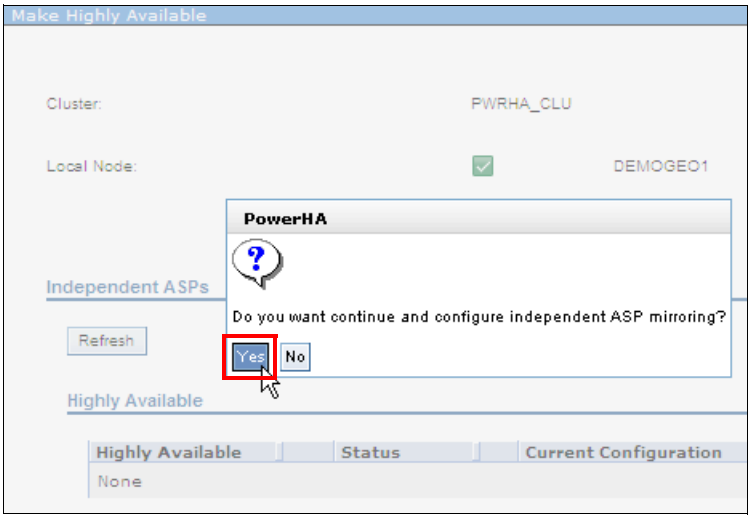


Figure 12-15 Prompt to continue setting up mirroring for the IASP

The Configure Mirroring wizard can also be started using the IASP context menu (Figure 12-16).

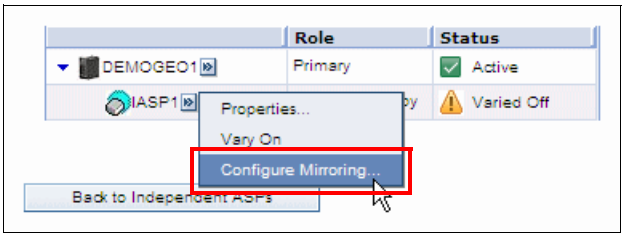


Figure 12-16 Configure Mirroring for an IASP

The geographic mirroring for the IASP can also be managed by PowerHA if it was set up elsewhere. When PowerHA detects that geographic mirroring was configured but that there are no ASP copy descriptions or an ASP session that can be managed from PowerHA GUI, it prompts you to create them (Figure 12-17).

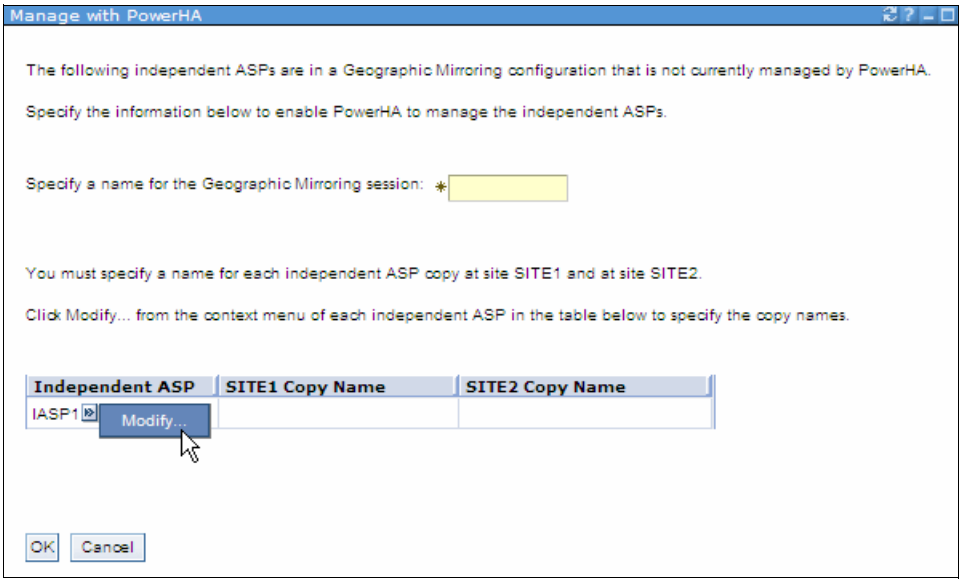


Figure 12-17 Manage geographic mirroring with PowerHA GUI

15. The Configure Mirroring wizard (Figure 12-18) can be used to set up an ASP session, ASP copy descriptions, and so on, to be used for geographic mirroring. On the Welcome page, click **Next** to choose the mirroring configuration options.

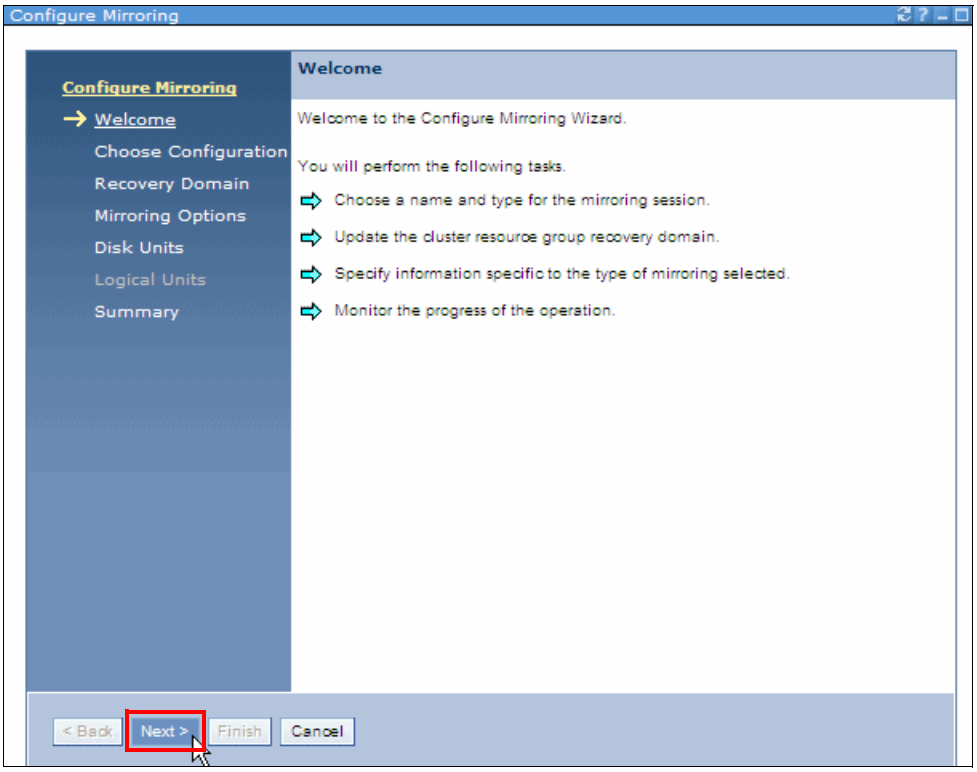


Figure 12-18 Configure Mirroring wizard Welcome panel

16. On the Choose Configuration page, select **Geographic Mirroring** as the type and specify a name for the ASP session (Figure 12-19).

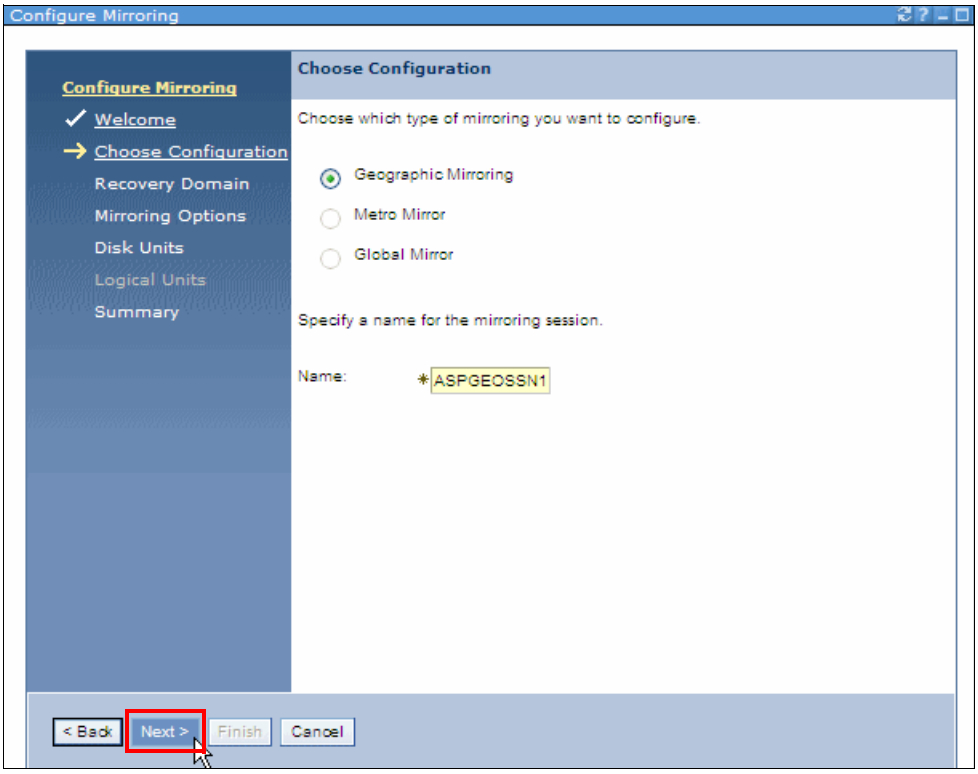


Figure 12-19 Specify mirroring type and a name for the ASP mirroring session

17. The Recovery Domain page (Figure 12-20) gives you a summary and also lets you make adjustments to the site names, data ports, and so on.

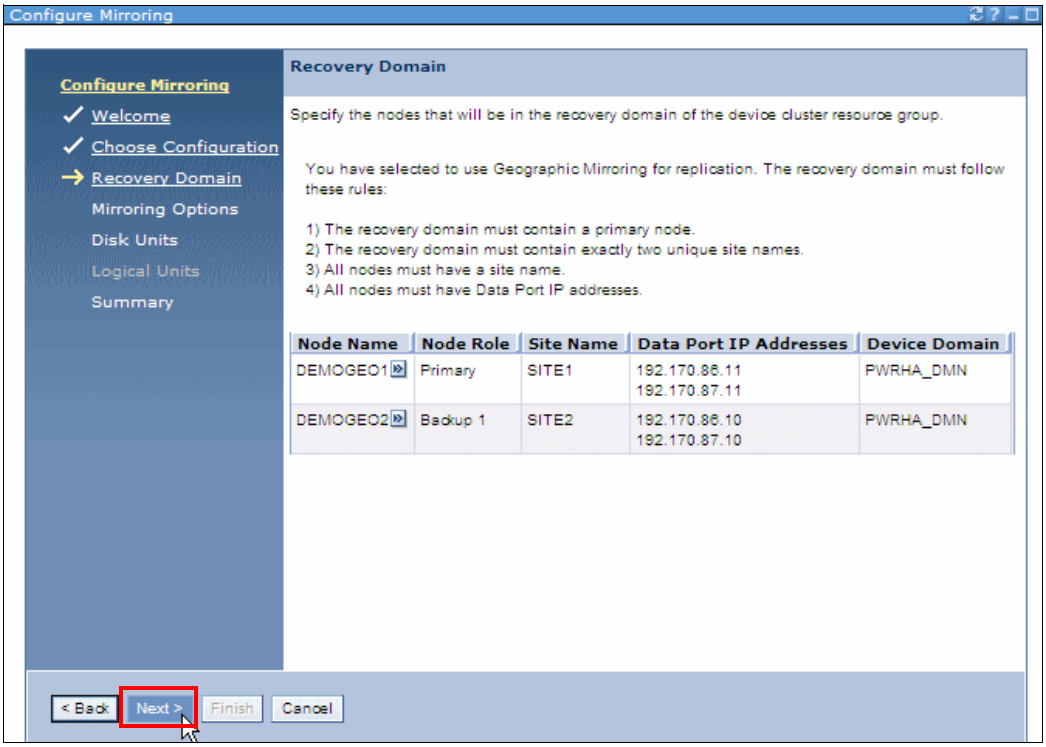


Figure 12-20 Recovery Domain summary

18. The Configure Mirroring wizard then takes you to the Mirroring Options panel, where you can add copy descriptions and specify advanced mirroring options. Choose **Modify** from the IASP context menu (Figure 12-21).

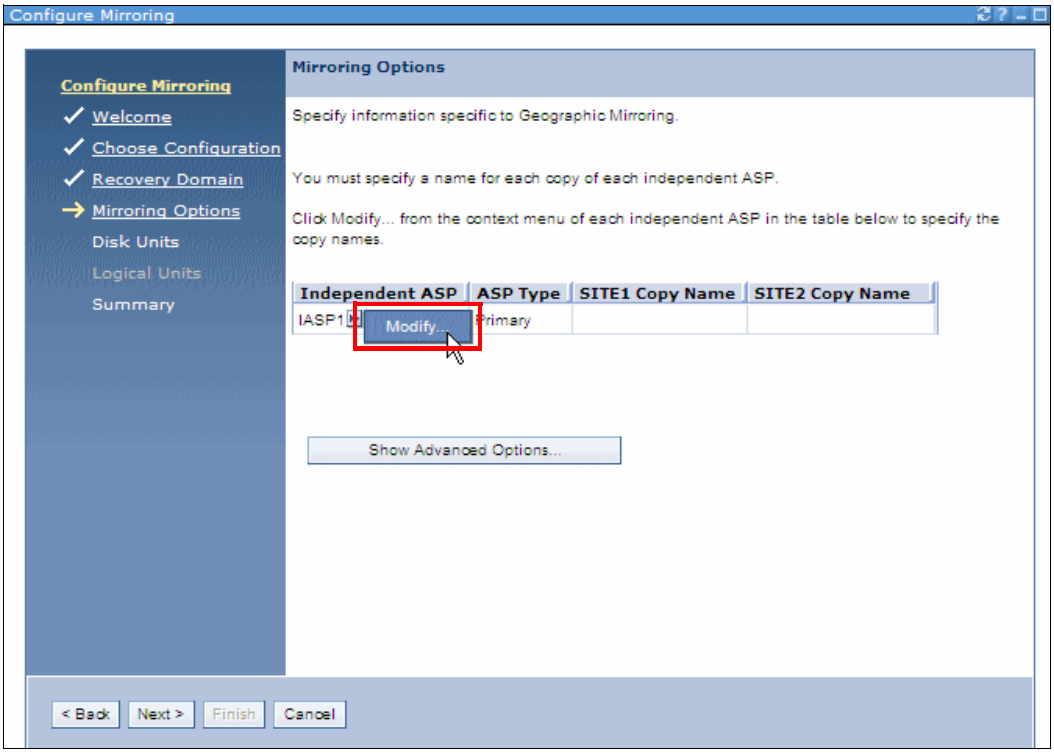


Figure 12-21 Context menu to modify IASP

19. Specify names for the ASP copy descriptions and click **Save** (Figure 12-22).

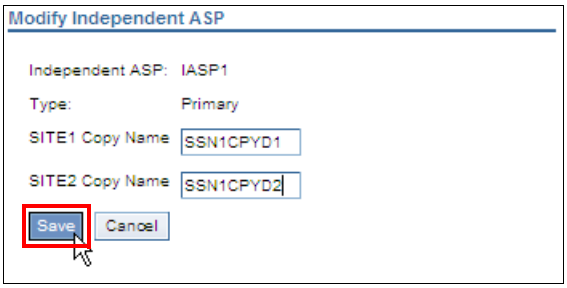


Figure 12-22 Add copy descriptions

20. On the Mirroring Options panel, click **Show Advanced Options** to change any of the mirroring attribute defaults. This is where you can specify whether you want synchronous or asynchronous delivery mode and synchronous or asynchronous data transmission. The synchronous or asynchronous mirroring mode determines whether the writes on the backup node are acknowledged back to the primary node after they are received in main memory (synchronous) or after they are written to the disk (asynchronous), respectively, on the mirror copy. Figure 12-23 shows the available advanced options for the mirroring session.

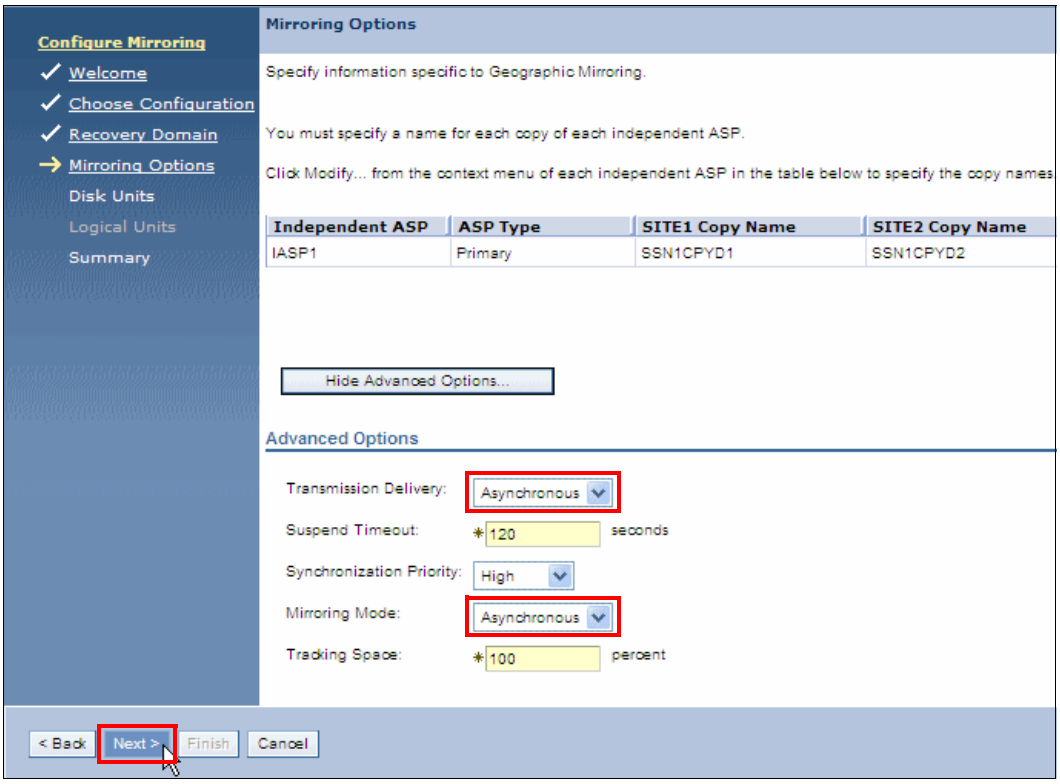


Figure 12-23 Advanced options for geographic mirroring session

21. The Configure Mirroring wizard then takes you to Disk Unit configuration page for creating the IASP mirror copy on the backup node. As shown in Figure 12-24, you can select the disk units to create the IASP from a list of available unconfigured disk units on the backup node. Click **Add** to add them to the selected Disk Units panel on the wizard.

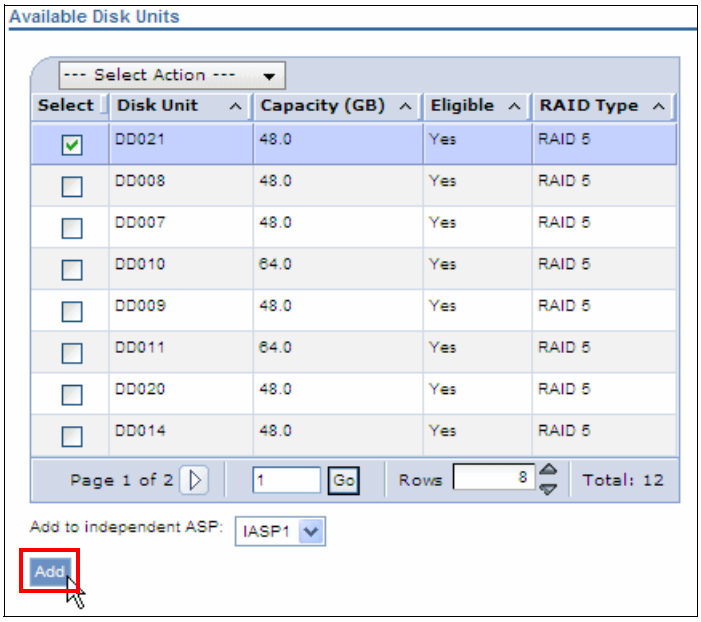


Figure 12-24 Select disk units to create the IASP mirror copy

22. Verify the total disk capacity on the Selected Disk Units pane (Figure 12-25) and make sure that you have adequate capacity to match the production copy. For more information about sizing considerations refer to “Disk subsystem considerations” on page 146.

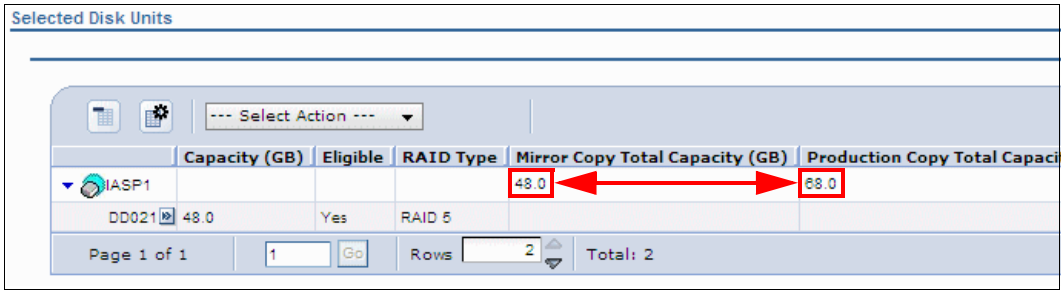


Figure 12-25 Selected Disk Units and their capacities

23. If the total capacity for the mirror copy does not match the capacity of the production copy, PowerHA issues a warning (Figure 12-26). For our scenario, we choose **Yes** to continue with the setup.

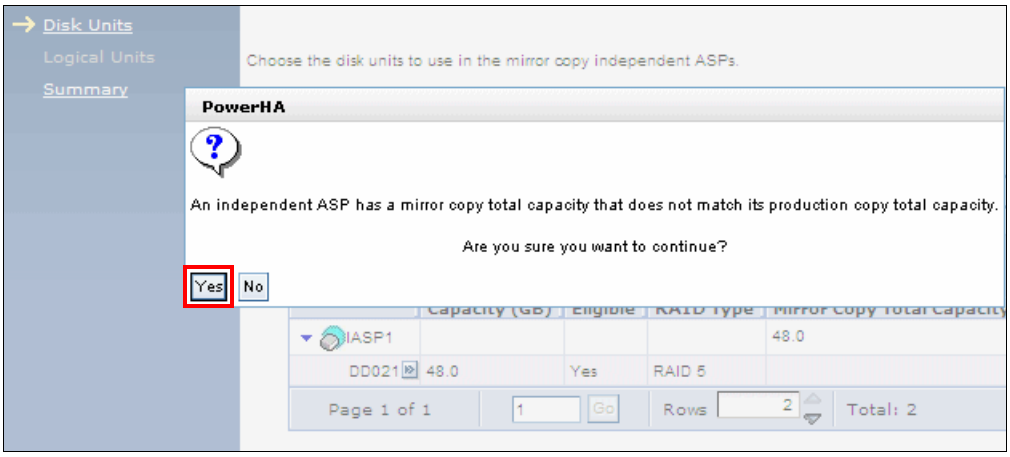


Figure 12-26 Capacity mismatch warning

24. The Configure Mirroring wizard shows you a summary page with all the details for the mirroring session, recovery domain, ASP copy descriptions, and the disk unit selection for the mirror copy (Figure 12-28 on page 252).

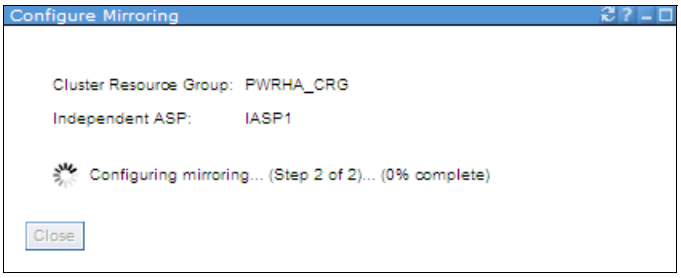


Figure 12-27 Configure Mirroring progress

Click **Finish** to complete the setup. You should see a configuration progress and successful completion message (Figure 12-27 on page 251).

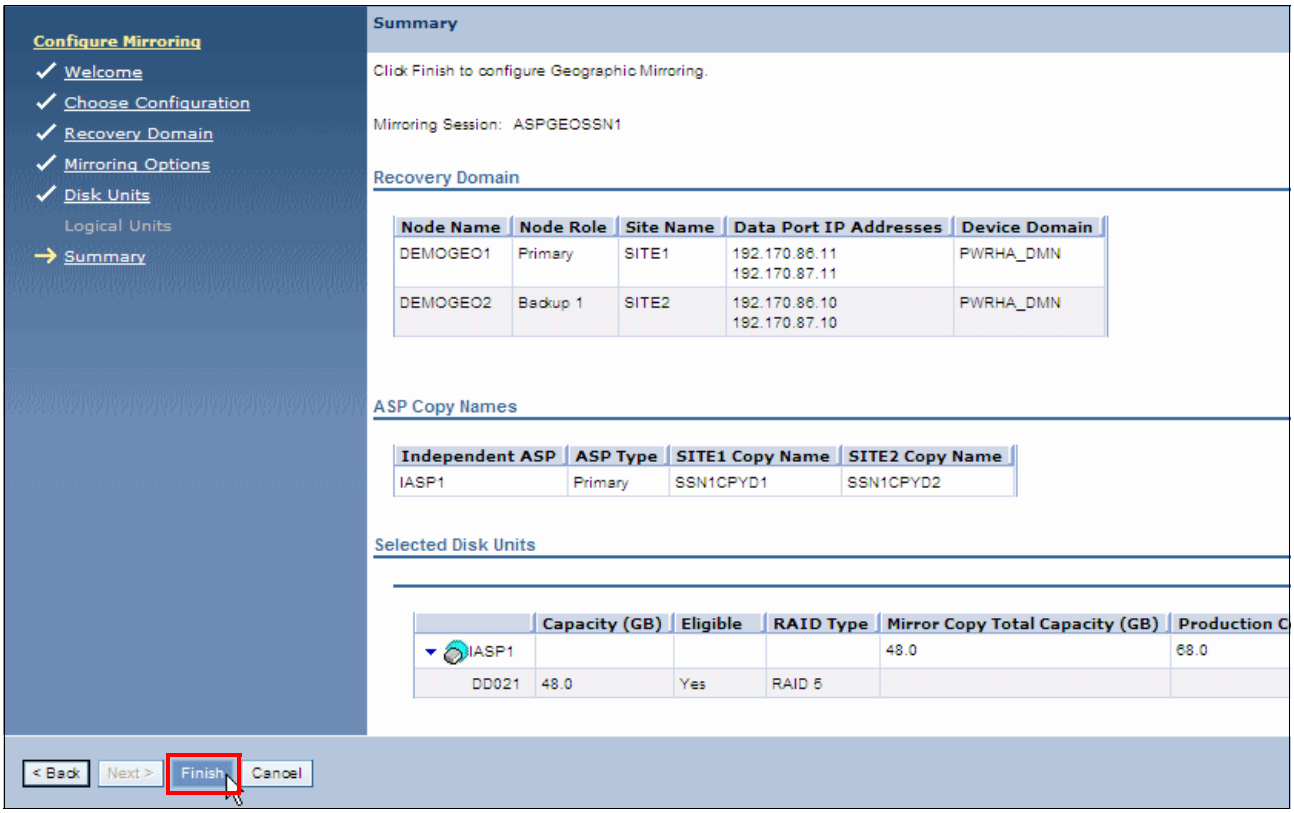


Figure 12-28 Configure Mirroring summary

Setting up geographic mirroring with CL commands

The setup shown above can also be done using CL commands, including the new **CFGGEOMIR** command (Example 12-1).

Example 12-1 CL commands used for similar geographic mirroring setup

```
CRTDEVASP DEVD(IASP1) RSRCNAME(IASP1)
ADDDEVDMNE CLUSTER(PWRHA_CLU) DEVDMN(PWRHA_DMN) NODE(DEMOGEO1)
ADDDEVDMNE CLUSTER(PWRHA_CLU) DEVDMN(PWRHA_DMN) NODE(DEMOGEO2)
CRTCRG CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG) CRGTYPE(*DEV) EXITPGM(*NONE)
   USRPRF(*NONE) RCYDMN(
      (DEMOGEO1 *PRIMARY *LAST SITE1 ('192.170.86.11' '192.170.87.11'))
      (DEMOGEO2 *BACKUP 1 SITE2 ('192.170.86.10' '192.170.87.10')))
   CFGOBJ((IASP1 *DEVD *ONLINE '10.0.0.1'))
   FLVMSGQ(QSYS/QSYSOPR) FLVWAITTIM(*NOMAX) FLVDFTACN(*CANCEL)
CFGGEOMIR ASPDEV(IASP2) ACTION(*CREATE)
   SRCSITE(SITE1) TGTSITE(SITE2)
   SSN(SSN1CPYD2/SSN1CPYD1/ASPGEOSSN1)
   DELIVERY(*ASYNC) UNITS(DD021)
   CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG) MODE(*ASYNC)
STRCRG CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG)
```

12.2 Managing geographic mirroring

The PowerHA GUI provides you with one-stop access to managing your entire HA environment including managing these:

- The cluster and its properties
- Cluster nodes
- Independent ASPs
- Cluster administrative domains
- Cluster resource groups
- TCP/IP interfaces

In the following sections we show you how to manage various aspects of your high-availability environment, including performing a planned switchover.

12.2.1 Administrative domain

In a high-availability environment it is necessary that the application and operational environment remain consistent among the nodes that participate in high availability. The cluster administrative domains interface on the PowerHA GUI allows you to maintain environment resiliency and ensures that the operational environment remains consistent across the nodes. The cluster administrative domain (Figure 12-29) provides the mechanism that keeps resources synchronized across systems and partitions within a cluster.

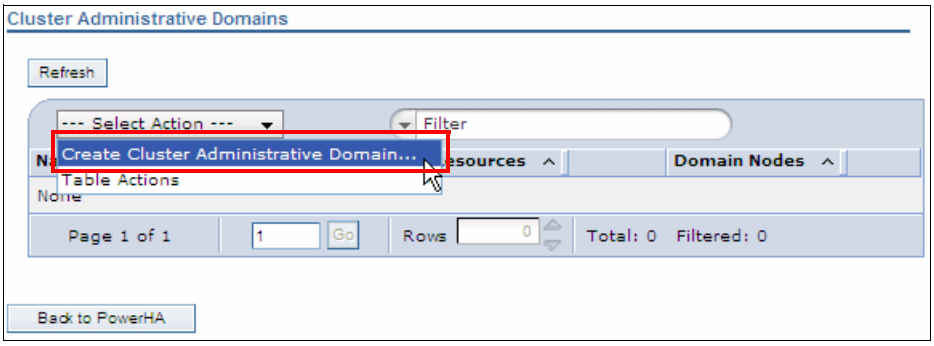


Figure 12-29 Create Cluster Administrative Domain option

You can add all the nodes that would participate in a cluster administrative domain (Figure 12-30).

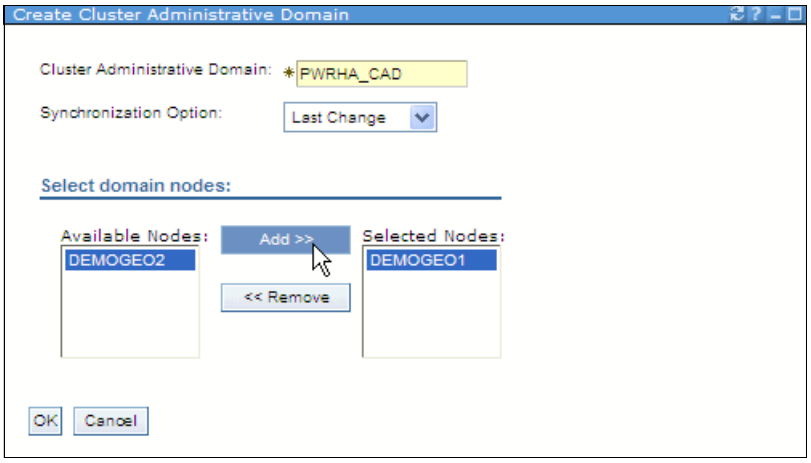


Figure 12-30 Add nodes to cluster administrative domain

Monitored resources

A monitored resource is a system object or a set of attributes not associated with a specific system object, such as the set of system environment variables. A resource represented by an MRE is monitored for changes by a cluster administrative domain. When a monitored resource is changed on one node in a cluster administrative domain, the change is propagated to the other active nodes in the domain. You can access the Monitored Resources panel from the Cluster Administrative Domain context menu (Figure 12-31).

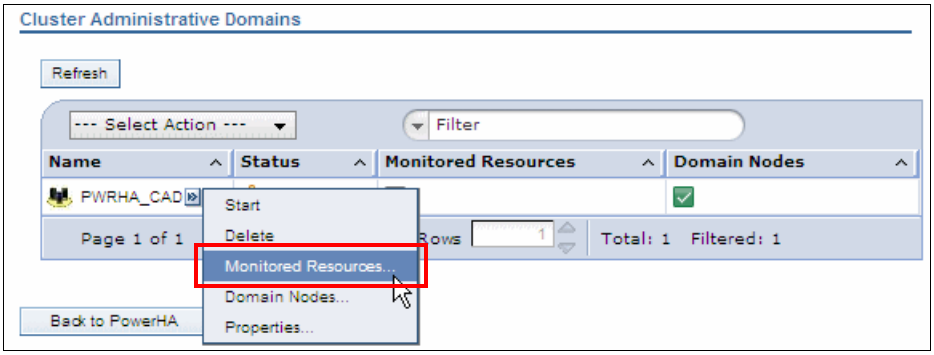


Figure 12-31 Monitored Resources in a Cluster Administrative Domain

On the Monitored Resources interface, click the **Select Action** drop-down menu and choose **Add Monitored Resources** (Figure 12-32).

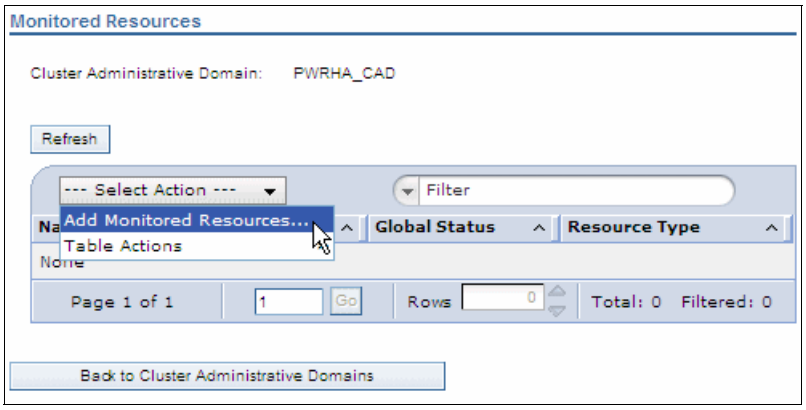


Figure 12-32 Add Monitored Resources option

On the Add Monitored Resources interface (Figure 12-33), you can type the name of the resource or select from a list. When adding an MRE for a system object, the resource name is the name of the system object. One or more attributes can be specified, and only attributes that are specified will be monitored for changes.

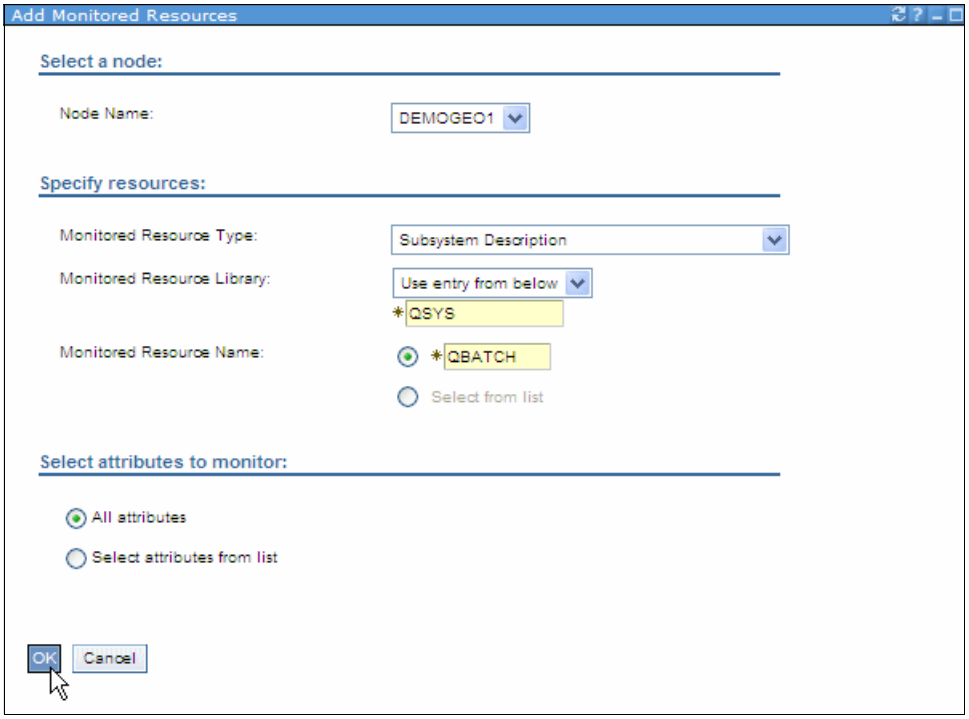


Figure 12-33 Add Monitored Resources

You can also access the attributes interface for a resource from the context menu of the Monitored Resource page (Figure 12-34).

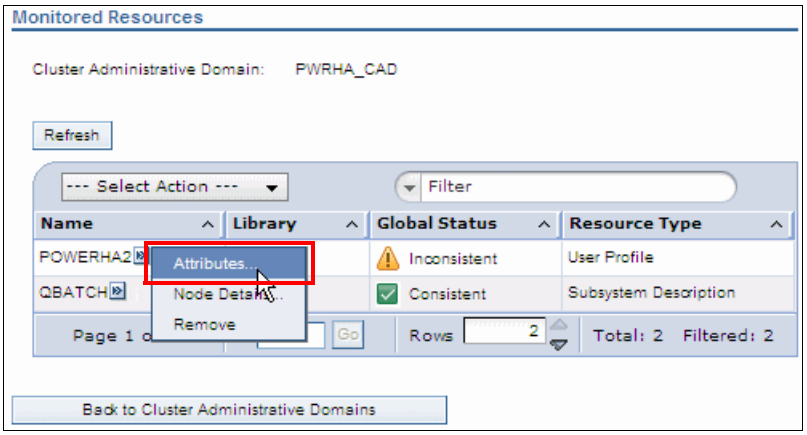


Figure 12-34 Access attributes option for an MRE

On the Attributes interface, you can see which attributes are consistent and which are inconsistent. As shown in Figure 12-35, you might find it easier to click the **Global Status** column to sort the list and identify the attributes that are in an inconsistent state.

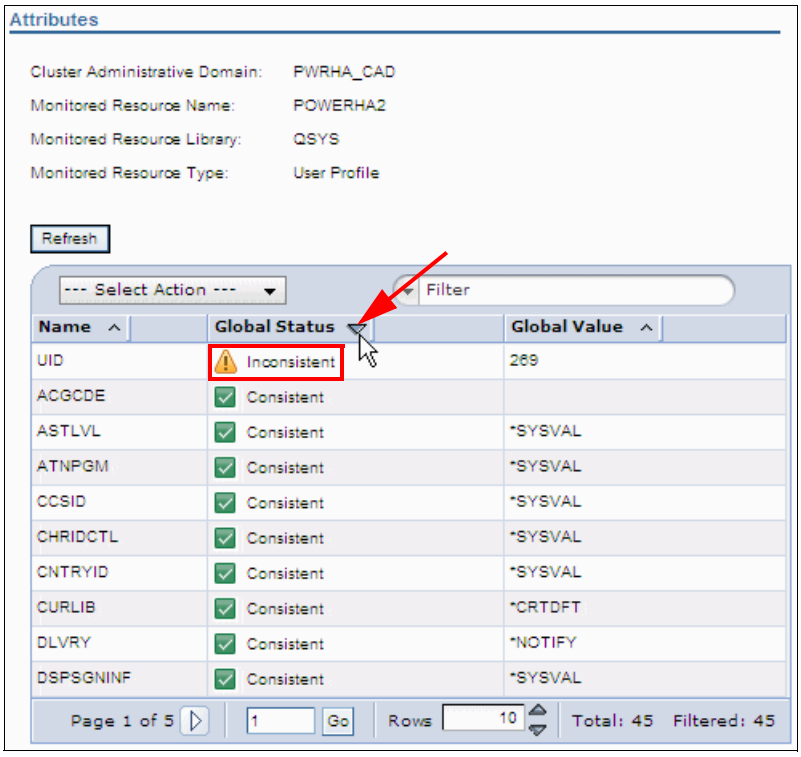


Figure 12-35 Attribute details for a user profile MRE

CL commands for administrative domain and MREs

The setup shown above can also be done using the commands shown in Example 12-2.

Example 12-2 CL commands used for Cluster Administrative Domain

```
CRTCAD CLUSTER(PWRHA_CLU) ADMDMN(PWRHA_CAD) DMNNODL(DEMOGEO1)
ADDCADNODE CLUSTER(PWRHA_CLU) ADMDMN(PWRHA_CAD) NODE(DEMOGEO2)
ADDCADMRE CLUSTER(PWRHA_CLU) ADMDMN(PWRHA_CAD) RESOURCE(QBATCH)
   RSCTYPE(*SBSD) RSCLIB(QSYS)
ADDCADMRE CLUSTER(PWRHA_CLU) ADMDMN(PWRHA_CAD) RESOURCE(POWERHA)
   RSCTYPE(*USRPRF) ATTRIBUTE(*ALL)
```

12.2.2 Planned switchover

Before attempting a planned switchover, it is very important that all the components in the HA environment are functional and in a green status. You can verify this from the main PowerHA GUI landing page that shows you the status of these:

- Cluster nodes: All cluster nodes should be active.
- Independent ASPs: All independent ASPs are available.
- Cluster administrative domains: Active and all monitored resources are consistent.
- Cluster resource groups: All CRGs are active.
- TCP/IP interfaces: All interfaces are active or in an expected status.

Figure 12-36 shows a fully functional HA environment.

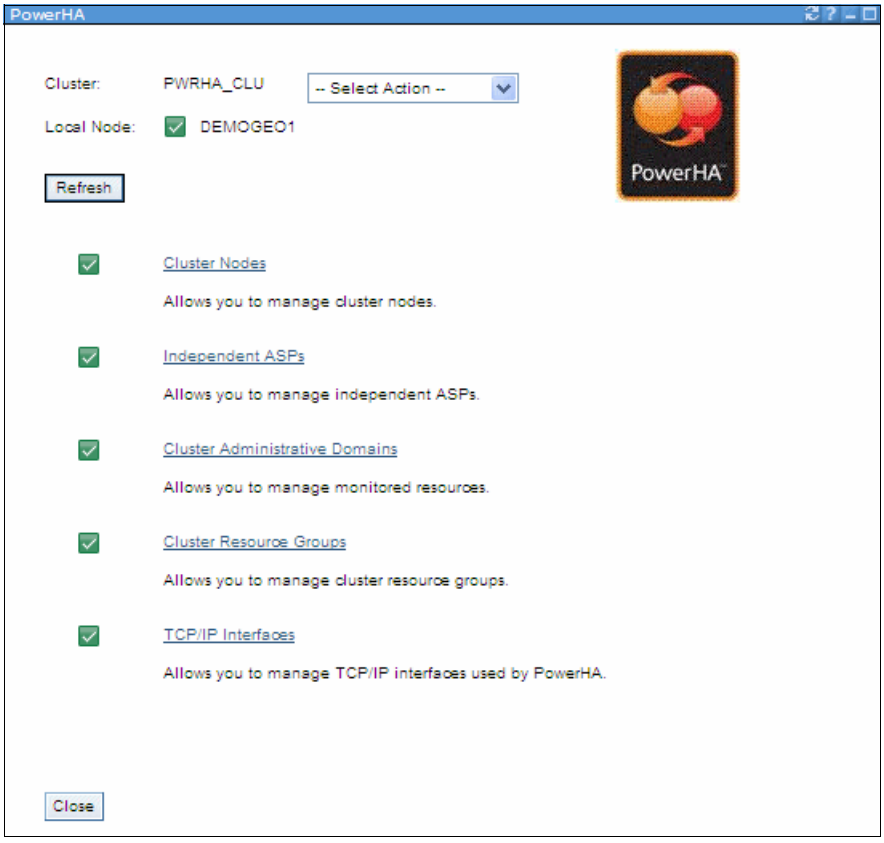


Figure 12-36 PowerHA main panel with status of all components

To perform a switchover, choose **Switchover** from the context menu of the cluster resource group (Figure 12-37).

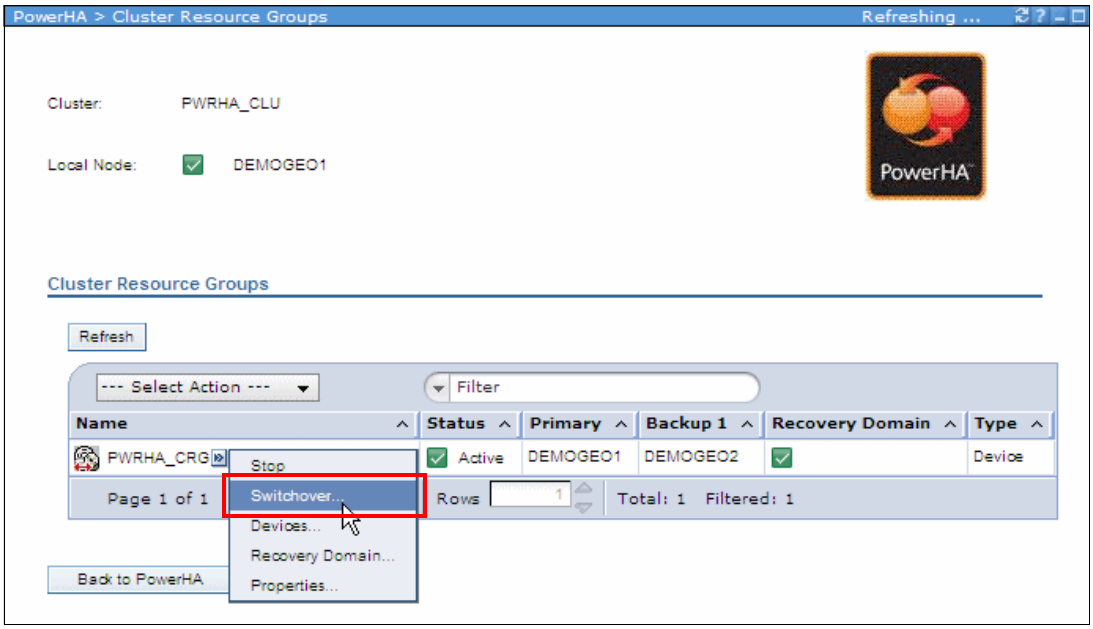


Figure 12-37 Switchover option for a cluster resource group

When you select the **Switchover** option, PowerHA gives you a preview of how the nodes and their roles in a cluster resource group will look before and after a switchover (Figure 12-38).

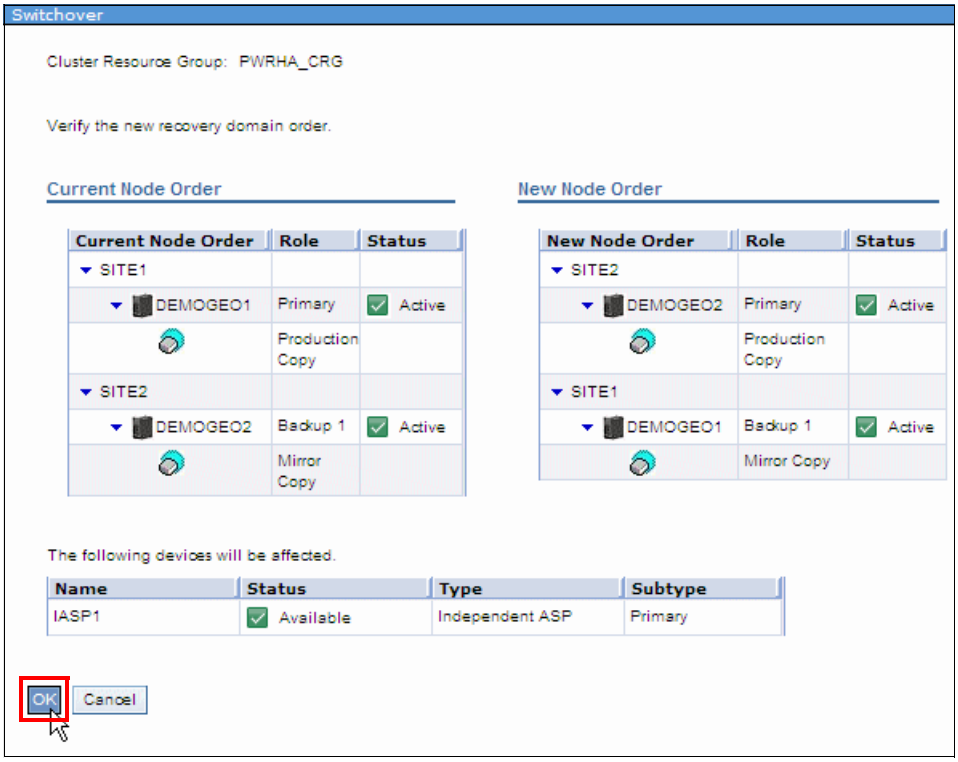


Figure 12-38 Switchover summary with current and new node order

On the recovery domain preview screen, verify that the roles of various nodes are as expected and click **OK** to proceed with the switchover.

PowerHA shows you various progress messages and gives you a success dialog if everything went well (Figure 12-39).

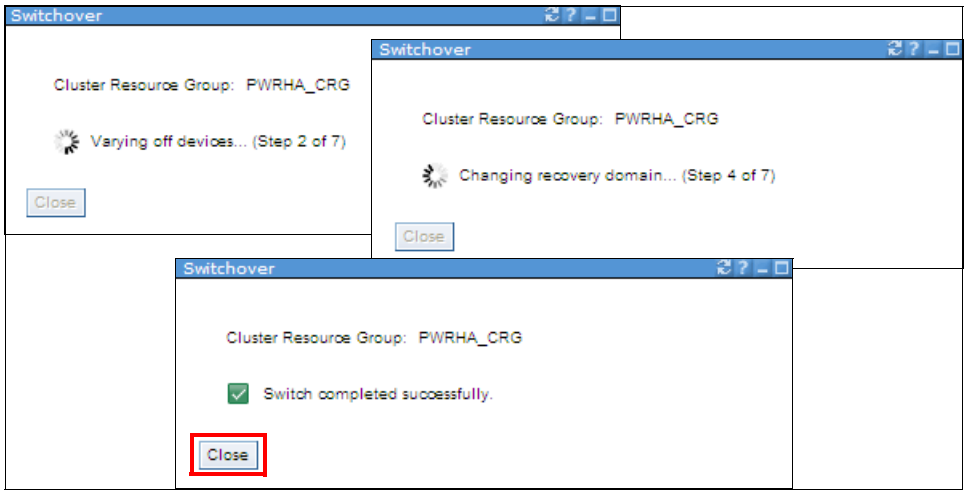


Figure 12-39 Switchover progress and completion messages

After you click **Close** on the Switch completed successfully dialog, you will be back on the cluster resource groups interface. Click **Refresh** to view the updated role status for all the nodes in the recovery domain (Figure 12-40).

Figure 12-40 CRG status after node role reversal

To verify the status of geographic mirroring and the current direction of replication, go to the Independent ASP details interface (Figure 12-41).

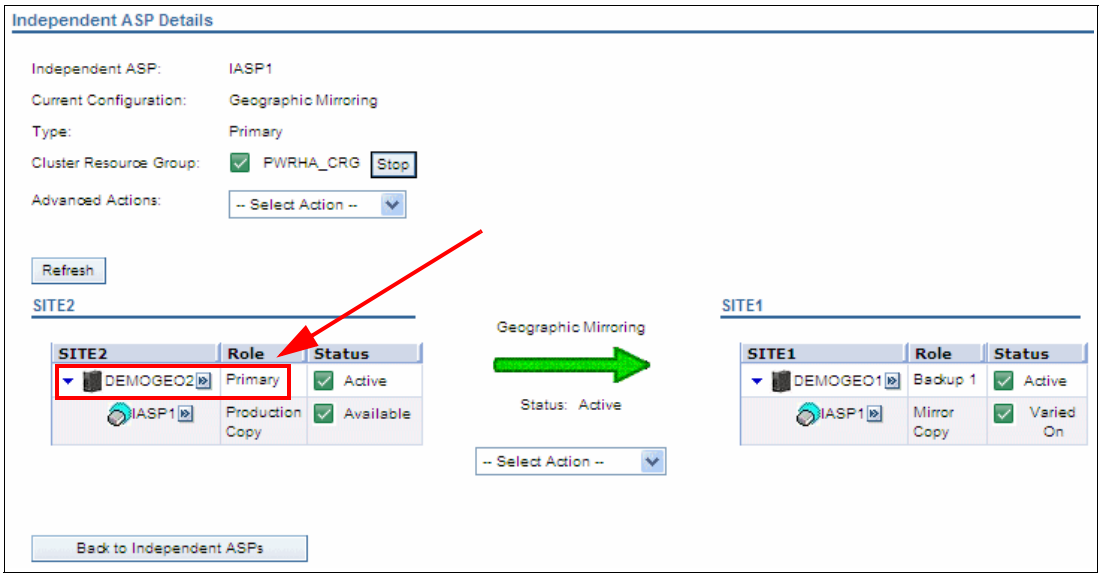


Figure 12-41 IASP details after switchover

12.2.3 Deconfiguring geographic mirroring

If you no longer want the capability to use geographic mirroring for a specific disk pool or disk pool group, you can select **Deconfigure Geographic Mirroring**. If you deconfigure geographic mirroring, the system stops geographic mirroring and deletes the mirror copy of the disk pools on the nodes in the mirror copy site.

To deconfigure geographic mirroring, the disk pool must be offline. First vary off the independent ASP and then select **Deconfigure Geographic Mirroring** from the IASP pop-up menu on the PowerHA GUI (Figure 12-42).

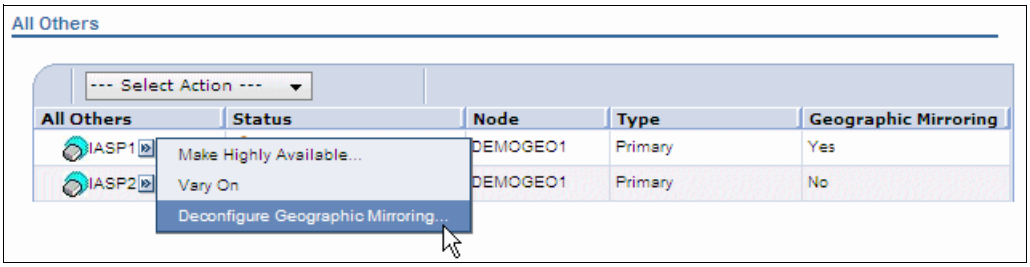


Figure 12-42 Deconfigure Geographic Mirroring menu option

Click **OK** on the confirmation dialog to continue deconfiguring geographic mirroring.

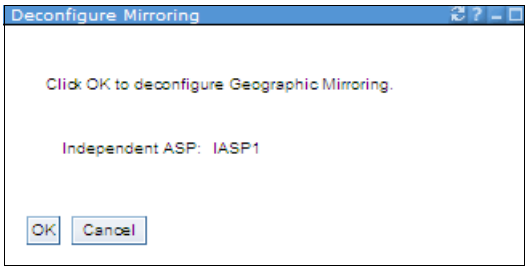


Figure 12-43 Confirmation dialog to deconfigure geographic mirroring

13

Configuring and managing DS8000 Copy Services

In this chapter we describe the scenario that uses DS8000 Copy Services with IBM PowerHA SystemMirror for i.

In various sections of this chapter, we discuss the following activities using an IBM i DS8000:

- Setting up a Copy Services environment
- Configuring Metro Mirror
- Configuring FlashCopy
- Configuring Global Mirror
- Configuring LUN-level switching
- Managing IBM i DS8000 Copy Services

13.1 Setting up IBM i DS8000 Copy Services

PowerHA SystemMirror for i controls IBM i clustering functions such as switchover and failover. When used with DS8000 Copy Services, it also needs to be able to control the Copy Services functions like PPRC failover/failback or FlashCopy. This requires the installation of the DS command-line interface (DS CLI) on the IBM i partitions. There is no need for a user to take care of the way the DSCLI works on an IBM i partition because it is used by PowerHA “under-the-cover”.

The DS CLI installation package can be downloaded from the following URL:

`ftp://ftp.software.ibm.com/storage/ds8000/updates/DS8K_Customer_Download_Files/CLI`

The install packages are provided as ISO image files, which can be used by any virtual CD drive tool on Windows.

For DS CLI installation procedures, refer to the *IBM System Storage DS Command-Line Interface User's Guide for the DS6000 series and DS8000 series*, GC53-1127, included in the install package and section 8.3 in *IBM i and IBM System Storage: A Guide to Implementing External Disks on IBM i*, SG24-7120.

For the IBM i partitions to be able to manage DS8000 Copy Services, we need to provide a DS storage user profile and its password. Whenever the password is changed on the DS8000, the configuration on the IBM i partition needs to be changed at the same time. For more information refer to section 9.5, “User management,” in *IBM System Storage DS8000: Architecture and Implementation*, SG24-8886.

Note: User profiles and passwords are case sensitive on the DS8000.

The communication via DS CLI requires an IP connection between the IBM i partition and the DS8000. This connection is initiated by IBM i to the DS8000, which listens on TCP port 1750.

Note: If there is a firewall between the IBM i partition and the DS8000, TCP port 1750 must be authorized.

13.1.1 Configuring IBM i DS8000 Metro Mirror (GUI and CL commands)

This section covers a scenario using a DS8000 Metro Mirror environment as the hardware replication technology.

Environment overview

For this scenario, we need to create several cluster items. Cluster, cluster monitor, IASP, and administrative domain setup are covered in Chapter 11, “Creating a PowerHA base environment” on page 199. They are common to all scenarios that we describe in this chapter.

These are cluster items that we specifically configure for our Metro Mirror environment:

- Cluster resource group
- Device domain
- ASPcopy descriptions and session

Table 13-1 and Figure 13-1 on page 266 show the setup that we use for our environment.

Table 13-1 Metro Mirror scenario settings

	Preferred production	**Preferred backup**
System name	DEMOPROD	DEMOHA
Cluster name	PWRHA_CLU	
Cluster resource group	PWRHA_CRG1	
Device domain	PWRHA_DMN	
Administrative domain	PWRHA_CAD	
IASP name, number	IASP1, 33	IASP1, 33
Cluster resource group site name	SITE1	SITE2
ASP copy description	IASP1_MM1	IASP1_MM2
ASP session	IASP1_MM	
Takeover IP	10.0.0.1	
Heartbeat Cluster IP	10.10.10.1	10.10.10.2
Management access IP	9.5.168.129	9.5.168.130
DS8000 device ID	IBM.2107-75AY031	IBM.2107-75AY032
DS8000 IP address[a]	9.5.168.55	9.5.168.55
Volume group ID	V2	V2
Volume IDs[b]	6000-6002, 6100-6102	6000-6002, 6100-6102
FC IO adapter resource name[c]	DC03	DC07
Host connection ID	3	1
Host connection WWPN	10000000C948031F	10000000C9523F21

a. IP addresses are identical because, in our environment, we use a single DS8000 configured with two distinct storage images.
b. Volume IDs are not needed to be the same on the source and target Metro Mirror relationships.
c. FC stands for Fibre Channel.

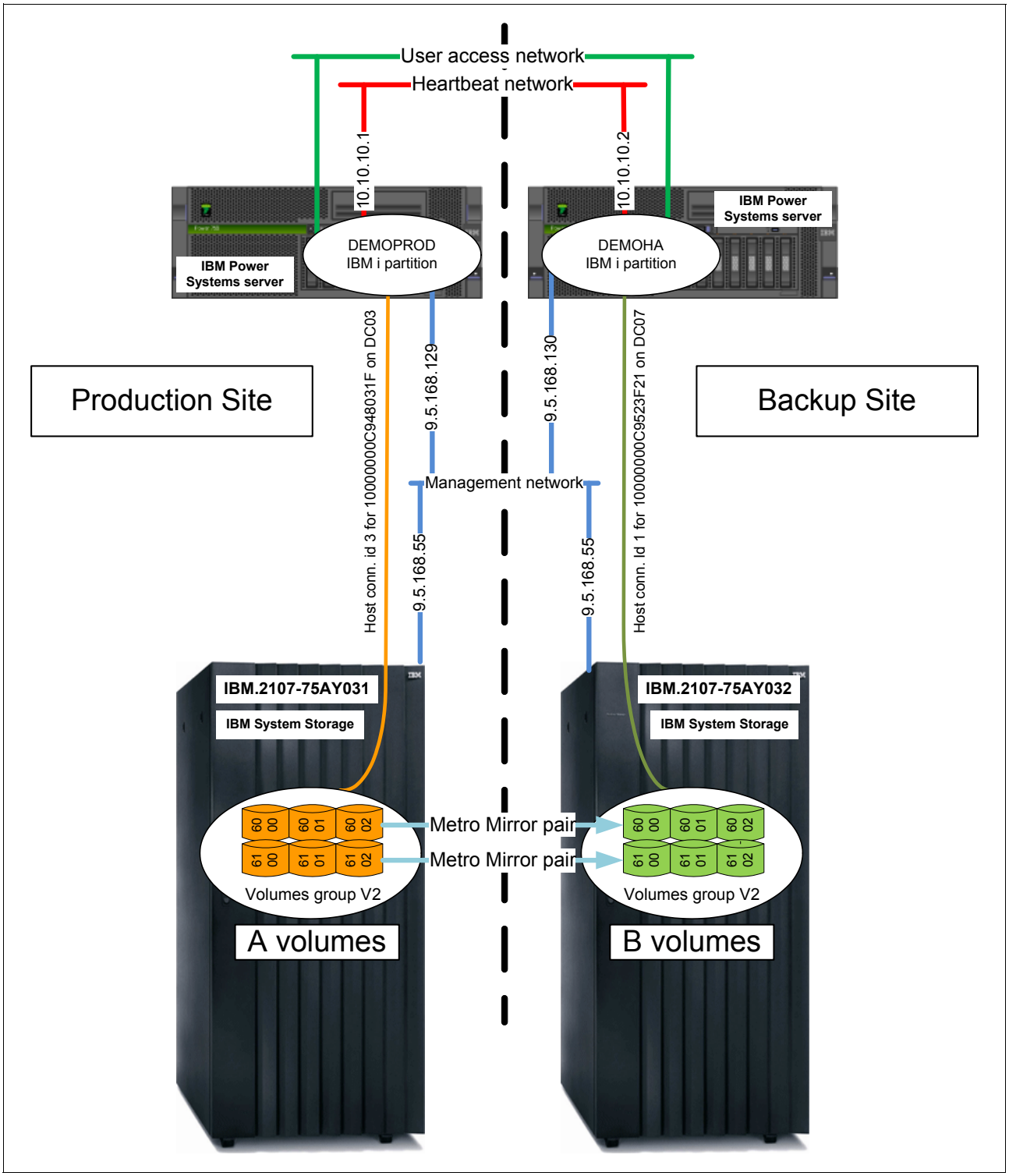


Figure 13-1 Metro Mirror scenario settings

Setting up a Metro Mirror environment

The Metro Mirror environment setup itself is not handled by IBM PowerHA SystemMirror for i. You have to configure the logical storage configuration, including the Copy Services setup, directly on the DS8000. Refer to 6.2, “Metro Mirror” on page 93, for more information.

Creating the PPRC paths

To find the possible links for a path, you can use the DS CLI command **lsavailpprcport**. In our scenario, the source:target LSSs are 60:60 and 61:61. The target WWNN is 5005076307FFCAB8. The I0040 port on the source DS8000 and the I0240 port on the target DS8000 are available. As you can see in Example 13-1, we can create the PPRC paths with **mkpprcpath**.

Example 13-1 Creating PPRC path

```
dscli> lsavailpprcport -l -dev IBM.2107-75AY031 -remotedev IBM.2107-75AY032 -remotewwnn
5005076307FFCAB8 60:60
Date/Time: September 22, 2011 4:27:57 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
Local Port Attached Port Type Switch ID Switch Port
==============================================
I0040      I0240         FCP  NA        NA
dscli> lsavailpprcport -l -dev IBM.2107-75AY031 -remotedev IBM.2107-75AY032 -remotewwnn
5005076307FFCAB8 61:61
Date/Time: September 22, 2011 4:28:02 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
Local Port Attached Port Type Switch ID Switch Port
==============================================
I0040      I0240         FCP  NA        NA
dscli> mkpprcpath -dev IBM.2107-75AY031 -remotedev IBM.2107-75AY032 -remotewwnn 5005076307FFCAB8
-srclss 60 -tgtlss 60 I0040:I0240
Date/Time: September 22, 2011 4:37:44 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
CMUC00149I mkpprcpath: Remote Mirror and Copy path 60:60 successfully established.
dscli> mkpprcpath -dev IBM.2107-75AY031 -remotedev IBM.2107-75AY032 -remotewwnn 5005076307FFCAB8
-srclss 61 -tgtlss 61 I0040:I0240
Date/Time: September 22, 2011 4:37:54 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
CMUC00149I mkpprcpath: Remote Mirror and Copy path 61:61 successfully established.
dscli> lspprcpath -l 60-61
Date/Time: September 22, 2011 4:38:51 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
Src Tgt State   SS   Port  Attached Port Tgt WWNN         Failed Reason PPRC CG
===============================================================================
60  60  Success FF60 I0040 I0240         5005076307FFCAB8 -             Disabled
61  61  Success FF61 I0040 I0240         5005076307FFCAB8 -             Disabled
```

Note: Make sure that the paths are established as *bi-directional*, that is, on both DS8000s from one to the other for the volumes' LSS.

Creating the Metro Mirror relationships

After the path exists from production DS8000 to backup DS8000, we can establish the Metro Mirror relationships using **mkpprc** (Example 13-2). In our Metro Mirror environment, Metro Mirror relationships are to be established between primary volume IDs 6000 - 6002 and secondary volume IDs 6000 - 6002 for a first set, and primary volume IDs 6100 - 6102 and secondary volume IDs 6100 - 6102 for a second set. The relationship is properly established when the status is *full duplex*.

Example 13-2 Creating Metro Mirror relationships

```
dscli> mkpprc -dev IBM.2107-75AY031 -remotedev IBM.2107-75AY032 -type mmir -mode full 6000-6002:6000-6002
6100-6102:6100-6102
Date/Time: September 22, 2011 4:46:48 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
CMUC00153I mkpprc: Remote Mirror and Copy volume pair relationship 6000:6000 successfully created.
CMUC00153I mkpprc: Remote Mirror and Copy volume pair relationship 6001:6001 successfully created.
CMUC00153I mkpprc: Remote Mirror and Copy volume pair relationship 6002:6002 successfully created.
CMUC00153I mkpprc: Remote Mirror and Copy volume pair relationship 6100:6100 successfully created.
CMUC00153I mkpprc: Remote Mirror and Copy volume pair relationship 6101:6101 successfully created.
```

```
CMUC00153I mkpprc: Remote Mirror and Copy volume pair relationship 6102:6102 successfully created.
dscli> lspprc 6000-61FF
Date/Time: September 22, 2011 4:47:12 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
ID        State        Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=================================================================================================
6000:6000 Copy Pending -      Metro Mirror 60        60             Disabled      Invalid
6001:6001 Copy Pending -      Metro Mirror 60        60             Disabled      Invalid
6002:6002 Copy Pending -      Metro Mirror 60        60             Disabled      Invalid
6100:6100 Copy Pending -      Metro Mirror 61        60             Disabled      Invalid
6101:6101 Copy Pending -      Metro Mirror 61        60             Disabled      Invalid
6102:6102 Copy Pending -      Metro Mirror 61        60             Disabled      Invalid
dscli> lspprc 6000-61FF
Date/Time: September 22, 2011 4:50:37 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
ID        State       Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=================================================================================================
6000:6000 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6001:6001 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6002:6002 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6100:6100 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6101:6101 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6102:6102 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
```

Note: For DS8000 releases prior to R4.3, to avoid a possible conflict with Volume Flash Init DS8000 capabilities for volumes in a remote Copy Services relationship, the formatting of volumes by DS8000 and IBM usually done when being added into an ASP must be handled as follows:

1. Initialize and format disks using System Service Tools (STRSST):
   a. Option 3. Work with disk units.
   b. Option 3. Work with disk unit recovery.
   c. Option 2. Disk unit recovery procedures.
   d. Option 1. Initialize and format disk unit option. Notice that this operation is only a marker, which disappears if there is an IPL.
2. Establish the Metro Mirror pair and wait for the source's volumes to become synchronized.
3. Add volumes to the ASP or create the IASP to which volumes are added.

Creating all configuration items with the GUI

The device cluster resource group, device domain, and ASP copy descriptions can be created with the PowerHA GUI using a specific wizard.

As for other PowerHA objects, the creation process starts from the main PowerHA panel:

1. Click **Independent ASPs** (Figure 13-2).

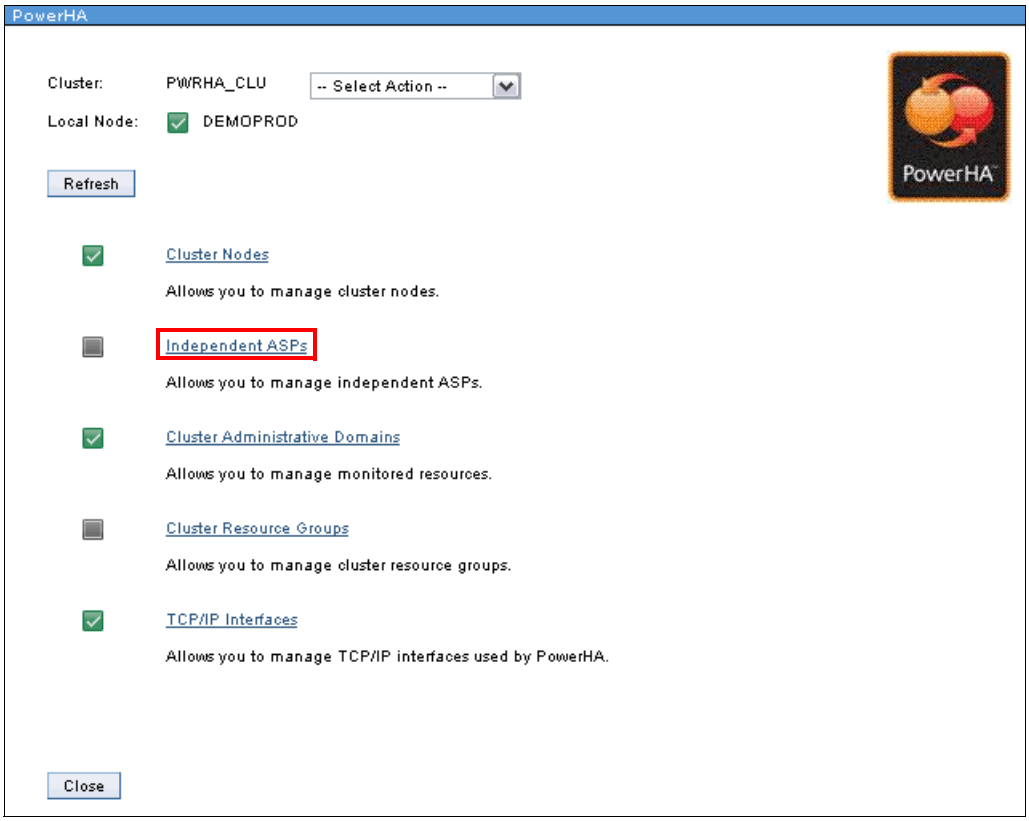


Figure 13-2 Starting Independent ASPs wizard

2. In our case, there is no highly available independent ASP. Click **Show all others** to retrieve all existing IASPs (Figure 13-3).

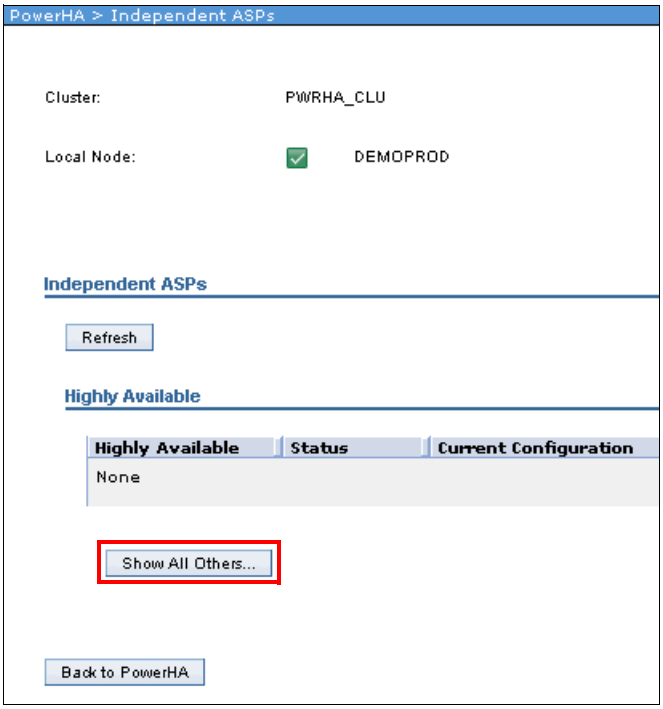


Figure 13-3 Starting Independent ASPs wizard

3. Select **Make Highly Available** for the IASP that you want to work with (Figure 13-4), in our case IASP1.

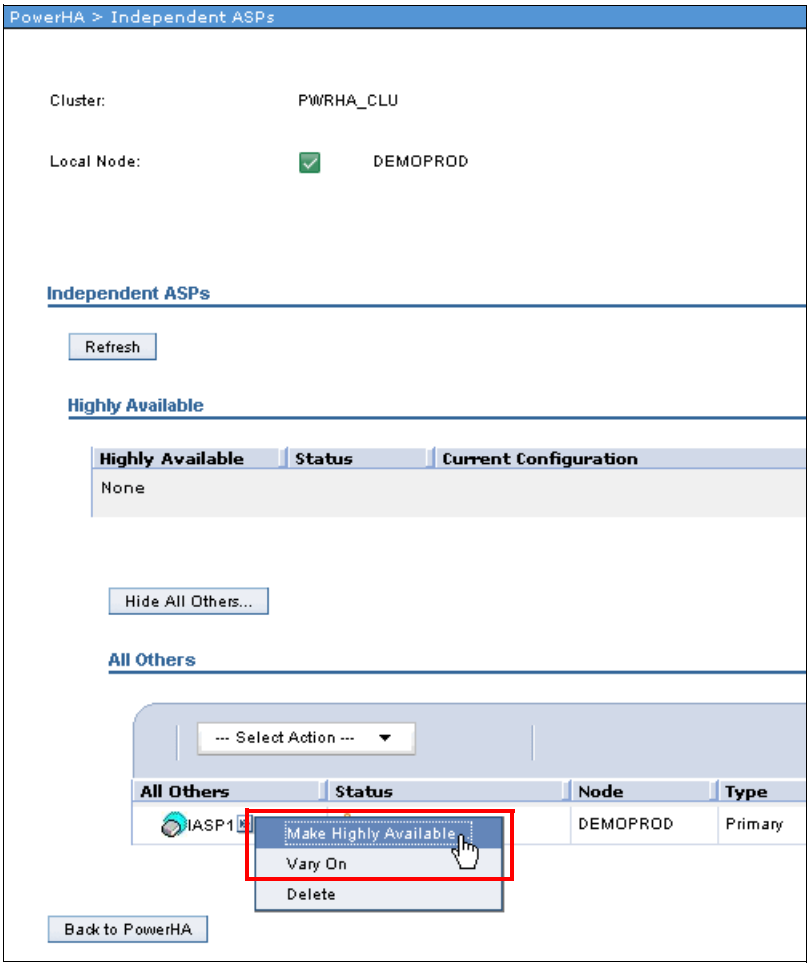


Figure 13-4 Launch the Make Highly Available wizard

4. Click **Next** (Figure 13-5) to start the Make Highly Available wizard.

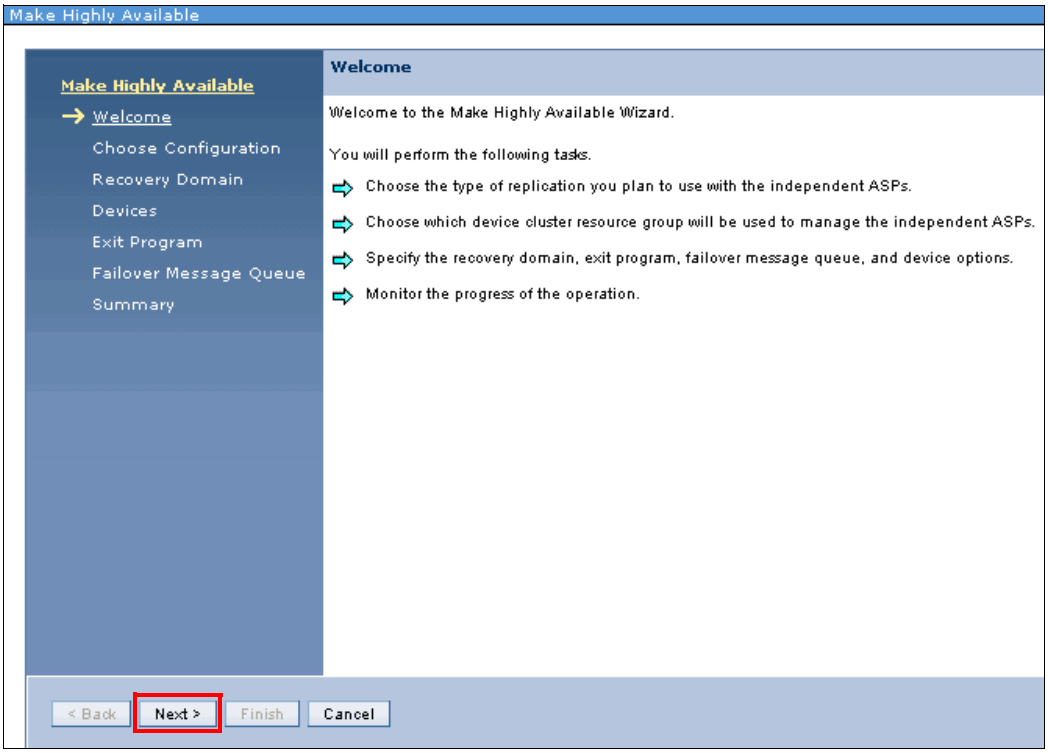


Figure 13-5 Start the Make Highly Available wizard

5. The first step requires providing the type of mirroring to configure and the name of the cluster resource group to be used or created. As shown in Figure 13-6, in our case, we select **Metro Mirror** and provide the CRG name PWRHA_CRG1.

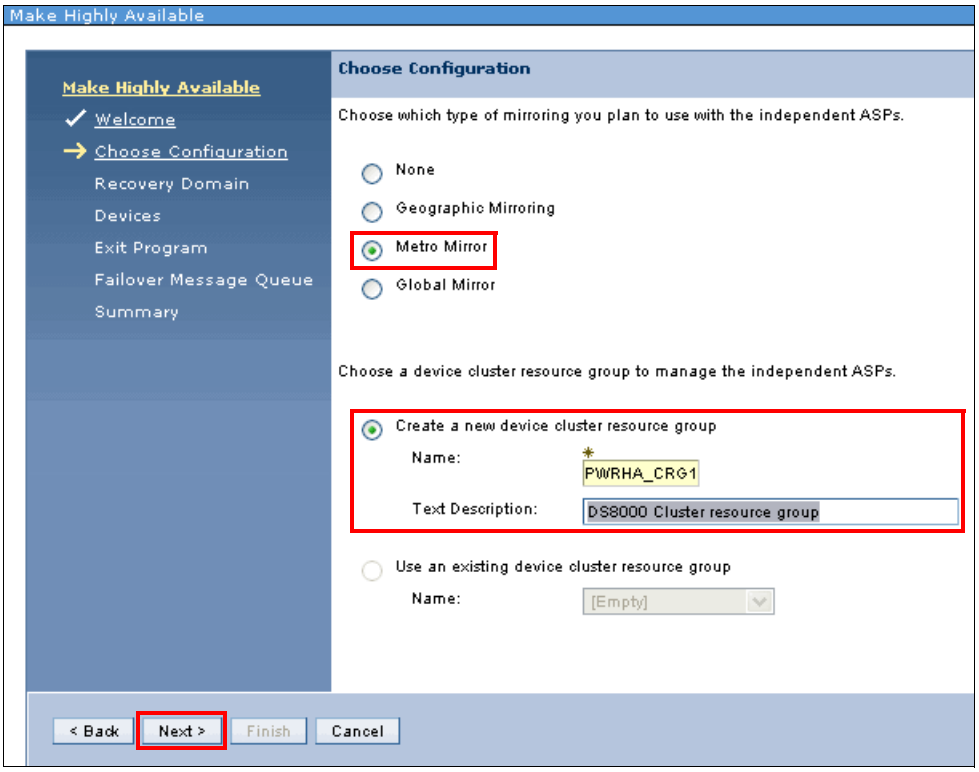


Figure 13-6 Specify type of mirroring and name of cluster resource group

6. As shown in Figure 13-7, the next step is to add the primary node into the recovery domain. In our case, DEMOPROD is this first node. Select the appropriate node in the list and provide a site name, SITE1 in our case, and click **Add**.

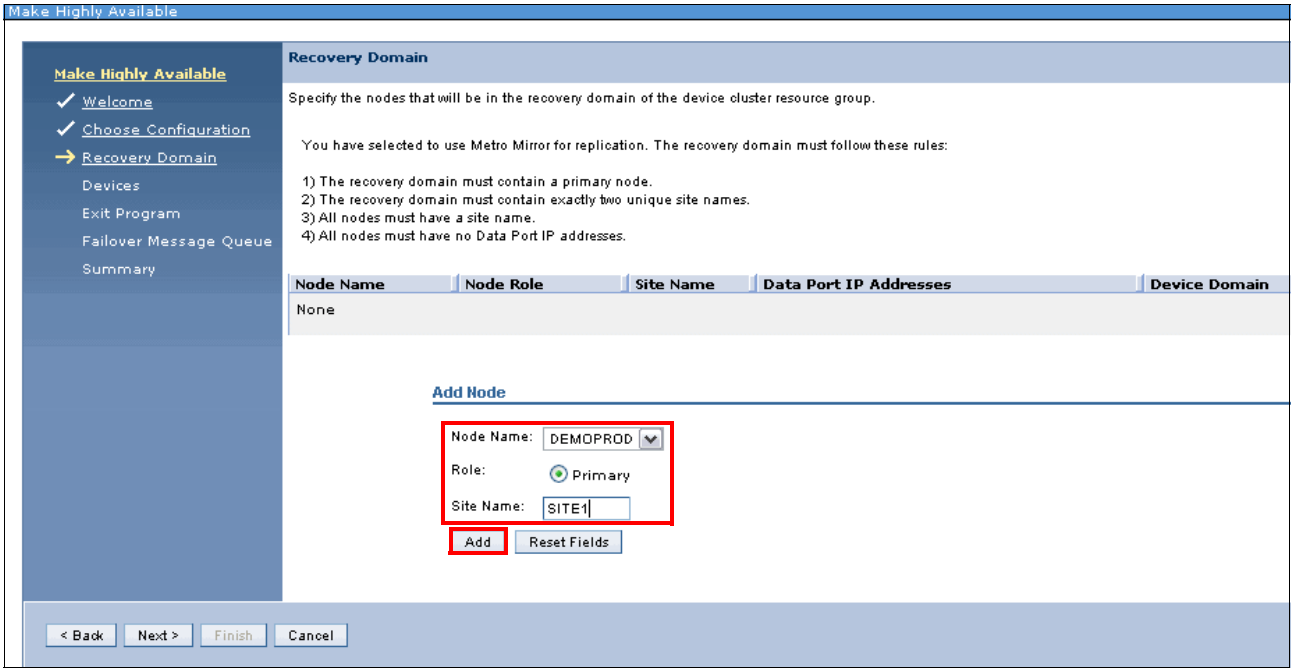


Figure 13-7 Add the primary node to the cluster resource group

7. We continue by adding all necessary nodes to the cluster resource group. In our case DEMOHA, which our backup node one, is located at the backup site with a site name of SITE2 (Figure 13-8).

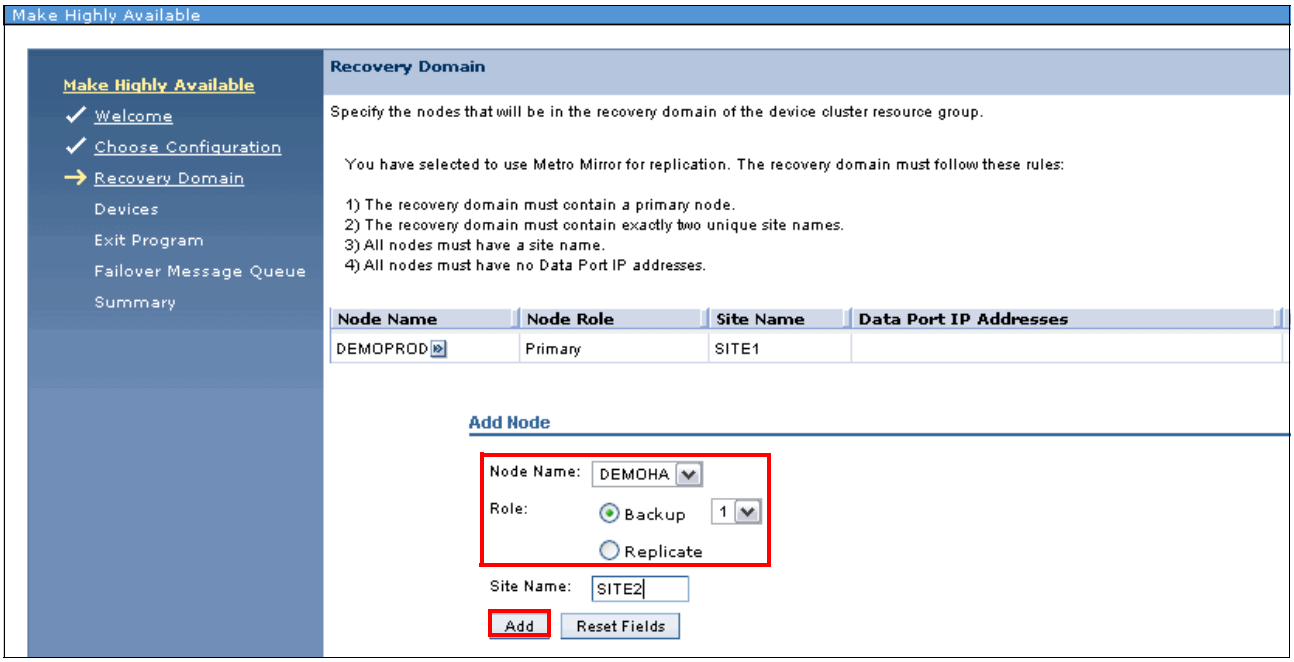


Figure 13-8 Add the second node to cluster resource group

8. After all nodes exist in the cluster resource group (Figure 13-9), click **Next** to continue.

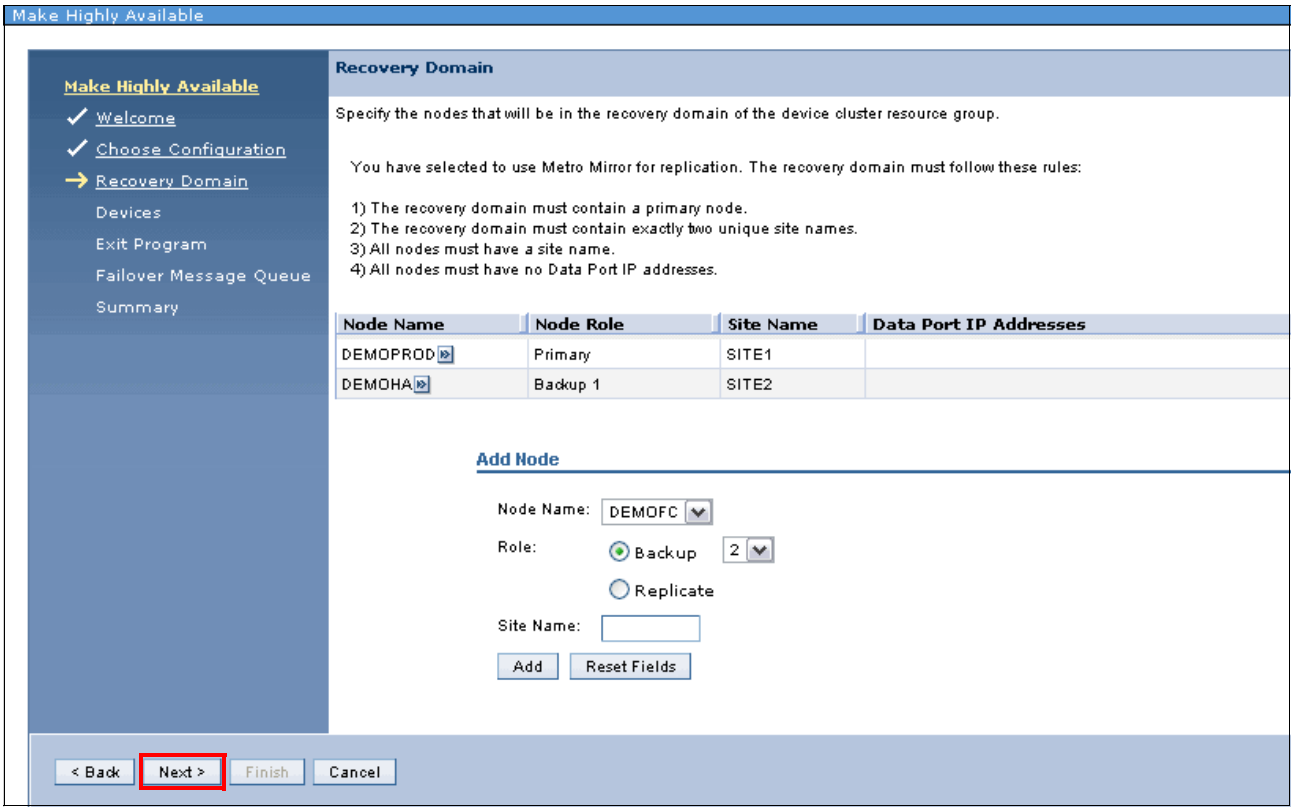


Figure 13-9 Continue when all nodes are in the cluster resource group

9. All nodes in a cluster resource group whose type is device must also be members of a *device domain*. There is no starting point for creating a device domain in the PowerHA GUI. Device domains are created by giving them a name (Figure 13-10). If there is already an appropriate existing device domain, we can select it from the drop-down list. In our environment, no device domain exists. Giving it a name and then clicking **OK** creates it.

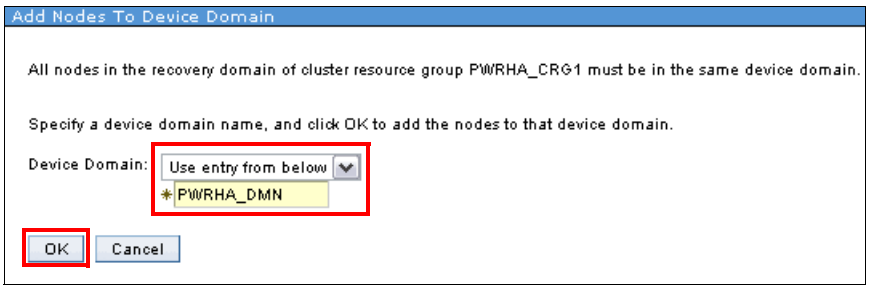


Figure 13-10 Create or select a device domain

10. As shown in Figure 13-11, each node is added to the device domain. Independent ASP and ASP device description creation occur at this time for all newly added nodes but the first one.

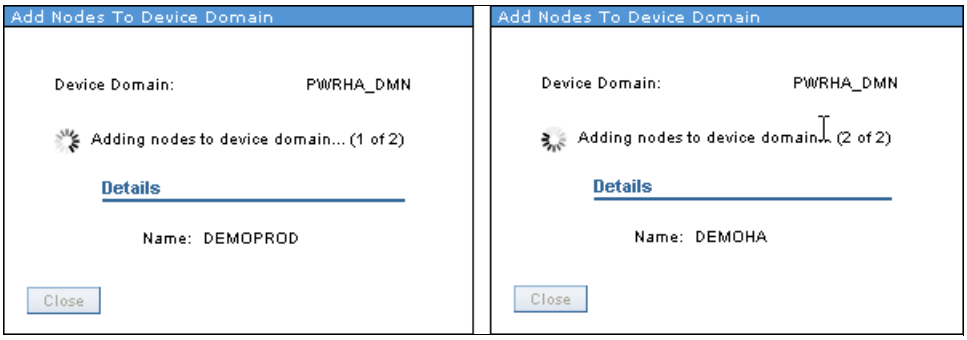


Figure 13-11 Nodes being added to device domain

11. When all nodes are included (Figure 13-12), we can click **Close** to continue.

Note: If PowerHA does not succeed to add all nodes into the device domain, you receive a Cancel/Retry panel. This provides you with the opportunity to analyze the reason for the failure, solve the cause, and retry adding nodes into the device domain.

Note: At this time, the cluster resource group is not yet created, but the device domain has been created and nodes have been included.

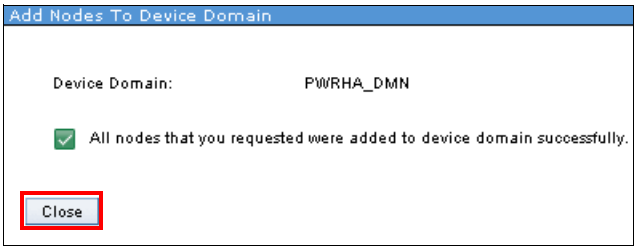


Figure 13-12 Nodes successfully added to device domain

12. After each node is a member of the same device domain (Figure 13-13), we can continue by clicking **Next**.

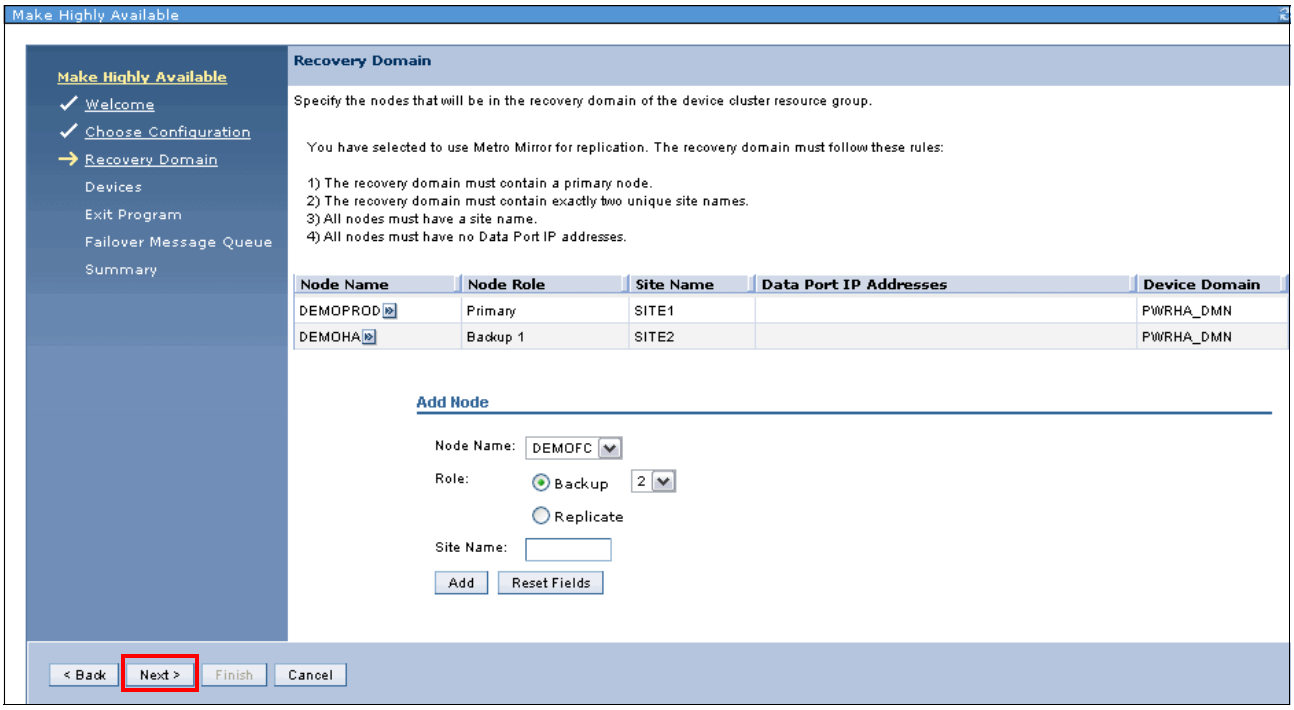


Figure 13-13 All nodes are included in both a device domain and the cluster resource group

13. The next step is to launch the device configuration wizard. As shown in Figure 13-14, select **Modify** for the device that you are working with (in our environment, IASP1).

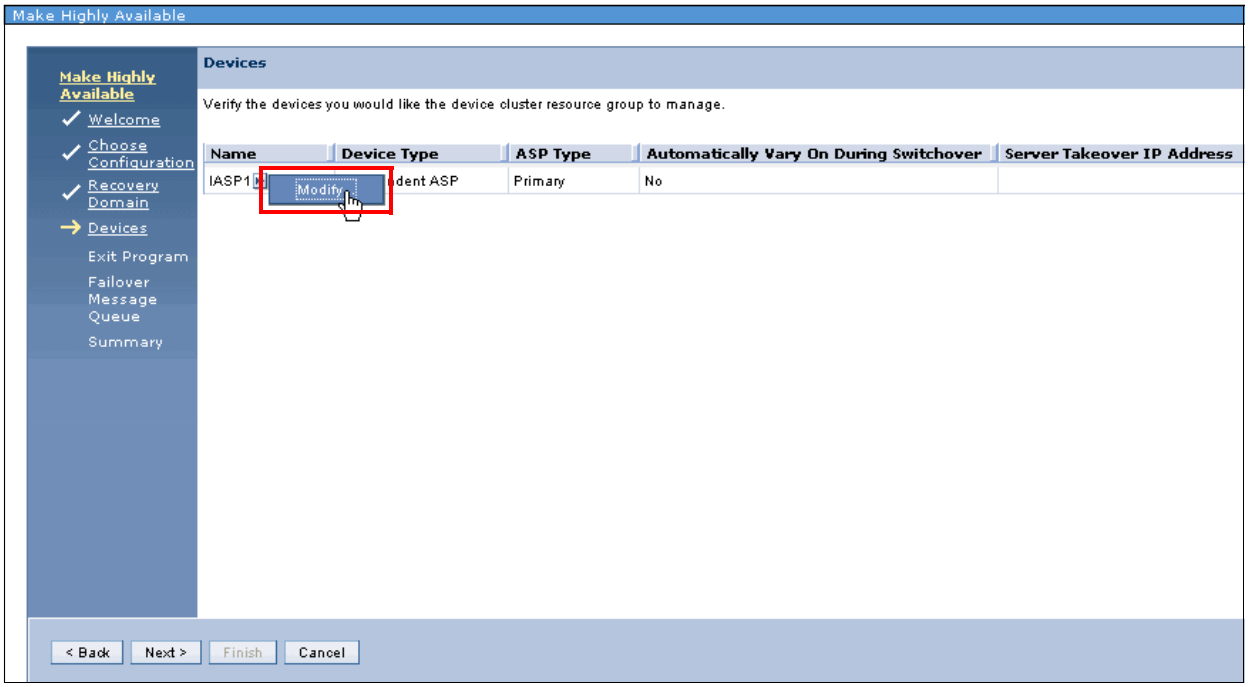


Figure 13-14 Launch the device configuration

14. As shown in Figure 13-15, you can select the automatic vary-on during a switchover and specify the server takeover IP address (the one that the users connect to). When done, click **Save**.

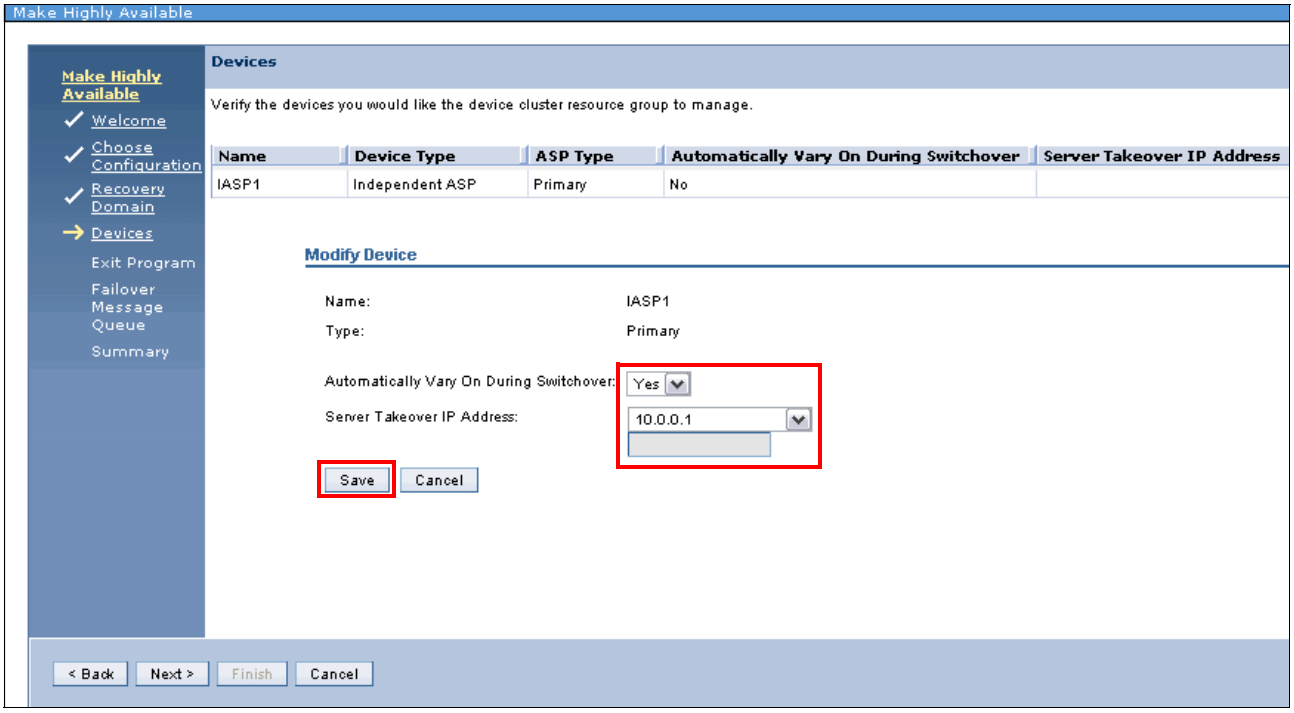


Figure 13-15 IASP settings

15. After the device is configured (Figure 13-16), click **Next** to continue.

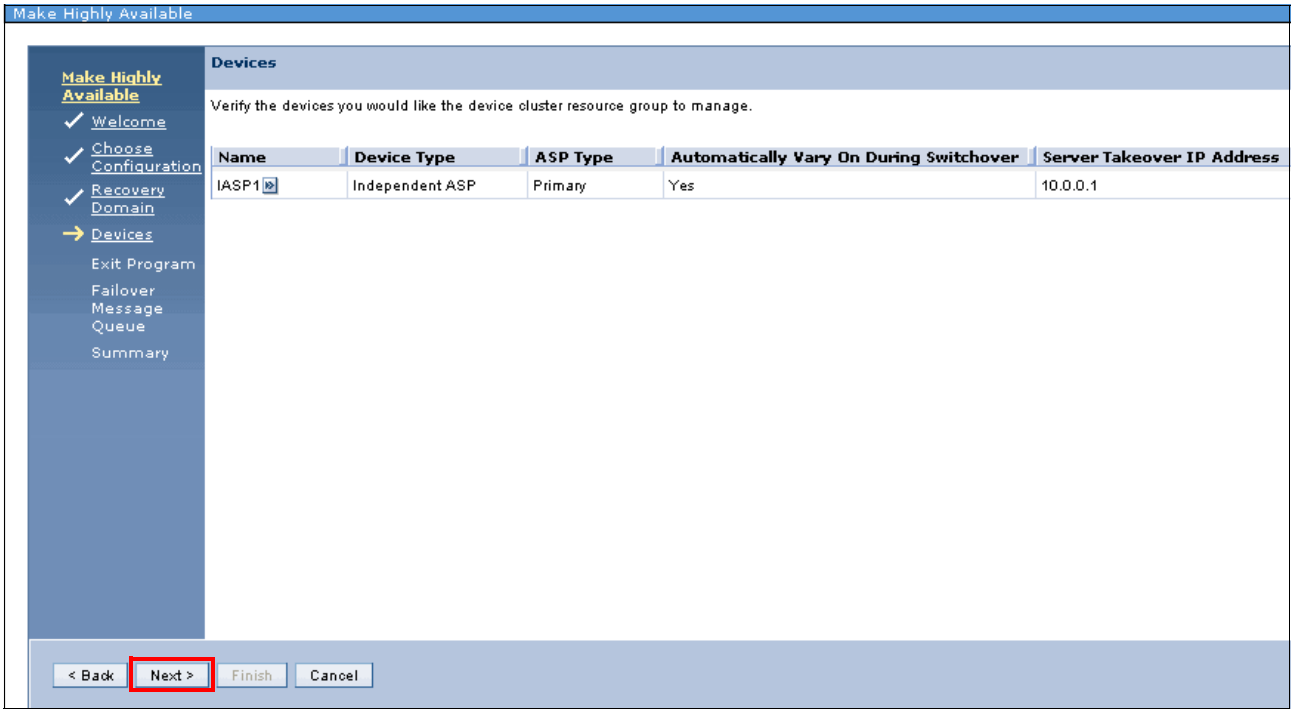


Figure 13-16 Continue when the device is configured

16. The next panel is related to providing the name of an exit program that might handle cluster resource group events like failover. In our case, there is no such program. We skip this step by clicking **Next** (Figure 13-17).

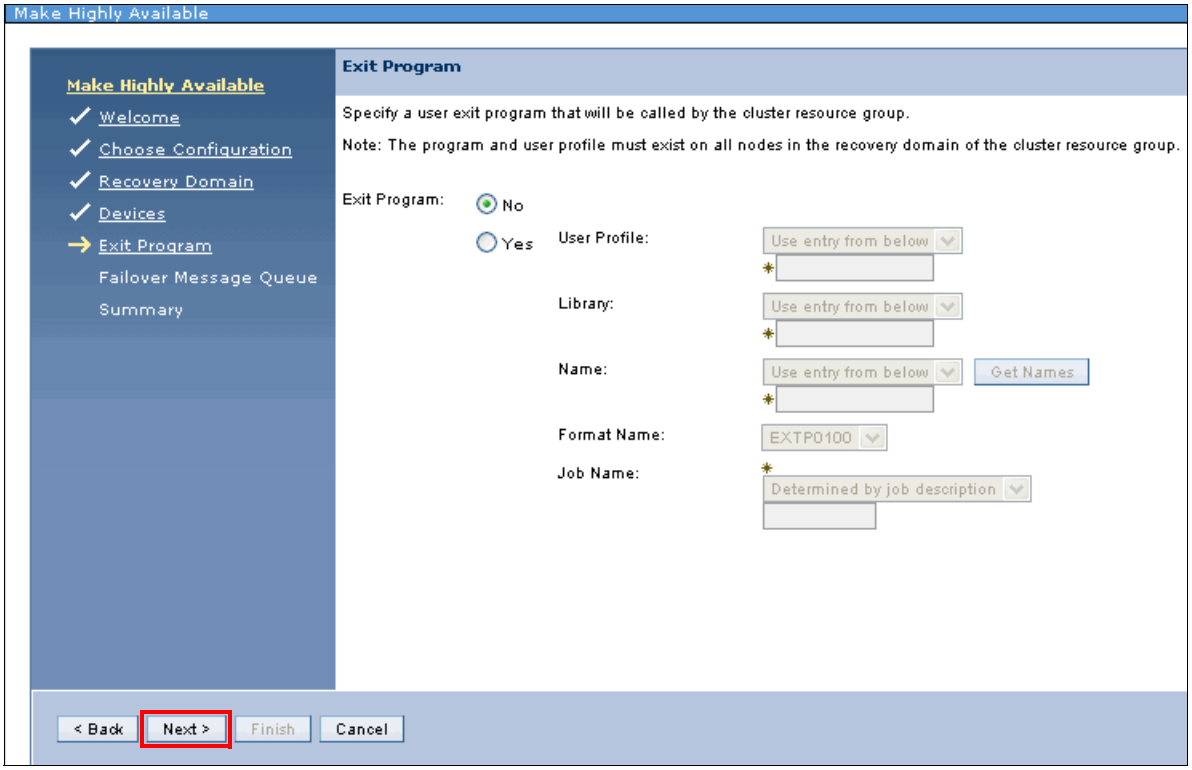


Figure 13-17 Specify an optional exit program

17. The last wizard panel is related to providing information for a *failover message queue*. As shown in Figure 13-18, for our environment we do not provide a CRG failover message queue, and just have to click **Next** to go to the summary wizard panel. If no CRG message queue is specified here, the cluster message queue is used for all CRGs in the device domain.

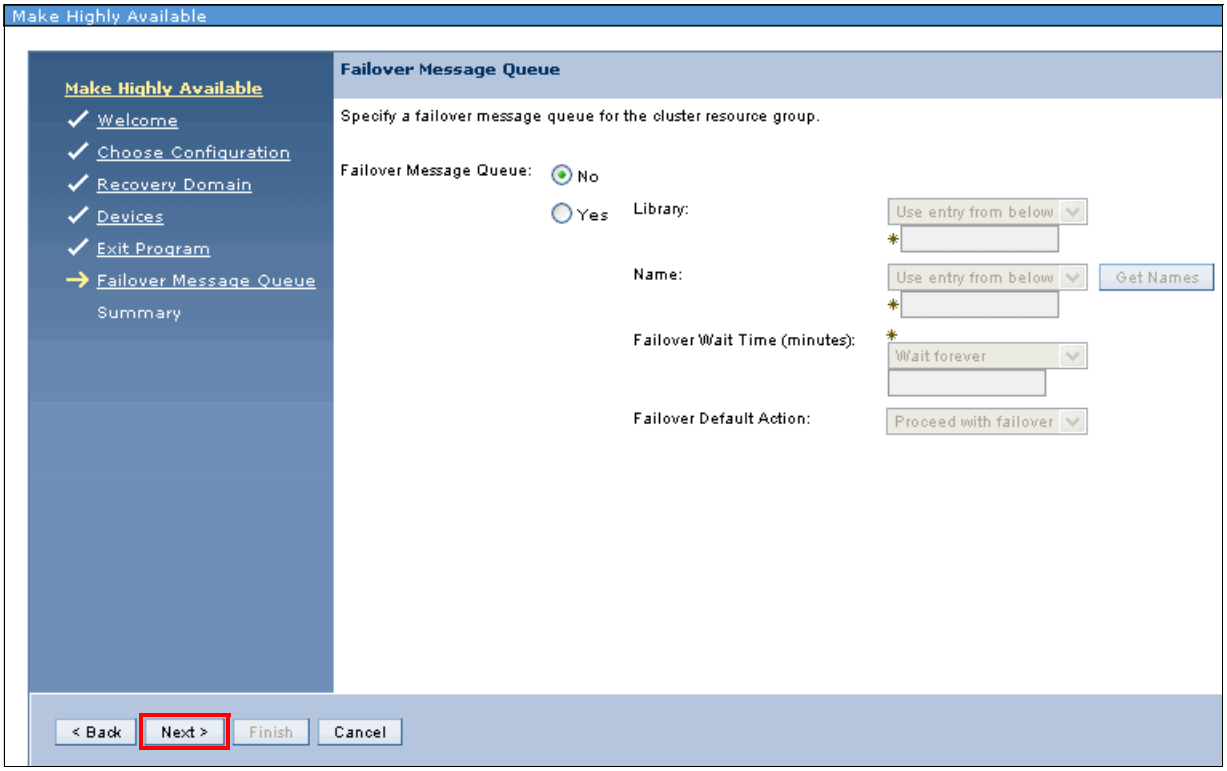


Figure 13-18 Specify an optional failover message queue

18. The last panel is a summary allowing you to review the provided information and go back to the related panel if something is wrong. Click **Finish** to create the CRG (Figure 13-19).

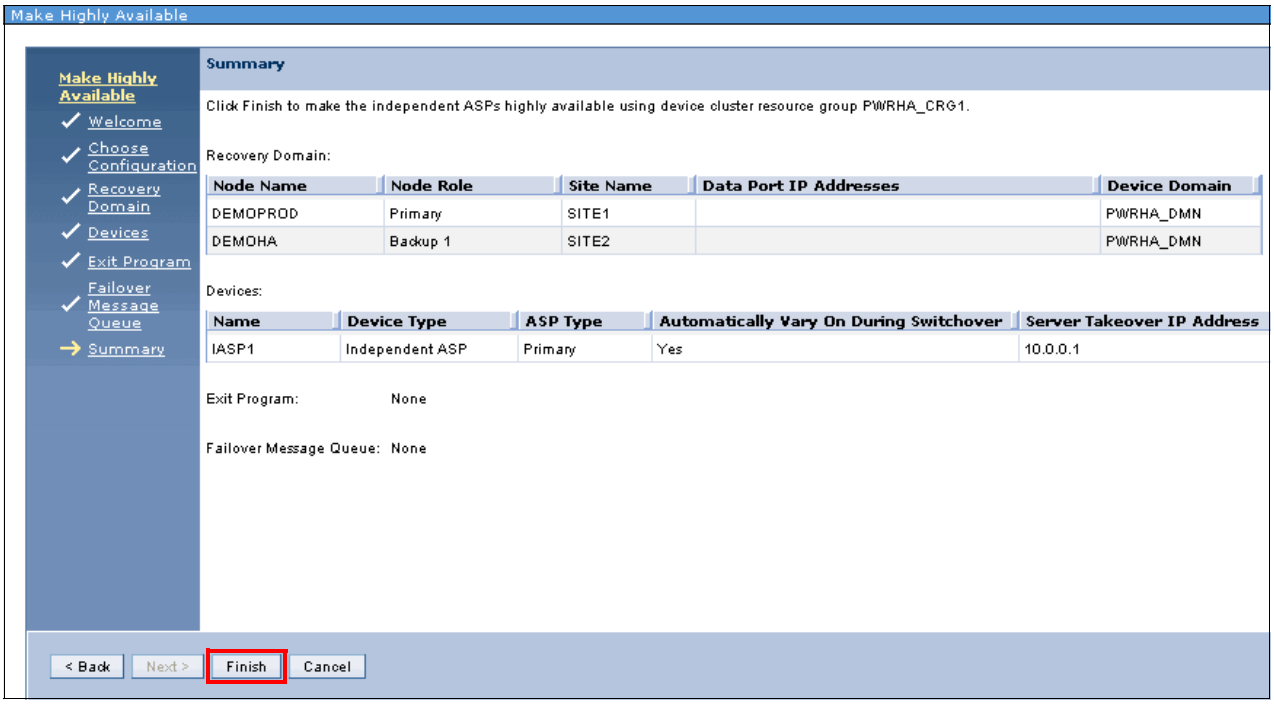


Figure 13-19 Finish the operation

The wizard displays several panels with the progress indicator (Figure 13-20).

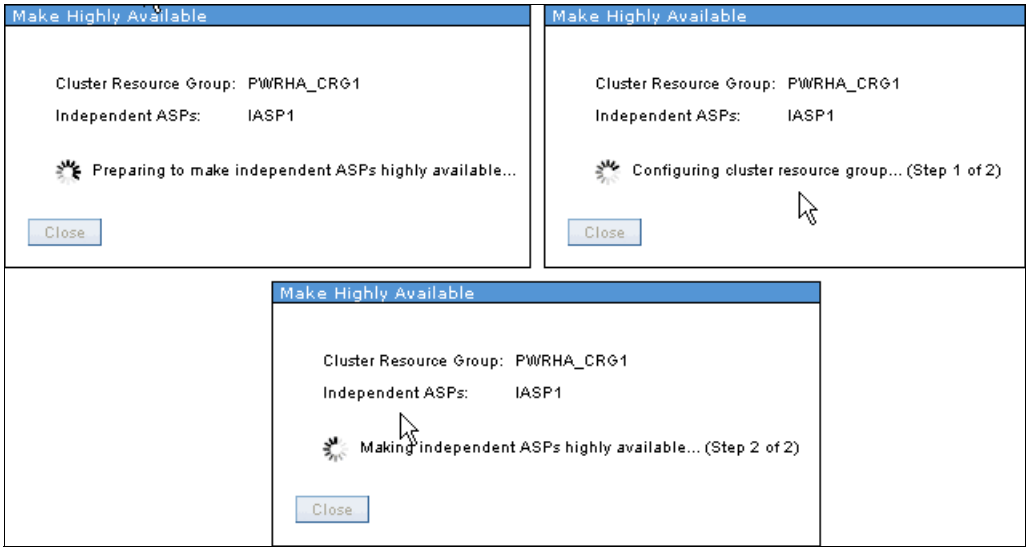


Figure 13-20 Cluster resource group is being created

19. At the end of this step, the cluster resource group has been created and the device domain is configured. Clicking **Close** (Figure 13-21) ends this part of the wizard.

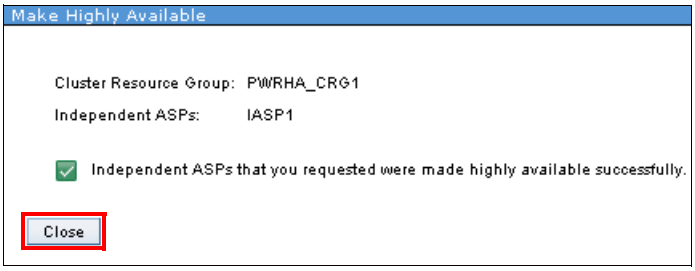


Figure 13-21 Cluster resource group is created

20. The next panel asks whether you want to continue the wizard to create the ASP copy descriptions for both independent ASP copies included in the cluster resource group. Click **Yes** to launch the Configure Mirroring wizard (Figure 13-22).

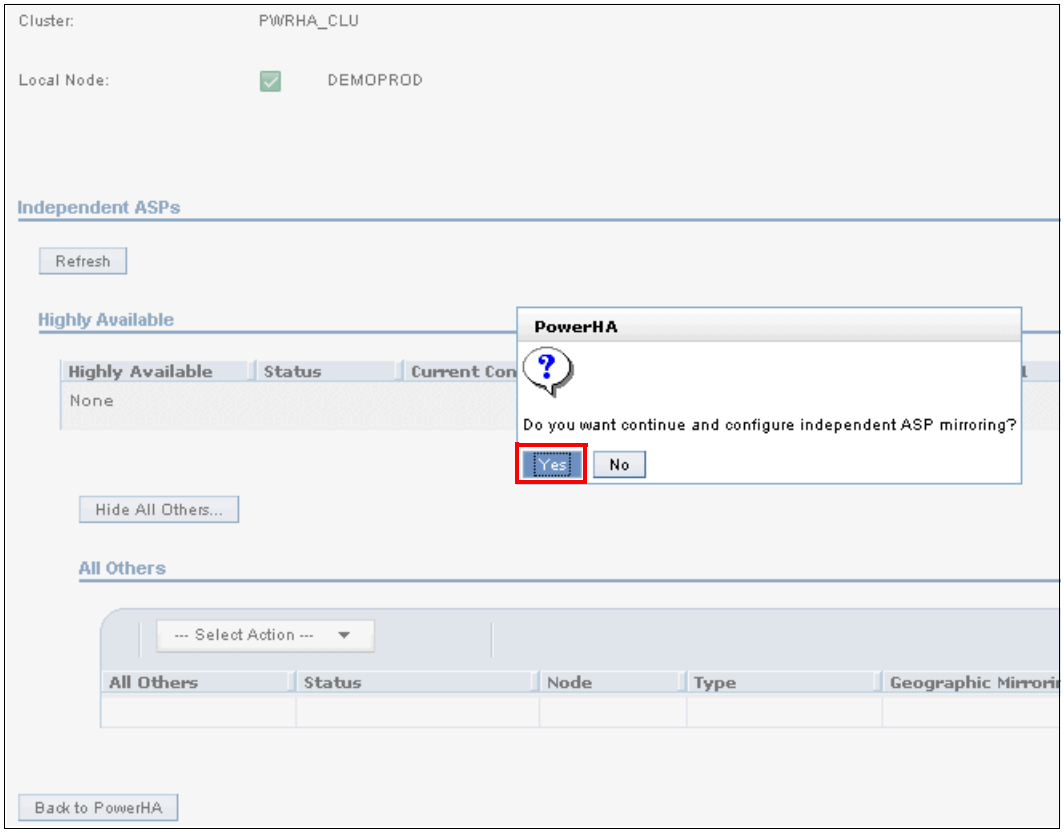


Figure 13-22 Question for configuring mirroring

Note: If you do not click Yes at this time, you can still create the ASP copy descriptions later through the PowerHA GUI by using the Configure mirroring option on the Independent ASP details panel (Figure 13-23 on page 283). You will then use the same wizard described in step 21 on page 284.

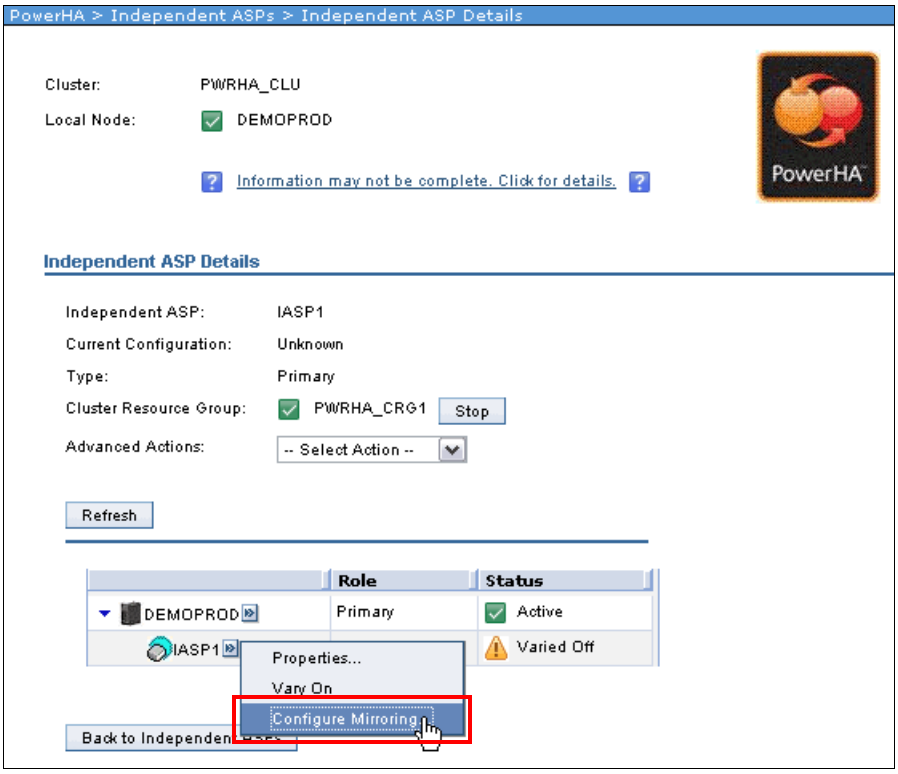


Figure 13-23 Select Configure Mirroring from IASP Details panel

21. On the Configure Mirroring Welcome panel, click **Next** to start the wizard.

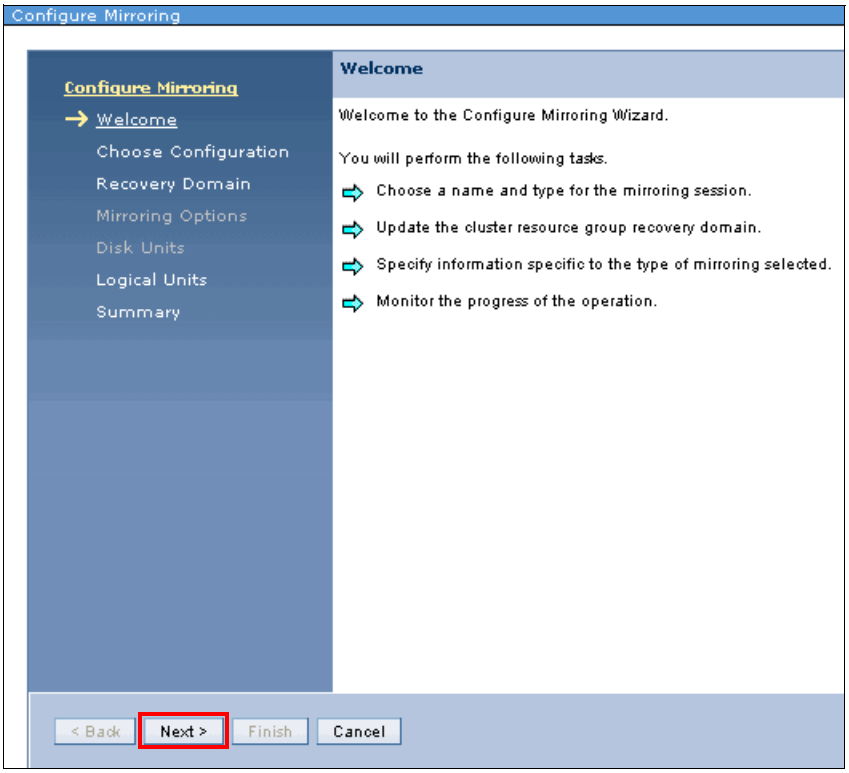


Figure 13-24 Starting Configure Mirroring wizard

22. The first panel is related to choosing the mirroring configuration. As we come from a previous wizard where we decided to use Metro Mirror, no other selection is possible. As shown in Figure 13-25, we provide the name of the new *ASP session*, IASP1_MM in our case, then we click **Next** to continue.

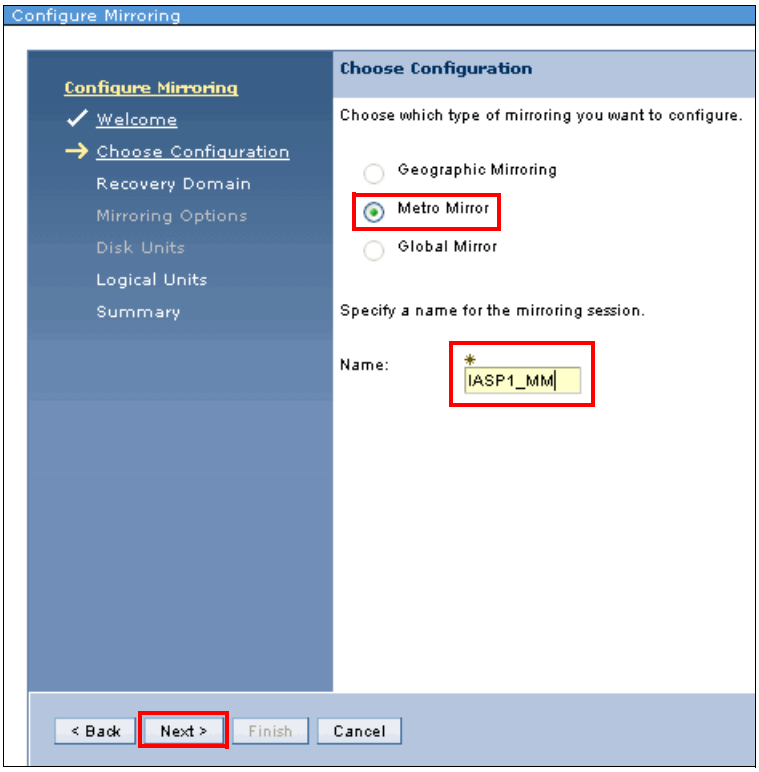


Figure 13-25 Choose configuration

23. The recovery domain information is already populated with values from the cluster resource group. Click **Next** (Figure 13-26).

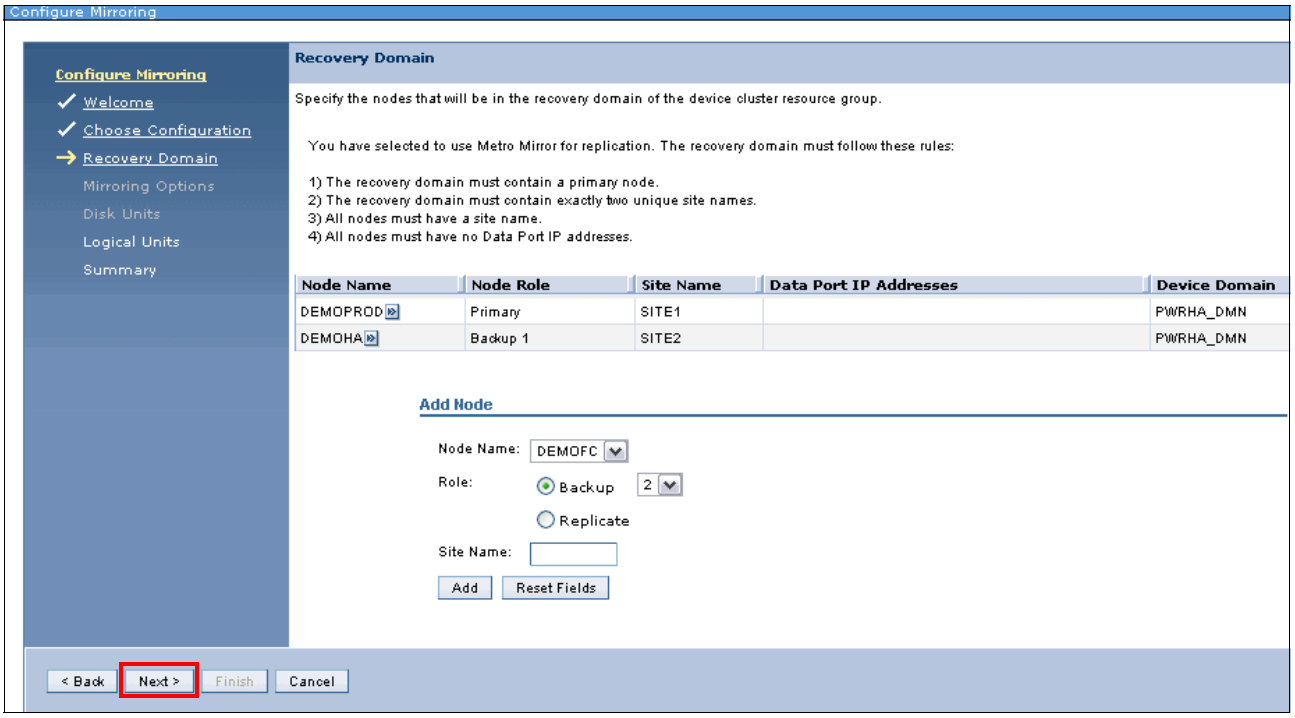


Figure 13-26 Recovery domain reminder

24. The next step is related to creating ASP copy descriptions for both the source copy and the target copy. As shown in Figure 13-27, select **Modify Source Copy** for the IASP (in our case, this is still IASP1).

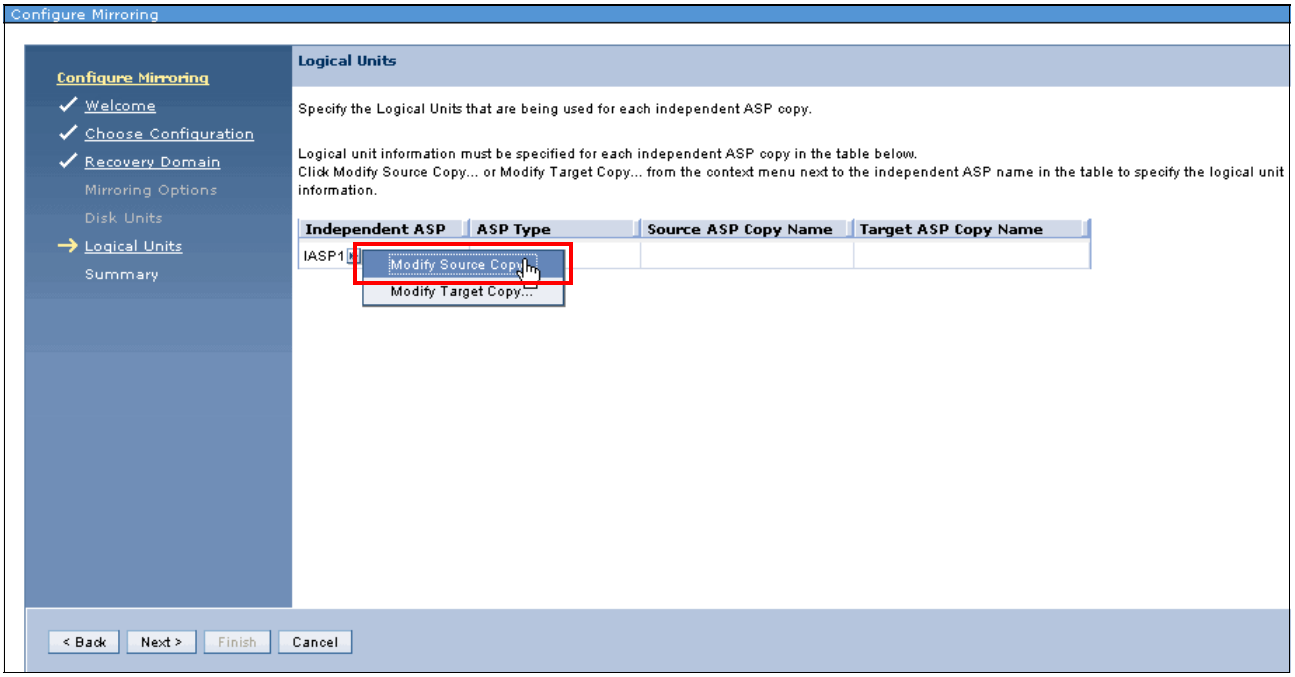


Figure 13-27 Beginning of copy descriptions creation

25. On the panel shown in Figure 13-28, you will, at the same time, either select or provide the copy description name (in our case we provide IASP1_MM1), provide the DS8000 device ID (IBM.2107-75AY031), the user profile (powerha), the password, the IP address (9.5.168.55), and the first range of volumes (6000 - 6002). Click **Add** to validate the panel.

> **Note:** For the DS8000 user ID and password, make sure that your browser settings do not override the information that you provide with one that could already exist in the saved passwords for Mozilla Firefox or AutoComplete settings for Microsoft Internet Explorer, for example.

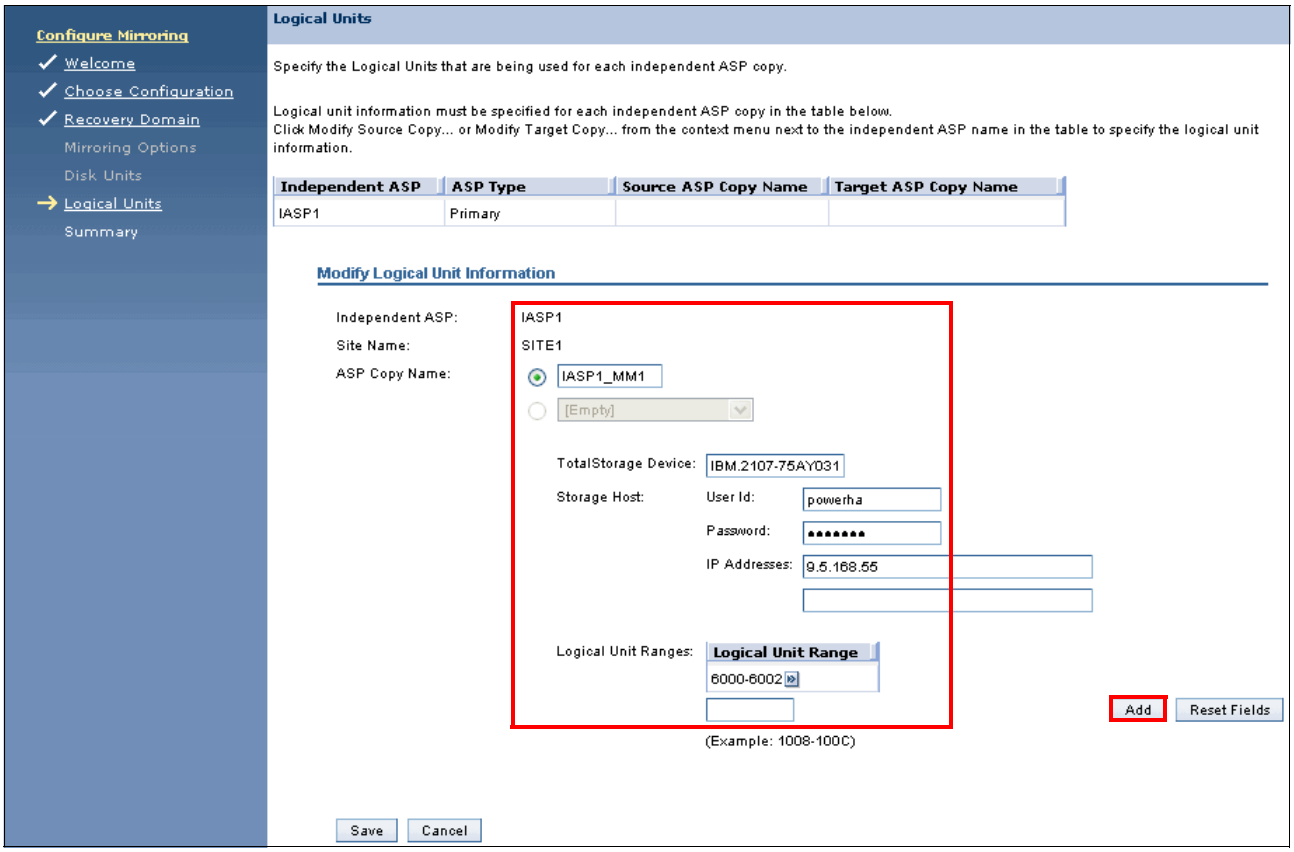


Figure 13-28 Specify first logical unit range for source copy description

26. You will need to add the logical unit ranges as needed. For each new range (in our case 6100-6102), click **Add**. When a range is not possible for a single volume, you still have to enter it as a range, for example, 6000-6000.

27. When all logical unit ranges are OK, click **Save** to validate the source copy description (Figure 13-29).

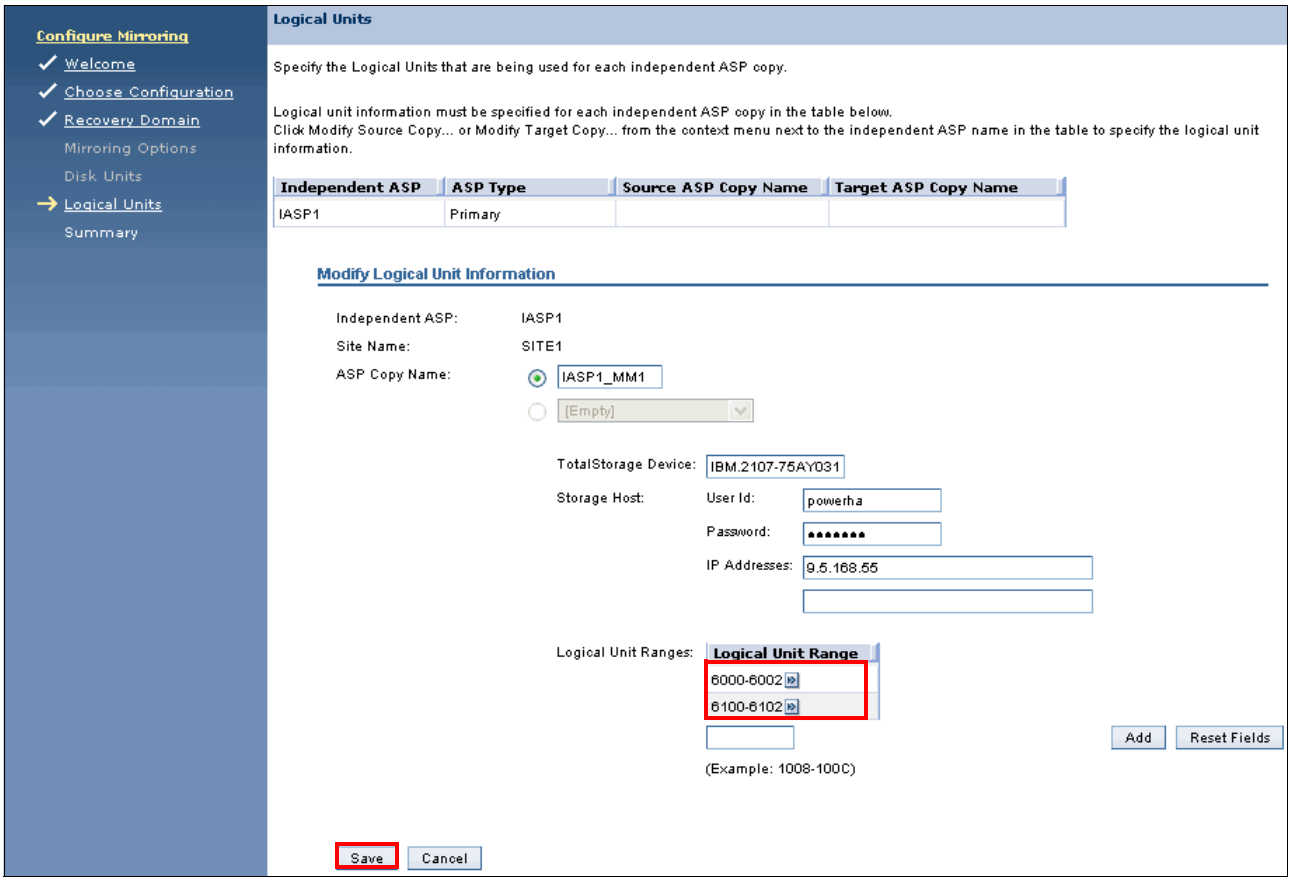


Figure 13-29 Validate logical unit ranges for source copy description

28. After the source copy description, we have to handle the target copy description creation. The steps are the same as for the source copy description. Only the information differs:

    a. As shown in Figure 13-30, we select **Modify Target Copy** against our IASP.

    b. We provide the required information for the target copy description (Figure 13-31 on page 290). In our case, the target copy description name is IASP1_MM2. It applies to DS8000 device ID IBM.2107-75AY032. We connect to this DS8000 with the user profile powerHA at the IP address 9.5.168.55. The first logical unit range is 6000 - 6002 and the second one is 6100 - 6102.

    c. When all logical volume ranges are OK, click **Save** to continue (Figure 13-32 on page 291).

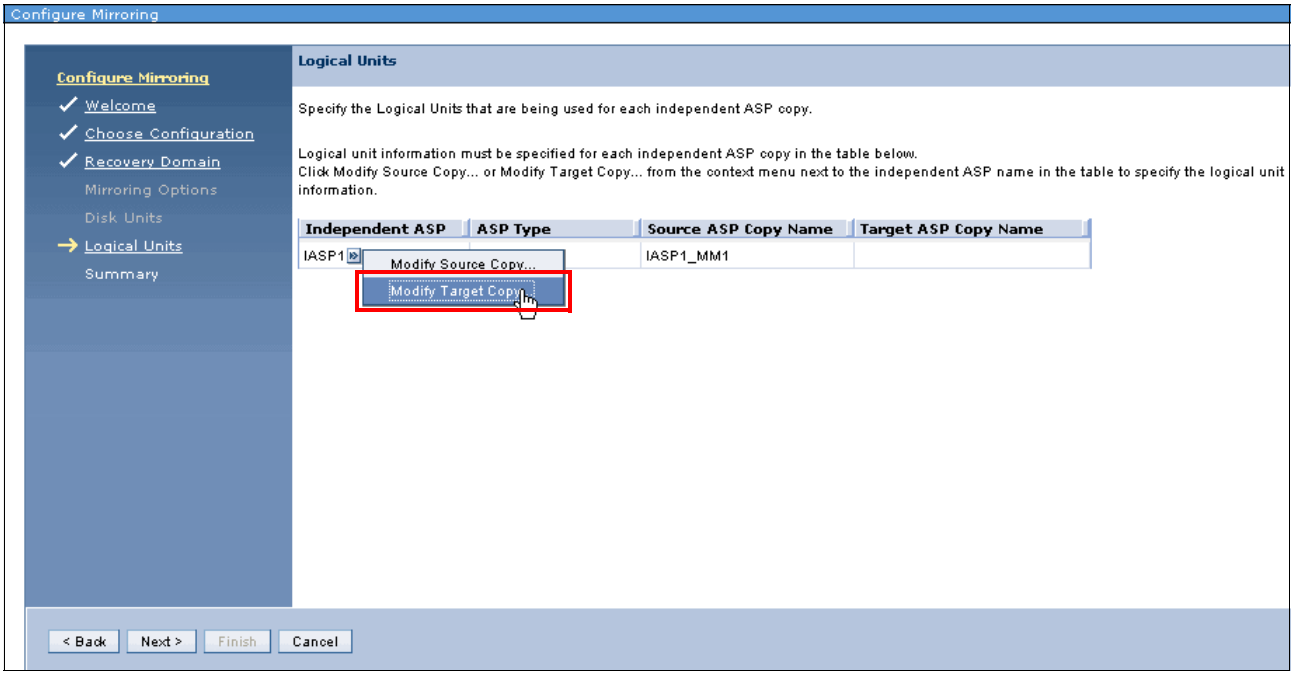


Figure 13-30 Continue with target copy description creation

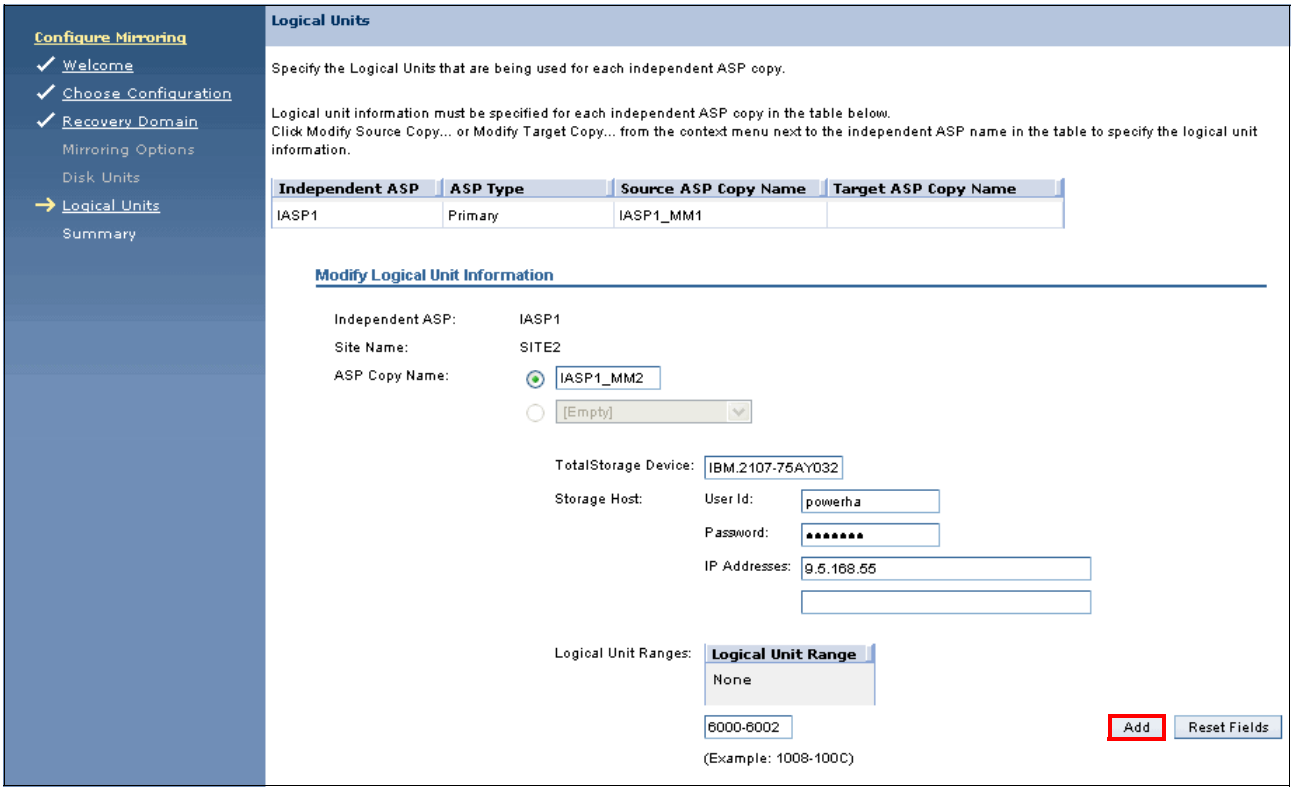


Figure 13-31 Specify first logical unit range for target copy description

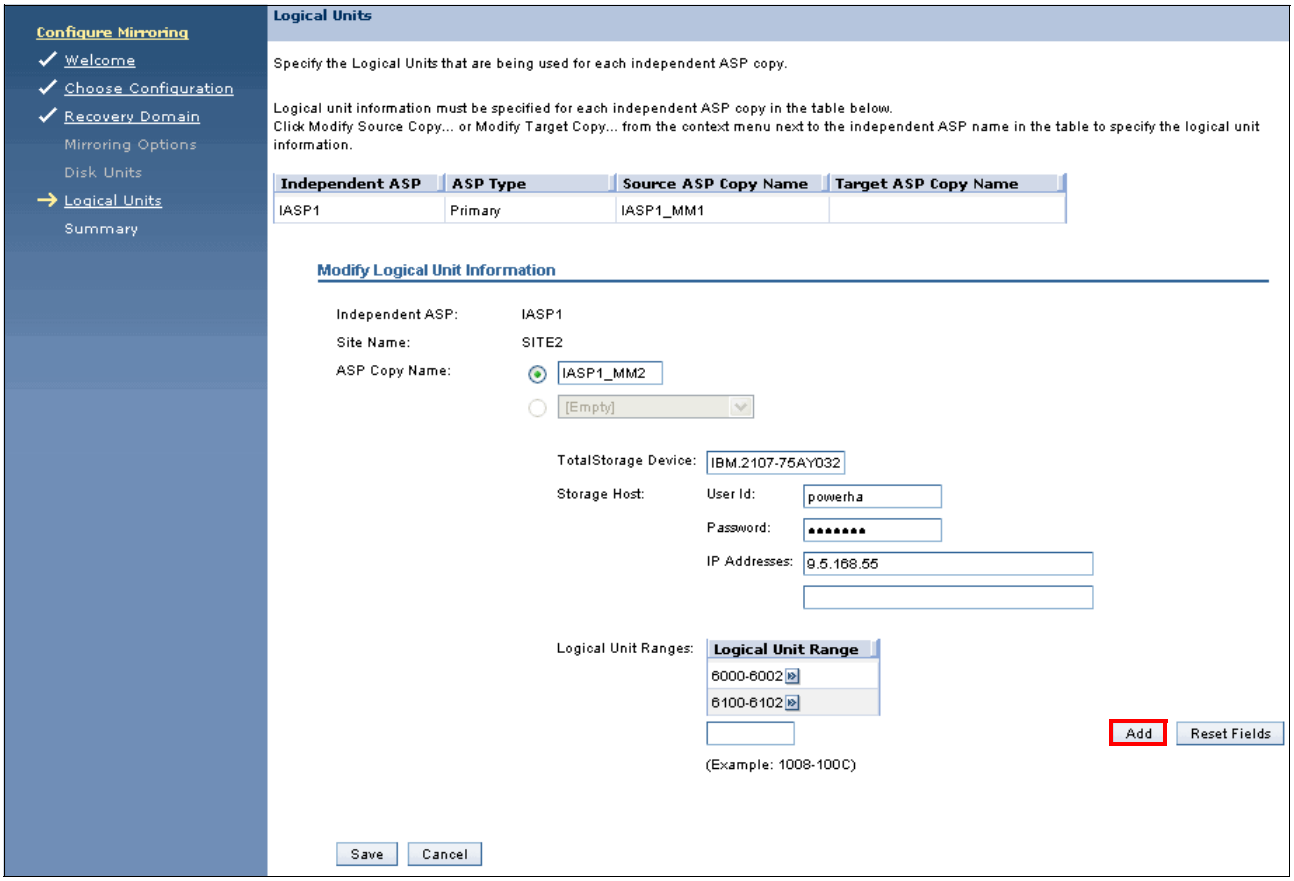


Figure 13-32 Validate logical unit ranges for target copy description

29. When both source and target copy descriptions are complete (Figure 13-33), click **Next** to review the configuration.

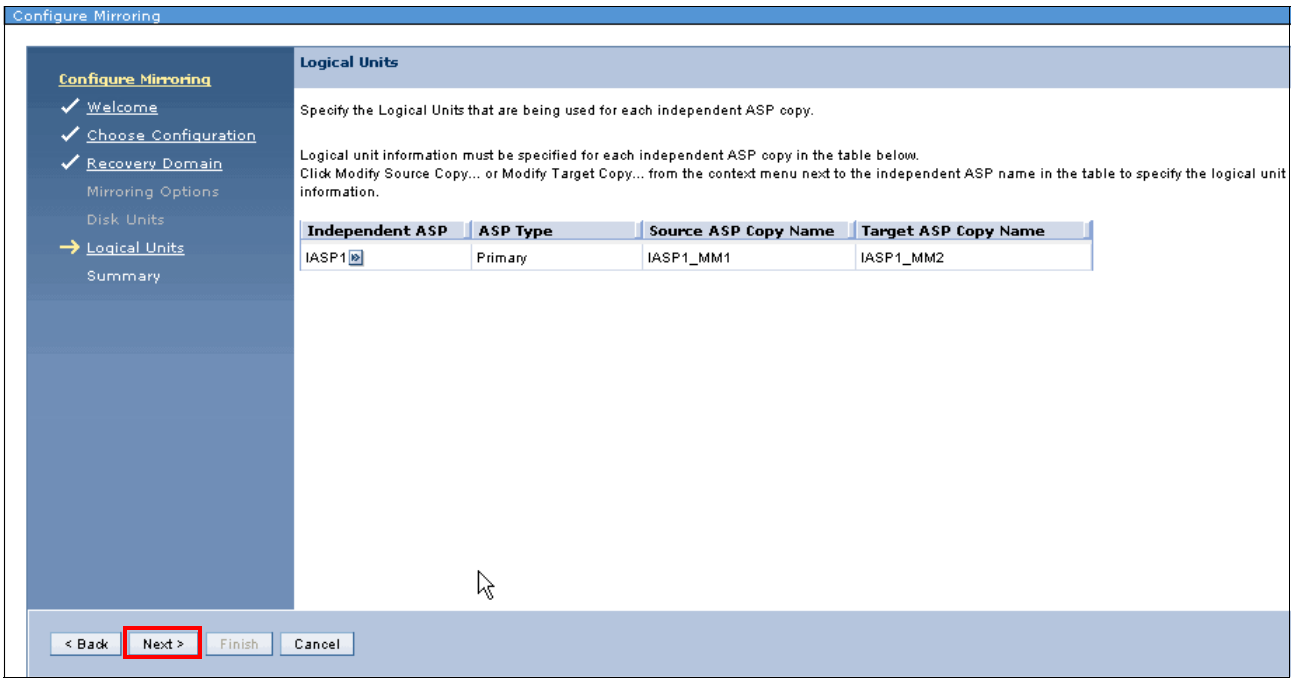


Figure 13-33 Copy descriptions are complete

30. When you reviewed your intended configuration, click **Finish** to proceed with the creation of the ASP copy descriptions (in our case IASP1_MM1 and IASP1_MM2) and the starting of the ASP session for Metro Mirror (in our case IASP1_MM) (Figure 13-34).

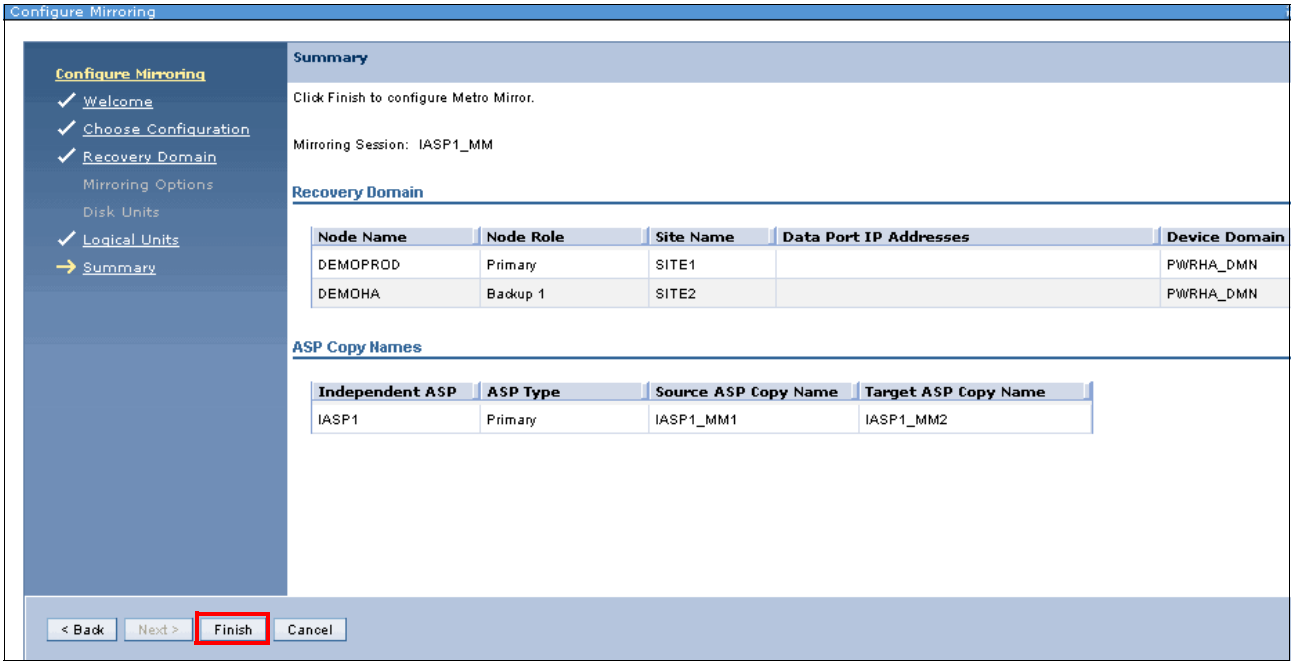


Figure 13-34 Submit copy descriptions creation

31. The wizard displays several panels with progress indicator (Figure 13-35).

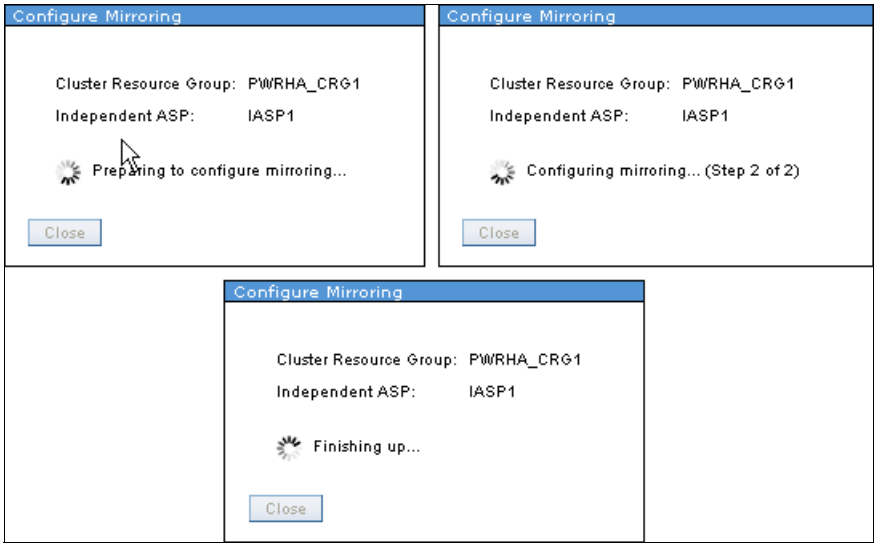


Figure 13-35 Copy descriptions are being created and session started

32. The cluster resource group, device domain, and copy descriptions have been created and the CRG is ready to be started when needed. Click **Close** to go back to the wizard starting point (Figure 13-36). Independent ASP is now highly available and ready to be varied on (Figure 13-37).

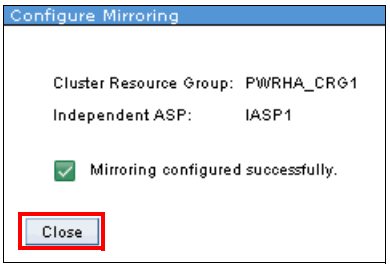


Figure 13-36 Mirroring configuration complete

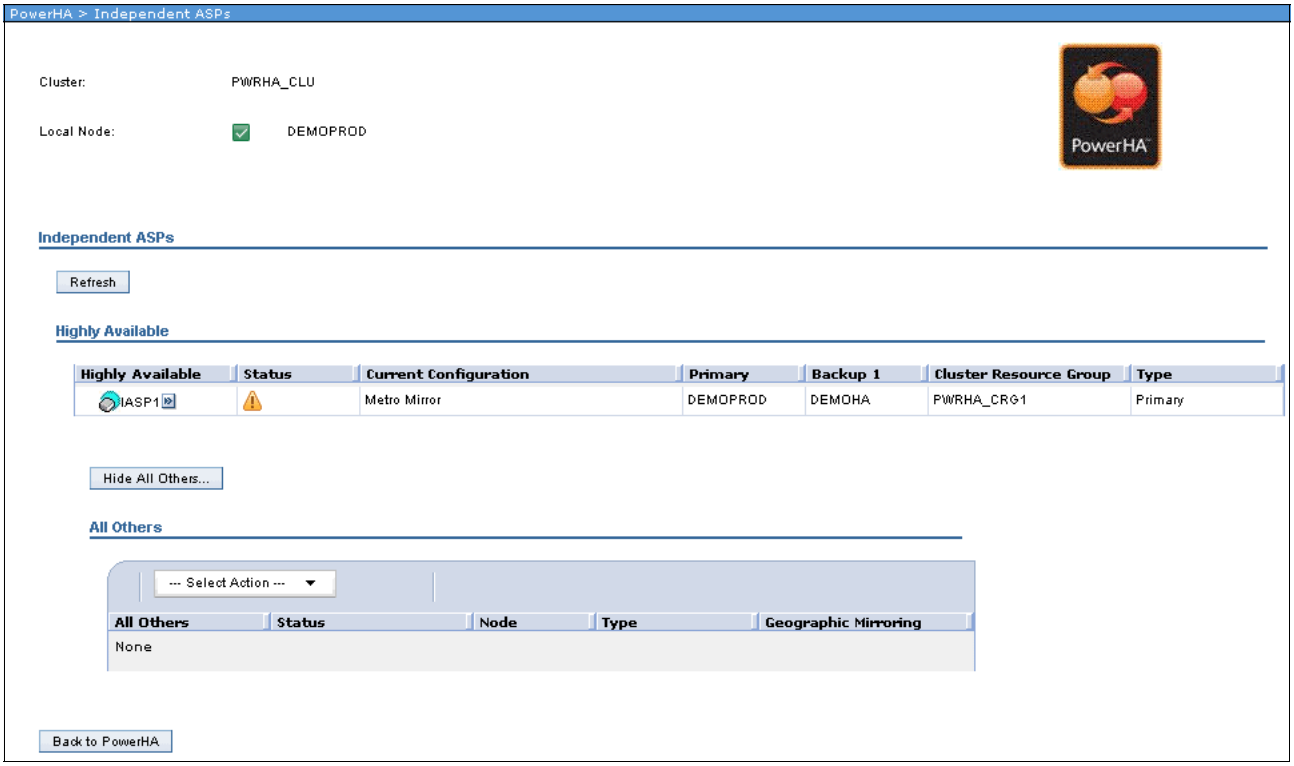


Figure 13-37 IASP is highly available ready

Using CL commands

The corresponding CL commands for creating a device domain and a cluster resource group are **ADDDEVDMNE** and **CRTCRG**.

The *device domain* must be properly populated with the nodes participating in the cluster resource group before we can create it.

As shown in Example 13-3, one command is necessary for each node to add it into the device domain. For our environment DEMOPROD owns the IASP and therefore must be the first node added into the device domain, and DEMOHA is the second node. Therefore, for any existing Independent ASP on DEMOPROD, the corresponding device descriptions need to be manually created on DEMOHA using **CRTDEVASP**. This is in contrast to the PowerHA GUI, which creates the ASP device descriptions on the other nodes in the device domain automatically.

Example 13-3 Adding nodes to a device domain and creating ASP device description

```
ADDDEVDMNE CLUSTER(PWRHA_CLU) DEVDMN(PWRHA_DMN) NODE(DEMOPROD)
ADDDEVDMNE CLUSTER(PWRHA_CLU) DEVDMN(PWRHA_DMN) NODE(DEMOHA)

CRTDEVASP DEVD(IASP1) RSRCNAME(IASP1) RDB(IASP1)
```

After the appropriate device domain and all IASP devices descriptions have been created, we can create the *cluster resource group* (Example 13-4).

Example 13-4 Creating a cluster resource group

```
CRTCRG CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG1) CRGTYPE(*DEV)
   EXITPGM(*NONE) USRPRF(*NONE)
   RCYDMN((DEMOPROD *PRIMARY *LAST SITE1) (DEMOHA *BACKUP 1 SITE2))
   CFGOBJ((IASP1 *DEVD *ONLINE '10.0.0.1'))
   TEXT('DS8000 Cluster Resource Group')
```

The related command for adding a Metro Mirror ASP copy description is **ADDASPCPYD** (Example 13-5). You have to create one for each IASP copy (that is, one for the Metro Mirror source and one for the target site).

Example 13-5 Adding Metro Mirror ASP copy descriptions

```
ADDASPCPYD ASPCPY(IASP1_MM1) ASPDEV(IASP1) CRG(PWRHA_CRG1) LOCATION(*DEFAULT)
   SITE(SITE1) STGHOST('powerha' ('password') ('9.5.168.55'))
   LUN('IBM.2107-75AY031' ('6000-6002' '6100-6102') ())
ADDASPCPYD ASPCPY(IASP1_MM2) ASPDEV(IASP1) CRG(PWRHA_CRG1) LOCATION(*DEFAULT)
   SITE(SITE2) STGHOST('powerha' ('password') ('9.5.168.55'))
   LUN('IBM.2107-75AY032' ('6000-6002' '6100-6102') ())
```

Starting ASP session and cluster resource group using CL commands

When creating the ASP copy descriptions through the GUI, at the end of the procedure, the session is automatically started. However, if you need to start it again, **STRASPSSN** can be used. Also, **STRCRG** can be used to start the cluster resource group when needed (Example 13-6).

Example 13-6 Starting the ASP session and cluster resource group

```
STRASPSSN SSN(IASP1_MM) TYPE(*METROMIR) ASPCPY((IASP1_MM1 IASP1_MM2))
STRCRG CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG1)
```

13.1.2 Configuring IBM i DS8000 FlashCopy (CL commands)

This section discusses a scenario using the DS8000 Copy Services FlashCopy function for creating a point-in-time copy of an IASP and making it available for another IBM i partition for backup purposes.

Environment overview

For this scenario, we need to create several cluster items. Cluster, cluster monitors, IASP, and administrative domain setup are covered in Chapter 11, “Creating a PowerHA base environment” on page 199. They are common to all scenarios that we describe in this chapter.

These are cluster items that we specifically configure for FlashCopy:

- Device domain
- FlashCopy ASP copy descriptions and sessions

Table 13-2 and Figure 13-38 on page 296 show the setup that we are going to use for our FlashCopy environment.

Table 13-2 Settings for FlashCopy scenario

	Source IASP	**Target IASP**
System name	DEMOHA	DEMOFC
Cluster name	PWRHA_CLU	
Device domain	PWRHA_DMN	
IASP name, number	IASP2,34	IASP2,34
Management access IP	9.5.168.129	9.5.168.130
DS8000 device ID	IBM.2107-75AY032	IBM.2107-75AY032
DS8000 IP address	9.5.168.55	9.5.168.55
Volume group ID	V26	V48
Volumes IDs[a]	A010-A013	A020-A021
FC IO adapter resource name[b]	DC05	DC04
Host connection ID	5	2
Host connection WWPN	10000000C94122A2	10000000C9523E9D

a. Volumes need to be the same size on both sides.
b. FC stands for Fibre Channel.

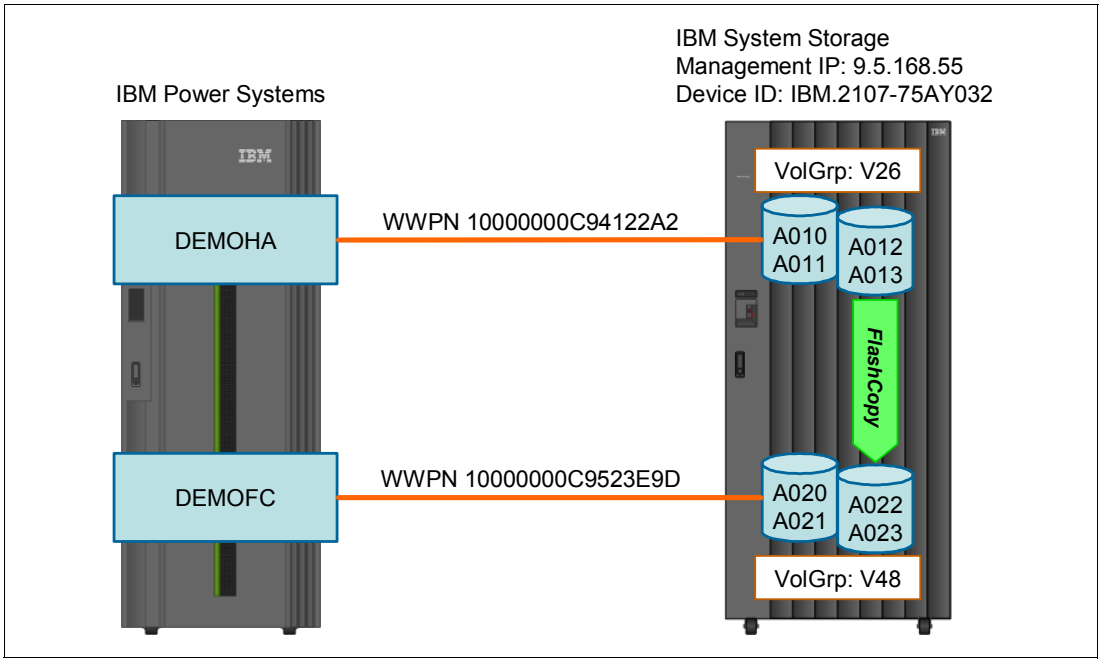


Figure 13-38 The schematics of our environment

Volume groups in the DS8000

We assume that we already have the IASP2 configured on the system DEMOHA, and we have the volumes for the target copy prepared. We list them using the DS CLI **showvolgrp** command (Figure 13-39).

```
dscli> showvolgrp v26
Date/Time: September 28, 2011 10:13:25 PM CEST IBM DSCLI Version: 7.6.10.530
DS: IBM.2107-75AY032
Name V48
ID   V26
Type OS400 Mask
Vols A010 A011 A012 A013
dscli> showvolgrp v48
Date/Time: September 28, 2011 10:13:30 PM CEST IBM DSCLI Version: 7.6.10.530
DS: IBM.2107-75AY032
Name V48fc
ID   V48
Type OS400 Mask
Vols A020 A021 A022 A023
```

Figure 13-39 Volumes for source and target FlashCopy

IASP device description and ASP copy descriptions

On the DEMOHA system there is IASP2 already configured. To be able to use the copy of it on the backup DEMOHA system, we need to create an IASP2 device description on it using **CRTDEVASP**, as follows:

```
CRTDEVASP DEVD(IASP2) RSRCNAME(IASP2)
```

Then we need to create the ASP copy descriptions, which describe both copies of the IASP2 (that is, the FlashCopy source and target volumes) by using **ADDASPCPYD** on either one of the two systems. Figure 13-40 and Figure 13-41 on page 298 are examples of our commands.

```
Add ASP Copy Description (ADDASPCPYD)

 Type choices, press Enter.

 ASP copy . . . . . > ASPCPYD1      Name
 ASP device . . . . . > IASP2        Name
 Cluster resource group . . . . .   *NONE        Name, *NONE
 Cluster resource group site  . .   *NONE        Name, *NONE
 Storage host:
   User name  . . . . . > USER_ID
   Password . . . . . > PASSWORD
   Internet address . . . . . > 9.5.168.55



 Location . . . . . > DEMOHA       Name, *DEFAULT, *NONE

Logical unit name:
   TotalStorage device  . . . . .   IBM.2107-75AY032
   Logical unit range . . . . . .   A010-A013     Character value
               + for more values
   Consistency group range  . . .                Character value
               + for more values
 Recovery domain:
   Cluster node . . . . . .   *NONE         Character value, *NONE
   Host identifier  . . . . . .                 Character value
               + for more values
   Volume group . . . . . .                     Character value
               + for more values
               + for more values

                                                                      Bottom
 F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
 F24=More keys
```

Figure 13-40 ADDASPCPYD for DEMOHA system

```
                    Add ASP Copy Description (ADDASPCPYD)

Type choices, press Enter.

ASP copy . . . . . > ASPCPYD2      Name
ASP device . . . . . > IASP2       Name
Cluster resource group . . . . .   *NONE       Name, *NONE
Cluster resource group site  . .   *NONE       Name, *NONE
Storage host:
  User name  . . . . . > USER_ID
  Password . . . . . > PASSWORD
  Internet address . . . . . > '9.5.168.55'



Location . . . . . > DEMOFC       Name, *DEFAULT, *NONE

Logical unit name:
  TotalStorage device  . . . . . > 'IBM.2107-75AY032'
  Logical unit range . . . . . > 'A020-A023'   Character value
              + for more values
  Consistency group range  . . .               Character value
              + for more values
Recovery domain:
  Cluster node . . . . . .   *NONE       Character value, *NONE
  Host identifier  . . . . . .             Character value
              + for more values
  Volume group . . . . . .                 Character value
              + for more values
              + for more values

                                                                      Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 13-41 ADDASPCPYD for DEMOFC system

Here we describe parameters that we used in the ASP copy descriptions:

- Username and password

  This specifies the DS user ID and password. In the case of the ASP copy description used for FlashCopy, all source and all target volumes need to be on the same TotalStorage device. Therefore, the user and password for both ASP copy descriptions will be the same.

- Internet address

  This is the IP address for managing the TotalStorage system.

- Location

  This parameter specifies the IBM i cluster node where the given copy of the IASP can be used. In our case the source for the FlashCopy is being used by DEMOHA and the target copy of the IASP will be used by DEMOFC.

Doing the FlashCopy of an IASP

After completing the configuration steps described in previous sections, we are ready to start the FlashCopy operation for IASP2 on the DEMOHA system:

1. Vary off the IASP or quiesce the database activity for the IASP on the production system.

   A quiesce brings the database to a consistent state by flushing modified data from main memory to disk and suspending database operations and transactions. The vary-on of the FlashCopied IASP will still be abnormal, but a lengthy database recovery is avoided, which makes the vary-on process shorter. This operation must be run on the production system. We run:

   ```
   CHGASPACT ASPDEV(IASP1) OPTION(*SUSPEND) SSPTIMO(60)
   ```

2. Start the ASP FlashCopy session.

   In this step we establish a point-in-time copy of the IASP with the FlashCopy operation. On the backup system we run the following command:

   ```
   STRASPSSN SSN(ASPSSN2) TYPE(*FLASHCOPY) ASPCPY((ASPCPYD1 ASPCPYD2))
   ```

   Figure 13-42 shows an example of the command starting the FlashCopy process in our environment.

```
                        Start ASP Session (STRASPSSN)

Type choices, press Enter.

Session  . . . . . . . . . . . . > IASPSSN      Name
Session type . . . . . . . . . . > *FLASHCOPY    *GEOMIR, *METROMIR...
ASP copy:
  Preferred source . . . . . . . > IASCPYD1      Name
  Preferred target . . . . . . . > IASCPYD2      Name
              + for more values
FlashCopy type . . . . . . . . .   *NOCOPY       *COPY, *NOCOPY
Persistent relationship  . . . .   *NO           *YES, *NO




                                                                      Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 13-42 Start ASP session

Here we explain parameters used by **STRASPSSN** that might need more clarification:

- Session type

  In our case we use *FLASHCOPY.

- Preferred source and target

  This specifies the ASP copy descriptions the we created previously.The correct order is important, as the source copy description points to the volumes that will be treated as a source for the FlashCopy and the target description points to the target volumes.

– Flashcopy type

  This parameter specifies whether all content of the source volumes will be copied onto the target volumes. The differences between those settings are described in “Full volume copy” on page 110 and “Nocopy option” on page 110.

– Persistent relationship

  This specifies whether the FlashCopy relation is persistent. When the relationship is persistent, the DS storage system maintains the FlashCopy relationship and keeps track of the changes to the IASP volumes even when all tracks have already been copied. You need to set it to *YES when you are going to use incremental FlashCopy or FlashCopy reverse.

3. Resume the IASP activity on the production system.

   To resume the normal operations for a previously quiesced IASP on the production system, you need to run the following command:

   ```
   CHGASPACT ASPDEV(IASP2) OPTION(*RESUME)
   ```

   Otherwise, if the IASP has been varied off on the production system before taking the FlashCopy, it can now be varied on again using **VRYCFG**.

4. Vary on the IASP copy on the backup system.

   Immediately after the FlashCopy session is started you can make the IASP copy available to your backup system by issuing the following **VRYCFG** command on the backup system:

   ```
   VRYCFG CFGOBJ(IASP2)
          CFGTYPE(*ASP)
          STATUS(*ON)
   ```

   When the IASP2 becames AVAIALABLE you can use it.

5. End the FlashCopy session.

   When you finish using your IASP2 copy you can end the FlashCopy session on the target system as follows:

   a. Vary off the IASP2:

      ```
      VRYCFG CFGOBJ(IASP2) CFGTYPE(*ASP) STATUS(*OFF)
      ```

   b. End the ASP session:

      ```
      ENDASPSSN SSN(ASPSSN2)
      ```

Monitoring the status of the FlashCopy session

You can check the status of the FlashCopy session with **DSPASPSSN** (Figure 13-43). In the output you can see the status of the IASP2 on each system. The UNKNOWN state for a remote system is normal.

```
                              Display ASP Session                   DEMOFC
                                                        09/30/11  13:42:13
Session . . . . . . . . . . . . . . . . . . :   ASPSSN2
  Type  . . . . . . . . . . . . . . . . . . :     *FLASHCOPY
  Persistent  . . . . . . . . . . . . . . . :     *NO
  FlashCopy type  . . . . . . . . . . . . . :     *NOCOPY
  Number sectors copied . . . . . . . . . . :     1536
  Number sectors remaining to be copied . . :     67107328


                             Copy Descriptions

ASP
device         Name                 Role          State     Node
IASP2          ASPCPYD1           SOURCE        UNKNOWN     DEMOHA
IASP2          ASPCPYD2           TARGET        VARYOFF     DEMOFC


                                                                  Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel   F19=Automatic refresh

                             Display ASP Session                   DEMOHA
                                                        09/30/11  13:42:07
Session . . . . . . . . . . . . . . . . . . :   ASPSSN2
  Type  . . . . . . . . . . . . . . . . . . :     *FLASHCOPY
  Persistent  . . . . . . . . . . . . . . . :     *NO
  FlashCopy type  . . . . . . . . . . . . . :     *NOCOPY
  Number sectors copied . . . . . . . . . . :     1536
  Number sectors remaining to be copied . . :     67107328


                             Copy Descriptions

ASP
device         Name                 Role          State     Node
IASP2          ASPCPYD1           SOURCE      AVAILABLE     DEMOHA
IASP2          ASPCPYD2           TARGET        UNKNOWN     DEMOFC


                                                                  Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel   F19=Automatic refresh
```

Figure 13-43 ASP session status for FlashCopy on target and source system

On the DS8000 you can check the status of the FlashCopy volume relationships with **lsflash** (Figure 13-44).

```
dscli> lsflash -fmt default a010-a013
Date/Time: October 3, 2011 4:12:40 PM CEST IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
ID        SrcLSS SequenceNum Timeout ActiveCopy Recording Persistent Revertible SourceWriteEnabled TargetWriteEnabled BackgroundCopy
==========================================================================================================================================
A010:A020 A0     0           60      Disabled   Enabled   Enabled    Disabled   Enabled            Enabled            Enabled
A011:A021 A0     0           60      Disabled   Enabled   Enabled    Disabled   Enabled            Enabled            Enabled
A012:A022 A0     0           60      Disabled   Enabled   Enabled    Disabled   Enabled            Enabled            Enabled
A013:A023 A0     0           60      Disabled   Enabled   Enabled    Disabled   Enabled            Enabled            Enabled
```

Figure 13-44 lsflash for FlashCopy session

Using FlashCopy of the Metro Mirror target

For this scenario we use the configuration shown in Figure 13-45 and Table 13-3 on page 303. In addition, the FlashCopy target volumes will be space efficient.

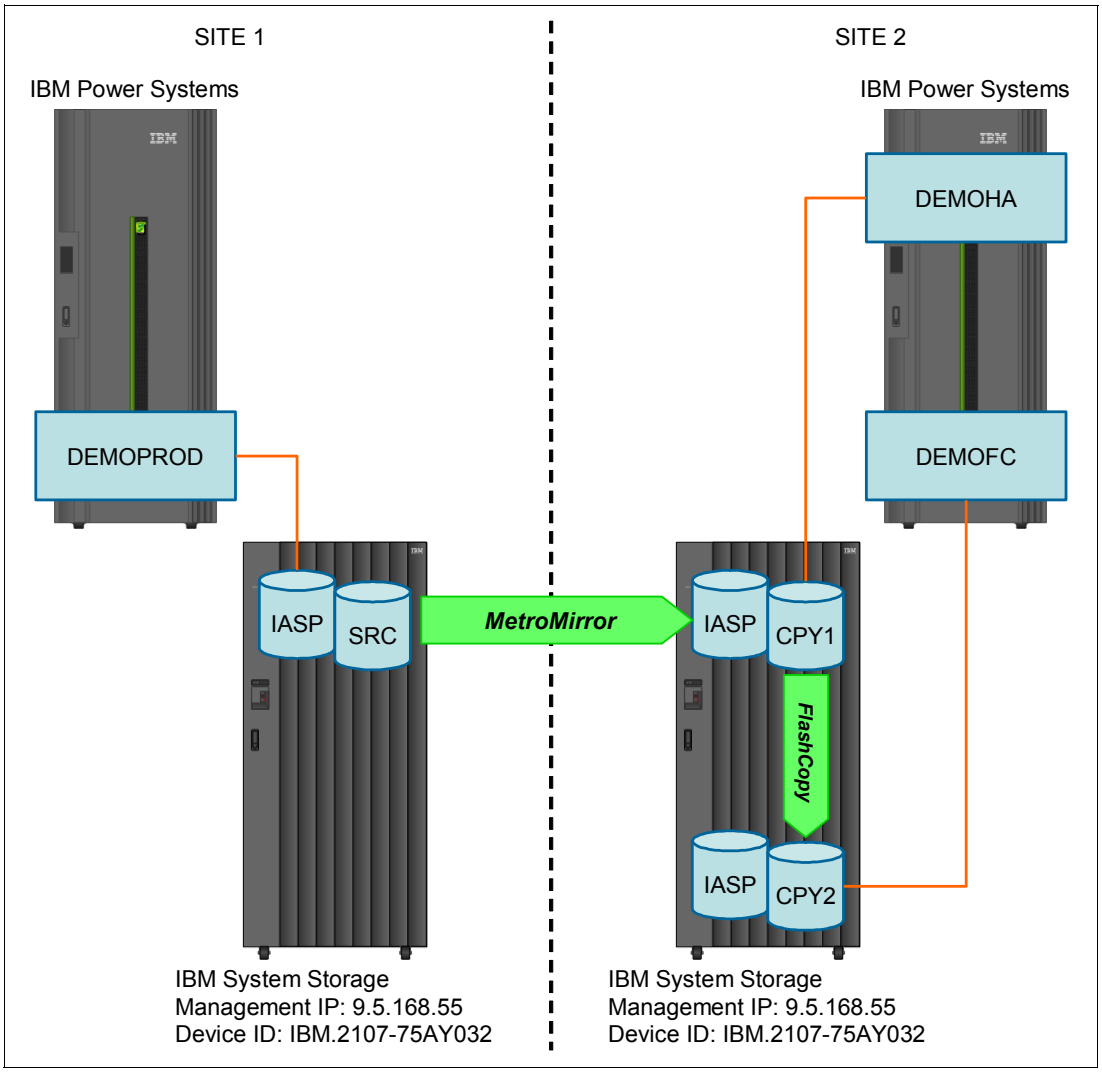


Figure 13-45 Environment for Metro Mirror target FlashCopy

Table 13-3 Parameters for Metro Mirror target FlashCopy

	Source IASP	**Target IASP**
System name	DEMOHA	DEMOFC
Cluster name	PWRHA_CLU	
Device domain	PWRHA_DMN	
IASP name, number	IASP1, 33	IASP1, 33
Management access IP	9.5.168.129	9.5.168.130
DS8000 device ID	IBM.2107-75AY032	IBM.2107-75AY032
DS8000 IP address	9.5.168.55	9.5.168.55
Volume group ID	V2	V48
Volumes IDs[a]	6000-6002 6100-6102	A600-A602 A700-A702
FC IO adapter resource name[b]	DC07	DC04
Host connection ID	1	2
Host connection WWPN	10000000C9523F21	10000000C9523E9D

a. Volumes need to be the same size on both sides.
b. FC stands for Fibre Channel.

We have configured Metro Mirror replication between DEMOPROD and DEMOHA systems. Figure 13-46 shows the status of this replication. On the DS storage system we can list the volumes used by IASP1 on the DEMOHA system and their properties (we show only one volume's details) (Figure 13-47 on page 304).

```
dscli> lspprc -fmt default 6000-6002 6100-6102
Date/Time: October 3, 2011 4:37:56 PM CEST IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
ID        State              Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================================
6000:6000 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6001:6001 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6002:6002 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6100:6100 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
6101:6101 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
6102:6102 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
```

Figure 13-46 Metro Mirror status

```
dscli> showvolgrp v2
Date/Time: September 30, 2011 9:34:05 PM CEST IBM DSCLI Version: 7.6.10.530 DS:
IBM.2107-75AY032
Name demoha_powerha
ID   V2
Type OS400 Mask
Vols 6000 6001 6002 6100 6101 6102
dscli> showfbvol 6000
Date/Time: September 30, 2011 9:34:11 PM CEST IBM DSCLI Version: 7.6.10.530 DS:
IBM.2107-75AY032
Name             demoprod_powerha
ID               6000
accstate         Online
datastate        Normal
configstate      Normal
deviceMTM        2107-A01
datatype         FB 520P
addrgrp          6
extpool          P0
exts             8
captype          iSeries
cap (2^30B)      8.0
cap (10^9B)      8.6
cap (blocks)     16777216
volgrp           V2
ranks            1
dbexts           -
sam              Standard
repcapalloc      -
eam              rotatevols
reqcap (blocks)  16777216
realextents      8
virtualextents   0
migrating        0
perfgrp          -
migratingfrom    -
resgrp           -
```

Figure 13-47 Volume group and volume details for IASP1

To use space-efficient (SE) volumes, we need to create the space-efficient storage pool. The SE storage is created in extension storage pool P5. The SE storage will have virtual size of 100 GB and the allocated physical size of 20 GB (Figure 13-48).

```
dscli> mksestg -repcap 20 -vircap 100 p5
Date/Time: October 5, 2011 6:21:37 PM CEST IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
CMUC00342I mksestg: The space-efficient storage for the extent pool P5 has been created successfully.
dscli> lssestg
Date/Time: October 5, 2011 6:26:10 PM CEST IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
extpool stgtype datastate configstate repcapstatus repcap(GiB/Mod1) vircap
==========================================================================
P0      fb      Normal    Normal      below                 300.0 2200.0
P1      fb      Normal    Normal      below                 300.0 2200.0
P3      fb      Normal    Normal      below                 300.0 3000.0
P5      fb      Normal    Normal      below                  20.0  100.0
```

Figure 13-48 Creation of the SE storage pool

Now we have to create the target volumes for the FlashCopy. The volumes need to be the same type as the source ones, in this case A01. Using the DS CLI **mkfbvol** commands (Figure 13-49), we create the volumes in the SE storage created earlier. We specify that the volumes are track space efficient (TSE). After creating the FlashCopy target volumes we assign them to a new volume group using the DS CLI **mkvolgrp** and **chvolgrp** command.

```
dscli> mkfbvol -extpool p5 -os400 a01 -sam tse 11a0-11a5
Date/Time: October 5, 2011 6:32:22 PM CEST IBM DSCLI Version: 7.6.10.530 DS:
IBM.2107-75AY032
CMUC00025I mkfbvol: FB volume 11A0 successfully created.
CMUC00025I mkfbvol: FB volume 11A1 successfully created.
CMUC00025I mkfbvol: FB volume 11A2 successfully created.
CMUC00025I mkfbvol: FB volume 11A3 successfully created.
CMUC00025I mkfbvol: FB volume 11A4 successfully created.
CMUC00025I mkfbvol: FB volume 11A5 successfully created.
dscli> mkvolgrp -type os400mask mmflashTGT
Date/Time: October 5, 2011 6:37:59 PM CEST IBM DSCLI Version: 7.6.10.530 DS:
IBM.2107-75AY032
CMUC00030I mkvolgrp: Volume group V49 successfully created.
dscli> chvolgrp -action add -volume 11A0-11A5 v49
Date/Time: October 5, 2011 6:38:43 PM CEST IBM DSCLI Version: 7.6.10.530 DS:
IBM.2107-75AY032
CMUC00031I chvolgrp: Volume group V49 successfully modified.
```

Figure 13-49 Creation of the volumes and adding them to a volume group

Now we need to connect the volumes to the DEMOFC system for which we use an already existing host connection for DEMOFC on the DS storage system and assign it our newly created volume group V50 containing the FlashCopy target volumes (Figure 13-50).

```
dscli> showhostconnect 2
Date/Time: October 5, 2011 6:40:18 PM CEST IBM DSCLI Version: 7.6.10.530 DS:
IBM.2107-75AY032
Name          demofc_powerha
ID            0002
WWPN          10000000C9523E9D
HostType      iSeries
LBS           520
addrDiscovery reportLUN
Profile       IBM iSeries - OS/400
portgrp       0
volgrpID      -
atchtopo      -
ESSIOport     all
speed         Unknown
desc          -
dscli> chhostconnect -volgrp v49 2
Date/Time: October 5, 2011 6:40:35 PM CEST IBM DSCLI Version: 7.6.10.530 DS:
IBM.2107-75AY032
CMUC00013I chhostconnect: Host connection 0002 successfully modified.
dscli> showhostconnect 2
Date/Time: October 5, 2011 6:40:39 PM CEST IBM DSCLI Version: 7.6.10.530 DS:
IBM.2107-75AY032
Name          demofc_powerha
ID            0002
WWPN          10000000C9523E9D
HostType      iSeries
LBS           520
addrDiscovery reportLUN
Profile       IBM iSeries - OS/400
portgrp       0
volgrpID      V49
atchtopo      -
ESSIOport     all
speed         Unknown
desc          -
```

Figure 13-50 Changing host connect for volume group V50 and DEMOFC system

Now we can see the disks in the DEMOFC system (Figure 13-51).

```
                Logical Hardware Resources Associated with IOP

Type options, press Enter.
  2=Change detail    4=Remove    5=Display detail    6=I/O debug
  7=Verify           8=Associated packaging resource(s)

                                                Serial          Part
Opt Description                    Type-Model   Number          Number
    Combined Function IOP           2844-001    53-7141345      0000039J1719
     Storage IOA                    2787-001    1F-C55OOBA      0000080P6417
      Disk Unit                     2107-A01    5O-11AOBB8
      Disk Unit                     2107-A01    5O-11A1BB8
      Disk Unit                     2107-A01    5O-11A2BB8
      Disk Unit                     2107-A01    5O-11A3BB8
      Disk Unit                     2107-A01    5O-11A4BB8
      Disk Unit                     2107-A01    5O-11A5BB8




F3=Exit    F5=Refresh    F6=Print    F8=Include non-reporting resources
F9=Failed resources      F10=Non-reporting resources
F11=Display logical address          F12=Cancel
```

Figure 13-51 Volumes in the DEMOFC system

We have created the required target volumes for the FlashCopy and connected them to our backup system DEMOFC. Now we need to create the objects needed to use FlashCopy with PowerHA SystemMirror for i:

1. Create the ASP copy descriptions.

   Figure 13-52 and Figure 13-53 on page 309 shows the commands used for creation of the ASP Copy descriptions needed for the FlashCopy.

```
                    Add ASP Copy Description (ADDASPCPYD)

 Type choices, press Enter.

 ASP copy . . . . . . . . . . . . > ASP1FC1       Name
 ASP device . . . . . . . . . . . > IASP1         Name
 Cluster resource group . . . . .   *NONE         Name, *NONE
 Cluster resource group site  . .   *NONE         Name, *NONE
 Storage host:
   User name  . . . . . . . . . . > userid
   Password . . . . . . . . . . . > password
   Internet address . . . . . . . > '9.5.168.55'



 Location . . . . . . . . . . . . > DEMOHA        Name, *DEFAULT, *NONE
Logical unit name:
   TotalStorage device  . . . . . > 'IBM.2107-75AY032'
   Logical unit range . . . . . . > '6000-6002'   Character value
              + for more values > '6100-6102'
   Consistency group range  . . .                Character value
              + for more values
 Recovery domain:
   Cluster node . . . . . . . . .   *NONE         Character value, *NONE
   Host identifier  . . . . . . .                Character value
              + for more values
   Volume group . . . . . . . . .                Character value
              + for more values
              + for more values



                                                                       Bottom
 F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
 F24=More keys
```

Figure 13-52 ADDASPCPYD for the source of the FlashCopy

```
                    Add ASP Copy Description (ADDASPCPYD)

 Type choices, press Enter.

 ASP copy . . . . . > ASP1FC2       Name
 ASP device . . . . . > IASP1        Name
 Cluster resource group . . . . .   *NONE         Name, *NONE
 Cluster resource group site  . .   *NONE         Name, *NONE
 Storage host:
   User name  . . . . . > userid
   Password . . . . . > password
   Internet address . . . . . > '9.5.168.55'



 Location . . . . . > DEMOFC        Name, *DEFAULT, *NONE
Logical unit name:
   TotalStorage device  . . . . . > 'IBM.2107-75AY032'
   Logical unit range . . . . . > '11A0-11A5'   Character value
               + for more values >
   Consistency group range  . . .                Character value
               + for more values
 Recovery domain:
   Cluster node . . . . . .   *NONE         Character value, *NONE
   Host identifier  . . . . . .                Character value
               + for more values
   Volume group . . . . . .                    Character value
               + for more values
               + for more values



                                                                      Bottom
 F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
 F24=More keys
```

Figure 13-53 ADDASPCPYD for the target FlashCopy

2. Start the ASP FlashCopy session.

   We start the FlashCopy with the following command:

   ```
   STRASPSSN SSN(ASPFC1) TYPE(*FLASHCOPY) ASPCPY((ASP1FC1 ASP1FC2))
   ```

   Figure 13-54 shows the status of the FlashCopy. The FlashCopy target on the DEMOFC system can be varied on and used. Figure 13-55 on page 311 shows the status on the DS8000.

```
                              Display ASP Session                    DEMOFC
                                                          10/05/11  12:00:50
Session . . . . . . . . . . . . . . . . . . . :   ASPFC1
  Type  . . . . . . . . . . . . . . . . . . . :     *FLASHCOPY
  Persistent  . . . . . . . . . . . . . . . . :     *NO
  FlashCopy type  . . . . . . . . . . . . . . :     *NOCOPY
  Number sectors copied . . . . . . . . . . . :     177536
  Number sectors remaining to be copied . . . :     100485760


                              Copy Descriptions

ASP
device         Name                 Role          State     Node
IASP1          ASP1FC1            SOURCE        UNKNOWN     DEMOHA
IASP1          ASP1FC2            TARGET      AVAILABLE     DEMOFC



                                                                      Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel   F19=Automatic refresh
```

Figure 13-54 ASP session status

```
dscli> lsflash 11a0-11a5
Date/Time: October 5, 2011 7:11:27 PM CEST IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
ID        SrcLSS SequenceNum Timeout ActiveCopy Recording Persistent Revertible SourceWriteEnabled TargetWriteEnabled BackgroundCopy
==========================================================================================================================================
6000:11A0 60     0           60      Disabled   Disabled  Disabled   Disabled   Enabled            Enabled            Disabled
6001:11A1 60     0           60      Disabled   Disabled  Disabled   Disabled   Enabled            Enabled            Disabled
6002:11A2 60     0           60      Disabled   Disabled  Disabled   Disabled   Enabled            Enabled            Disabled
6100:11A3 61     0           60      Disabled   Disabled  Disabled   Disabled   Enabled            Enabled            Disabled
6101:11A4 61     0           60      Disabled   Disabled  Disabled   Disabled   Enabled            Enabled            Disabled
6102:11A5 61     0           60      Disabled   Disabled  Disabled   Disabled   Enabled            Enabled            Disabled
dscli> lssestg
Date/Time: October 5, 2011 7:11:31 PM CEST IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
extpool stgtype datastate configstate repcapstatus repcap(GiB/Mod1) vircap
========================================================================
P0      fb      Normal    Normal      below                300.0 2200.0
P1      fb      Normal    Normal      below                300.0 2200.0
P3      fb      Normal    Normal      below                300.0 3000.0
P5      fb      Normal    Normal      below                 20.0  100.0
dscli> showsestg p5
Date/Time: October 5, 2011 7:11:33 PM CEST IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
extpool               P5
stgtype               fb
datastate             Normal
configstate           Normal
repcapstatus          below
%repcapthreshold      0
repcap(GiB)           20.0
repcap(Mod1)          -
repcap(blocks)        41943040
repcap(cyl)           -
repcapalloc(GiB/Mod1) 0.1
%repcapalloc          0
vircap(GiB)           100.0
vircap(Mod1)          -
vircap(blocks)        209715200
vircap(cyl)           -
vircapalloc(GiB/Mod1) 48.0
%vircapalloc          48
overhead(GiB/Mod1)    2.0
reqrepcap(GiB/Mod1)   20.0
reqvircap(GiB/Mod1)   100.0
```

Figure 13-55 lsflash for the FlashCopied volumes

3. End the FlashCopy session.

   When you finish your activities on the target FlashCopy IASP, you need to end the ASP session. To do so you need to vary off the IASP1 and then use **ENDASPSSN SSN(ASPFC1)**.

13.1.3 Configuring IBM i DS8000 Global Mirror (CL commands)

This section describes a scenario using a DS8000 Global Mirror environment as the hardware replication technology. Most of the PowerHA GUI panels wizard are common between Metro Mirror and Global Mirror. Differences exist only when selecting a Metro Mirror or Global Mirror mirroring solution, and for the ASP copy descriptions for which a *consistency group* is mandatory for Global Mirror. This is the reason that we decided to show you only the CL commands for this scenario and point out the differences between configuring Global Mirror versus Metro Mirror with the PowerHA GUI. Each time that names for items such as copy descriptions or sessions display, you have to provide the proper information. The main differences are for the following panels:

- The panel shown in Figure 13-6 on page 273, where you select Global Mirror in place of Metro Mirror.
- The panel shown in Figure 13-25 on page 285, where Global Mirror mirroring is selected due to previous wizard selection.
- Provide information for consistency group logical volumes similar to steps 25 through 30 in “Setting up a Metro Mirror environment” on page 266.

Environment overview

As for Metro Mirror scenarios, we need to create several cluster items. Cluster, cluster monitors, IASP, and administrative domain setup are covered in Chapter 11, “Creating a PowerHA base environment” on page 199. They are common to all scenarios.

These are Global Mirror specific cluster items:

- Cluster resource group
- Device domain, which is the same as a Metro Mirror environment
- Global Mirror copy descriptions and session

Table 13-4 and Figure 13-56 on page 313 describe and show our environment.

Table 13-4 Global Mirror scenario settings

	Preferred production	**Preferred backup**
System name	DEMOPROD	DEMOHA
Cluster name	PWRHA_CLU	
Cluster resource group	PWRHA_CRG2	
Device domain	PWRHA_DMN	
Administrative domain	PWRHA_CAD	
IASP name, number	IASP2, 182	IASP2, 182
Cluster resource group site name	SITE1	SITE2
ASP copy description	IASP2_GM1	IASP2_GM2
ASP session	IASP2_GM	
Takeover IP	10.0.0.1	
Heartbeat cluster IP	10.10.10.1	10.10.10.2
Management access IP	9.5.168.129	9.5.168.130
DS8000 device ID	IBM.2107-13ABGAA	IBM.2107-1375210
DS8000 IP address[a]	9.5.168.32	9.5.168.32
Volume group ID	V15	V13
Volumes IDs[b]	0200-0201, 0300-0301	0200-0201, 0300-0301
Consistency group volumes IDs	0250-0251, 0350-0351	0250-0251, 0350-0351
FC IO adapter resource name[c]	DC05	DC05
Host connection ID	07	01
Host connection WWPN	10000000C94802CB	10000000C94122A2

a. IP addresses are identical because in our environment we use a single management console for both DS8000s.
b. Volumes IDs are not needed to be the same on the source and the target Global Mirror relationships.
c. FC stands for Fibre Channel.

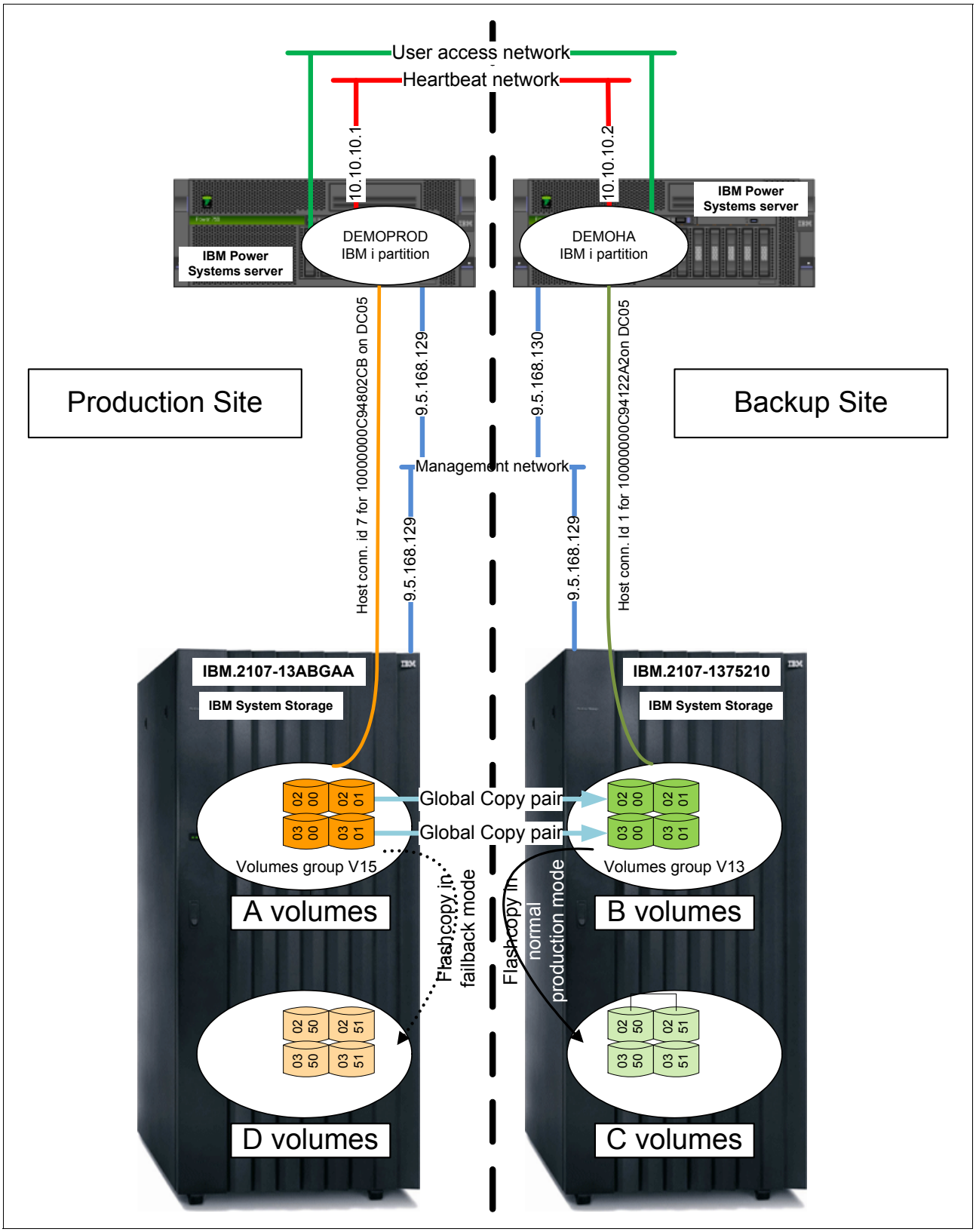


Figure 13-56 Global Mirror scenario settings

Setting up a Global Mirror environment

Global Mirror environment setup is not handled by IBM PowerHA SystemMirror for i. You have to set up all items directly to the DS8000. Refer to 6.3, “Global Mirror” on page 98, for more information.

Creating the PPRC path

The path management activities are exactly the same for the Global Copy part of Global Mirror as for Metro Mirror. Therefore, see “Creating the PPRC paths” on page 267 for detailed information about path commands.

Note: Make sure that the paths exist on both DS8000s from one to the other for the volumes’ LSS.

Creating the Global Copy relationships

After the path exists from Production DS8000 to Backup DS8000, we can establish the Global Copy relationships. For that, we use **mkpprc** with the parameter type set to *gcp* in place of *mmir* (Example 13-7). In our Global Mirror environment, Global Copy relationships are to be established between 0200 - 0201 and 0200 - 0201 volumes for the first set, and 0300 - 0301 and 0300 - 0301 volumes for the second set. The relationship is properly established when the first pass status is *True*.

Example 13-7 Creating Global Copy relationships

```
dscli> mkpprc -dev IBM.2107-13ABGAA -remotedev IBM.2107-1375210 -type gcp -mode full 0200-0201:0200-0201
0300-0301:0300-0301
Date/Time: October 3, 2011 11:48:40 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
CMUC00153I mkpprc: Remote Mirror and Copy volume pair relationship 0200:0200 successfully created.
CMUC00153I mkpprc: Remote Mirror and Copy volume pair relationship 0201:0201 successfully created.
CMUC00153I mkpprc: Remote Mirror and Copy volume pair relationship 0300:0300 successfully created.
CMUC00153I mkpprc: Remote Mirror and Copy volume pair relationship 0301:0301 successfully created.
dscli> lspprc 0200-03FF
Date/Time: October 3, 2011 11:49:11 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
ID        State        Reason Type        SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================
0200:0200 Copy Pending -      Global Copy 02        60             Disabled      False
0201:0201 Copy Pending -      Global Copy 02        60             Disabled      False
0300:0300 Copy Pending -      Global Copy 03        60             Disabled      False
0301:0301 Copy Pending -      Global Copy 03        60             Disabled      False
dscli> lspprc 0200-03FF
Date/Time: October 3, 2011 11:52:22 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
ID        State        Reason Type        SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================
0200:0200 Copy Pending -      Global Copy 02        60             Disabled      True
0201:0201 Copy Pending -      Global Copy 02        60             Disabled      True
0300:0300 Copy Pending -      Global Copy 03        60             Disabled      True
0301:0301 Copy Pending -      Global Copy 03        60             Disabled      True
```

Note: In a Global Copy relationship, the source volumes remain at *copy pending* status and the target volumes at *target copy pending* status. They never reach full duplex and target full duplex, because of the asynchronous process.

Creating the FlashCopy relationships on the backup site

After the Global Copy is established between the production DS8000 and the backup DS8000, we can establish the FlashCopy relationships (between B and C volumes) on the backup DS8000. For that, we use **mkflash** (Example 13-8 on page 315). In our Global Mirror

environment, Flash Copy relationships are to be established between 0200 - 0201 and 0250 - 0251 volumes for a first set, and 0300 - 0301 and 0350 - 0351 volumes for a second set. The relationship is properly established as soon as the command returns. Specific parameters for a Flash Copy relationship used in a Global Mirror session are as follows:

- persist (persistent)
- tgtinhibit (target write inhibit)
- record (record changes)
- nocp (no full copy)

Example 13-8 Global Mirror FlashCopy relationships

```
dscli> mkflash -persist -nocp -tgtinhibit -record 0200-0201:0250-0251 0300-0301:0350-0351
Date/Time: October 3, 2011 11:54:00 AM CDT IBM DSCLI Version: 7.6.10.530 DS:
IBM.2107-13ABGAA
CMUC00137I mkflash: FlashCopy pair 0200:0250 successfully created.
CMUC00137I mkflash: FlashCopy pair 0201:0251 successfully created.
CMUC00137I mkflash: FlashCopy pair 0300:0350 successfully created.
CMUC00137I mkflash: FlashCopy pair 0301:0351 successfully created.
```

Making a Global Mirror session

After FlashCopy relationships exist on the backup system, we can create the Global Mirror session. For automatic switchback to be managed by PowerHA, the session must be created on each DS8000 (in our case, IBM.2107-13ABGAA is the Production DS8000 and IBM.2107-1375210 is the Backup DS8000). To do that, we use **mksession** (Example 13-9). **mksession** applies to an LSS. At the same time, we include our volumes in the session on the production DS8000 only. No volume must be included in the session on the backup DS8000.

Example 13-9 Global Mirror session creation

```
dscli> mksession -dev IBM.2107-13ABGAA -lss 02 -volume 0200-0201 2
Date/Time: October 3, 2011 12:02:04 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
CMUC00145I mksession: Session 2 opened successfully.
dscli> mksession -dev IBM.2107-13ABGAA -lss 03 -volume 0300-0301 2
Date/Time: October 3, 2011 12:02:37 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
CMUC00145I mksession: Session 2 opened successfully.
dscli> lssession 02-03
Date/Time: October 3, 2011 12:02:44 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
LSS ID Session Status Volume VolumeStatus PrimaryStatus        SecondaryStatus   FirstPassComplete
AllowCascading
=================================================================================================================
======
02     02      Normal 0200   Join Pending Primary Copy Pending Secondary Simplex True              Disable
02     02      Normal 0201   Join Pending Primary Copy Pending Secondary Simplex True              Disable
03     02      Normal 0300   Join Pending Primary Copy Pending Secondary Simplex True              Disable
03     02      Normal 0301   Join Pending Primary Copy Pending Secondary Simplex True              Disable
dscli> mksession -dev IBM.2107-1375210 -lss 02 1
Date/Time: October 3, 2011 12:05:25 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-1375210
CMUC00145I mksession: Session 1 opened successfully.
dscli> mksession -dev IBM.2107-1375210 -lss 03 1
Date/Time: October 3, 2011 12:05:37 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-1375210
CMUC00145I mksession: Session 1 opened successfully.
dscli> lssession 02-03
Date/Time: October 3, 2011 12:36:59 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-1375210
LSS ID Session Status Volume VolumeStatus PrimaryStatus SecondaryStatus FirstPassComplete AllowCascading
=================================================================================================================
02     01      -      -      -           -             -               -                -
03     01      -      -      -           -             -               -                -
```

Starting Global Mirror session

After A volumes have been added to the session on the production system, we can start the session. For that, we use **mkgmir** (Example 13-10). **mkgmir** applies to an LSS. You use this command on only one LSS.

Example 13-10 Starting Global Mirror session

```
dscli> mkgmir -dev IBM.2107-13ABGAA -lss 02 -session 2
Date/Time: October 3, 2011 12:28:24 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
CMUC00162I mkgmir: Global Mirror for session 2 successfully started.
dscli> lssession 02-03
Date/Time: October 3, 2011 12:28:33 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
LSS ID Session Status         Volume VolumeStatus PrimaryStatus        SecondaryStatus   FirstPassComplete
AllowCascading
=============================================================================================================================
=
02     02      CG In Progress 0200   Active       Primary Copy Pending Secondary Simplex True              Disable
02     02      CG In Progress 0201   Active       Primary Copy Pending Secondary Simplex True              Disable
03     02      CG In Progress 0300   Active       Primary Copy Pending Secondary Simplex True              Disable
03     02      CG In Progress 0301   Active       Primary Copy Pending Secondary Simplex True              Disable
```

Using CL commands

The corresponding CL commands for creating a device domain and a cluster resource group for a Global Mirror scenario are exactly the same as the ones used for the Metro Mirror scenario. Refer to “Using CL commands” on page 293 for more information.

The related command for adding a Global Mirror ASP copy description is **ADDASPCPYD** (Example 13-11). You have to create one for each IASP copy (that is, one for the Global Mirror source and one for the target site). The difference with the Metro Mirror ASP copy description creation command resides in the consistency group volume specification. In our case, volumes 0250 - 0251 and 0350 - 0351 are those volumes on both source and target Global Mirror copy descriptions.

Example 13-11 Adding Global Mirror ASP copy descriptions

```
ADDASPCPYD ASPCPY(IASP2_GM1) ASPDEV(IASP2) CRG(PWRHA_CRG2) LOCATION(*DEFAULT)
   SITE(SITE1) STGHOST('qlpar' 'password' ('9.5.168.32'))
   LUN('IBM.2107-13ABGAA' ('0200-0201' '0300-0301') ('0250-0251' '0350-0351'))
ADDASPCPYD ASPCPY(IASP2_GM2) ASPDEV(IASP2) CRG(PWRHA_CRG2) LOCATION(*DEFAULT)
   SITE(SITE2) STGHOST('qlpar' 'password' ('9.5.168.32'))
   LUN('IBM.2107-1375210' ('0200-0201' '0300-0301') ('0250-0251' '0350-0351'))
```

Starting ASP session and cluster resource group using the CL commands

When creating the ASP copy descriptions through the GUI, at the end of the procedure the session is automatically started. However, if you need to start it again, use **STRASPSSN**. As shown in Example 13-12, **STRCRG** can be used to start the cluster resourcegGroup when needed.

Note: **STRASPSSN** fails if at least one of the consistency group volumes is included in a volume group. PowerHA assumes that there is a DS8000 configuration error in this case because the consistency group volumes must not be accessed by any host, which is supposed to be the case if they are members of a volume group.

Example 13-12 Starting the ASP session and cluster resource group

```
STRASPSSN SSN(IASP2_GM) TYPE(*GLOBALMIR) ASPCPY((IASP2_GM1 IASP2_GM2))
STRCRG CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG2)
```

13.1.4 Configuring IBM i DS8000 LUN-level switching

This section describes the scenario of DS8000 LUN-level switching between two IBM i servers or partitions. For this scenario, we need to create several cluster items. Clusters, cluster monitors, and the IASP setup are covered in Chapter 11, “Creating a PowerHA base environment” on page 199. They are common to all scenarios of this chapter.

Cluster items that we specifically configure for our LUN-level switching environment are cluster resource groups.

Environment overview

In this scenario, we use and switch DS8000 LUNs with volume group ID V48 between partition DEMOFC and partition DEMOHA. Table 13-5 provides more detailed information about the configuration.

Table 13-5 Environment overview

	Production	**Backup**
System name/node name	DEMOFC	DEMOHA
Cluster name	PWRHA_CLU	
Device CRG	PWRHA_CRG3	
Device domain	PWRHA_DMN	
IASP	IASP_LUN	
Site name	SITE1	SITE1
Heartbeat cluster IP	10.0.0.3	10.0.0.2
DS8000 device ID	IBM.2107-75AY032	
DS8000 IP	9.5.168.55	
Host connection ID	0002	0005
Volume group ID	V48	
Volumes IDs	A020-A023	

	Production	Backup
FC IOA resource name	DC04	DC05
Host connection WWPN	10000000C9523E9D	10000000C94122A2

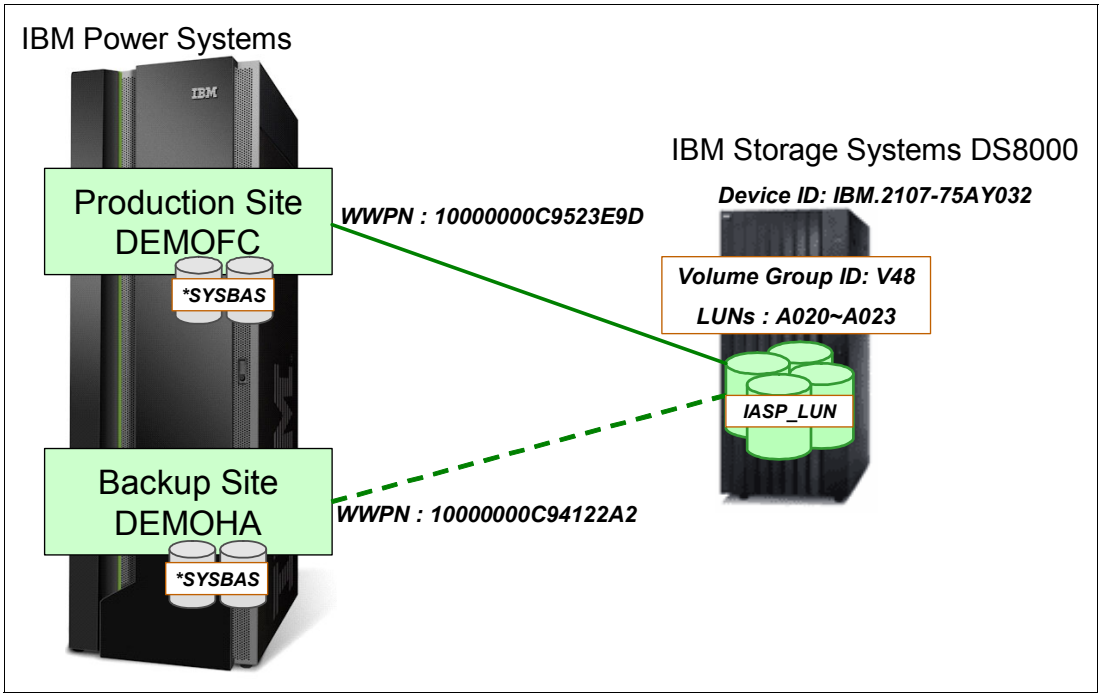


Figure 13-57 The schematics of our scenario

Configuration on the DS8000

For DS8000 LUN-level switching, you need to prepare host identifiers, volume groups, and LUNs on DS8000. In this scenario, A020-A023 LUNs are assigned into volume group V48, and host connection IDs 0002 and 0005 are used for DEMOFC and DEMOHA (Example 13-13).

Note: When configuring LUN-level switching, a volume group is assigned only to the IBM i production node. The host connection for the backup node is *not* allowed to have a volume group assigned, as it will be automatically assigned when LUNs are switched by PowerHA.

Example 13-13 Information about volume groups and host connection on DS8000

```
dscli> showvolgrp V48
Date/Time: September 29, 2011 9:28:20 AM CDT IBM DSCLI Version: 7.6.10.530 DS:
IBM.2107-75AY032
Name V48fc
ID   V48
Type OS400 Mask
Vols A020 A021 A022 A023
dscli> lshostconnect 2 5
Date/Time: September 28, 2011 11:27:31 AM CDT IBM DSCLI Version: 7.6.10.530 DS:
IBM.2107-75AY032
Name           ID   WWPN             HostType Profile              portgrp volgrpID
ESSIOport
```

```
================================================================================================
==
demofc_powerha 0002 10000000C9523E9D iSeries  IBM iSeries - OS/400       0 V48      all
DEMOHA_FC      0005 10000000C94122A2 iSeries  IBM iSeries - OS/400       0 -        all
```

Configuration on IBM i

Create a basic cluster environment for LUN-level switching, as described in Chapter 11, “Creating a PowerHA base environment” on page 199:

1. Create a cluster with two nodes, DEMOFC and DEMOHA, and a device domain (Example 13-14). The cluster nodes are automatically started as the default setting.

   Example 13-14 Creating a cluster environment

   ```
   CRTCLU CLUSTER(PWRHA_CLU)
         NODE((DEMOFC ('10.0.0.3')) (DEMOHA ('10.0.0.2')))
         START(*YES) VERSION(*CUR) HAVERSION(*CUR)
         CLUMSGQ(QSYS/QSYSOPR) FLVWAITTIM(*NOMAX) FLVDFTACN(*PROCEED)

   ADDDEVDMNE CLUSTER(PWRHA_CLU) DEVDMN(PWRHA_DMN) NODE(DEMOFC)
   ADDDEVDMNE CLUSTER(PWRHA_CLU) DEVDMN(PWRHA_DMN) NODE(DEMOHA)
   ```

2. For production node DEMOFC, create an IASP using LUNs previously assigned by DS8000. For backup node DEMOHA we create an IASP device description with the same name that we used on the production node (Example 13-15).

   Example 13-15 Creating IASP

   ```
   On Production site
   CFGDEVASP ASPDEV(IASP_LUN) ACTION(*CREATE) UNITS(DD026 DD025 DD024 DD023)

   On Backup site
   CRTDEVASP DEVD(IASP_LUN) RSRCNAME(IASP_LUN)
   ```

3. After creating a cluster environment and IASP, we create a device cluster resource group using **CRTCRG** (Example 13-16). At this time do not list the IASP as a configuration object for the CRG.

   Example 13-16 Create cluster resource group

   ```
   CRTCRG CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG3)
         CRGTYPE(*DEV) EXITPGM(*NONE) USRPRF(*NONE)
         RCYDMN((DEMOFC *PRIMARY *LAST SITE1) (DEMOHA *BACKUP *LAST SITE1))
   ```

   Note: Make sure that you set the *same* site name for each recovery domain node. In our scenario we use the site name SITE1 for node DEMOFC and node DEMOHA.

4. Create the ASP copy description for the IASP to be used with LUN-level switching using **ADDASPCPYD** (Example 13-17).

Example 13-17 Create ASP Copy Description

```
ADDASPCPYD ASPCPY(LUN_SWITCH) ASPDEV(IASP_LUN) CRG(PWRHA_CRG3) SITE(SITE1)
      STGHOST(powerha () ('9.5.168.55')) LOCATION(*DEFAULT)
      LUN('IBM.2107-75AY032' ('A020-A023') ())
      RCYDMN((DEMOFC (0002) (V48)) (DEMOHA (0005) (V48)))
```

On this command, specify the host identifiers and volume groups for each node in the recovery domain parameter (RCYDMN), even though we do not make the volume group assign a host connection ID for the backup node DEMOHA (Example 13-13 on page 318).

5. Add the IASP to the configuration object list of device CRGs created in step 3 using **ADDCRGDEVE** (Example 13-18). The reason why we did not add the configuration object list when we created the CRG is that we migth get misleading error messages in the job log. This happens when the IASP is listed in the configuration object list of CRGs before the copy description is created. After the copy description is created, PowerHA tells the CRG that PowerHA is in control of the switching, which causes the CRG to skip its IOP and tower switching checks. If the IASP is listed in the configuration object list of CRGs before the copy description is created, then we see misleading error messages in the job log. After the copy description is created, PowerHA tells the CRG that PowerHA is in control of the switching, which causes the CRG to skip its IOP and tower switching checks.

Example 13-18 Add the configuration object to CRG

```
ADDCRGDEVE CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG3) CFGOBJ((IASP_LUN))
```

6. After adding the IASP to the configuration object list of the device CRGs, we start the CRG for LUN-level switching between the production site and the backup site using **STRCRG** (Example 13-19).

Example 13-19 Start CRG

```
STRCRG CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG3)
```

Now your IBM PowerHA SystemMirror for i DS8000 LUN-level switching configuration is complete and your IASP is highly available for planned and unplanned switches, as described in 13.2.2, “Using CL commands for DS8000 LUN-level switching” on page 346.

13.2 Managing IBM i DS8000 Copy Services

In this section we describe how to manage the DS8000 environment for switchover and switchback scenarios for Metro and Global Mirror.

13.2.1 Switchover and switchback for a Metro Mirror or Global Mirror planned outage

In this section we describe the procedure for a planned site switch for Metro Mirror using the PowerHA GUI, and for Global Mirror using CL commands.

From the user standpoint, the steps for performing a switch for both scenarios are the same. The difference is only related to DS8000 commands done under-the-covers by PowerHA to perform the actions.

Operations to perform the switchback are also exactly the same as the switch except in the case of asymmetrical Global Mirror configuration. Switchback for this specific one will be described in detail using CL commands.

Using the PowerHA GUI for a Metro Mirror planned switchover

Note: Before starting any switchover make sure that no jobs are using the IASP any longer.

Before starting the switch, we look at our current configuration. DEMOPROD is the current production system, and DEMOHA is the current first backup system.

In Figure 13-58 and Figure 13-59 on page 322, you can see the IP address configuration on both DEMOPROD and DEMOHA systems. On DEMOPROD the takeover IP address 10.0.0.1 is active. On DEMOHA, this IP address is inactive. The takeover IP address is usually the one to which all users and systems connect.

```
                         Work with TCP/IP Interfaces
                                                              System:   DEMOPROD
Type options, press Enter.
  1=Add   2=Change   4=Remove   5=Display   9=Start   10=End

     Internet         Subnet           Interface     Alias
Opt  Address          Mask             Status        Name

     9.5.168.129      255.255.255.0    Active        *NONE
     9.5.168.133      255.255.255.0    Failed        *NONE
     9.5.168.134      255.255.255.0    Failed        *NONE
     10.0.0.1         255.255.255.0    Active        POWERHA_TAKEOVER_IP
     10.10.10.1       255.255.255.0    Active        POWERHA
     127.0.0.1        255.0.0.0        Active        LOCALHOST
     192.168.86.10    255.255.255.0    Active        *NONE
     192.168.87.10    255.255.255.0    Active        *NONE




                                                                        Bottom
F3=Exit      F5=Refresh   F6=Print list   F11=Display line information
F12=Cancel   F17=Top      F18=Bottom
```

Figure 13-58 Metro Mirror: IP configuration on preferred production before switchover

```
                          Work with TCP/IP Interfaces
                                                                 System:   DEMOHA
Type options, press Enter.
  1=Add   2=Change   4=Remove   5=Display   9=Start   10=End

     Internet         Subnet           Interface     Alias
Opt  Address          Mask             Status        Name

     9.5.168.130      255.255.255.0    Active        *NONE
     9.5.168.133      255.255.255.0    Active        *NONE
     9.5.168.134      255.255.255.0    Active        *NONE
     10.0.0.1         255.255.255.0    Inactive      POWERHA_TAKEOVER_IP
     10.10.10.2       255.255.255.0    Active        POWERHA
     127.0.0.1        255.0.0.0        Active        LOCALHOST
     192.168.86.11    255.255.255.0    Active        *NONE
     192.168.87.11    255.255.255.0    Active        *NONE




                                                                        Bottom
F3=Exit      F5=Refresh   F6=Print list   F11=Display line information
F12=Cancel   F17=Top      F18=Bottom
```

Figure 13-59 Metro Mirror: IP configuration on preferred backup before switchover

Figure 13-60 shows the current status for DS volume members of the Metro Mirror relationships. With **lspprc** on the current production DS8000 *75AY031*, all volumes are *full duplex*, which is the normal status for source volumes. On the current backup DS8000 75AY032, all volumes are *target full duplex*, which is the normal status for target volumes.

```
dscli> lspprc -dev IBM.2107-75AY031 6000-61FF
Date/Time: September 29, 2011 10:18:16 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
ID        State       Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=======================================================================================================
6000:6000 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6001:6001 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6002:6002 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6100:6100 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6101:6101 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6102:6102 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid

dscli> lspprc -dev IBM.2107-75AY032 6000-61FF
Date/Time: September 29, 2011 10:28:06 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
ID        State              Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================================
6000:6000 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6001:6001 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6002:6002 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6100:6100 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
6101:6101 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
6102:6102 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
```

Figure 13-60 Metro Mirror volumes status before switchover

In this section we describe how to manage cluster resource groups and how to do a switchover.

Note: Any cluster node can be selected to perform a switchover as long as it is in active status.

1. Switchover activity is related to the cluster resource group. Therefore, to enter the switchover wizard (Figure 13-61), click **Cluster Resource Groups**.

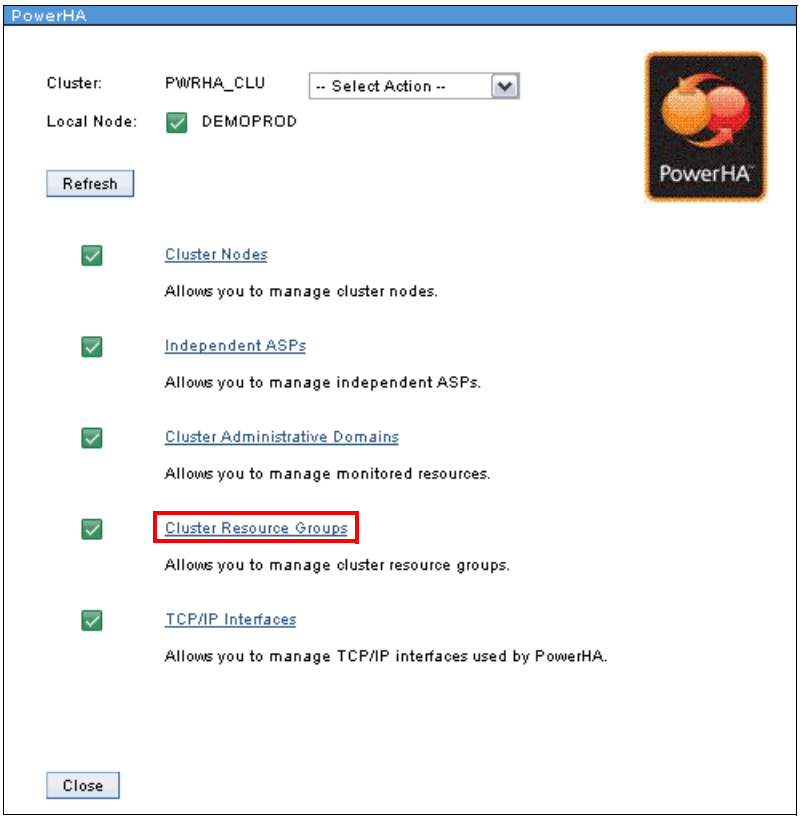


Figure 13-61 Metro Mirror switchover starting point

2. Select **Switchover** against the cluster resource group that you are working on. In our case, this is PWRHA_CRG1 (Figure 13-62).

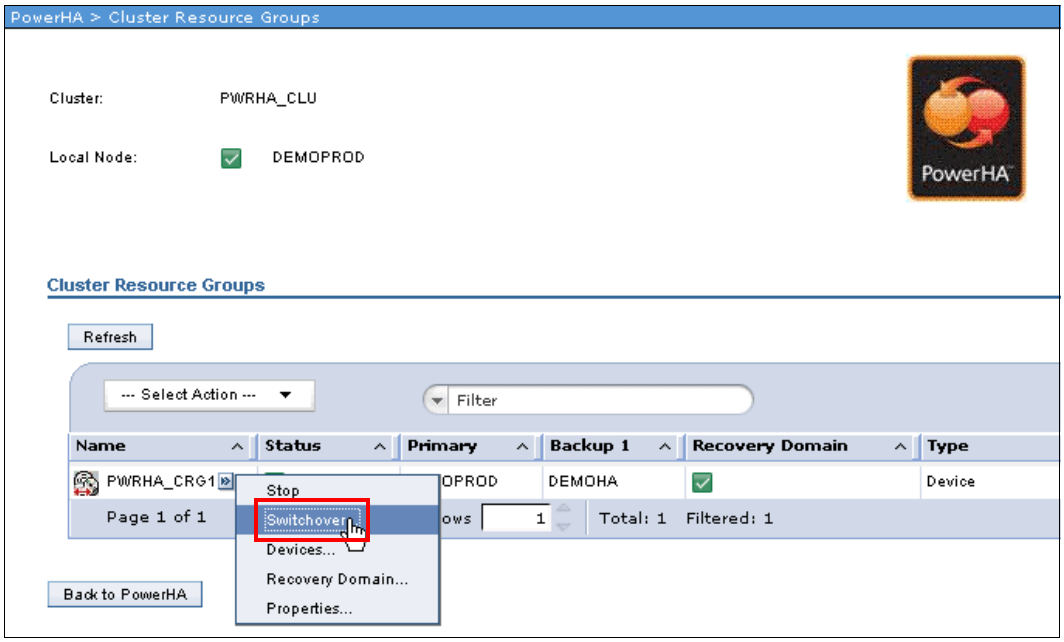


Figure 13-62 Metro Mirror launch switchover

3. Before starting the switchover, we have the opportunity to review the CRG and recovery domain current status, and what will be the result of the operation. In our case (Figure 13-63), all statuses are correct, and after the switchover DEMOPROD will be the first backup site and DEMOHA will be the production site. If everything is correct, click **OK**.

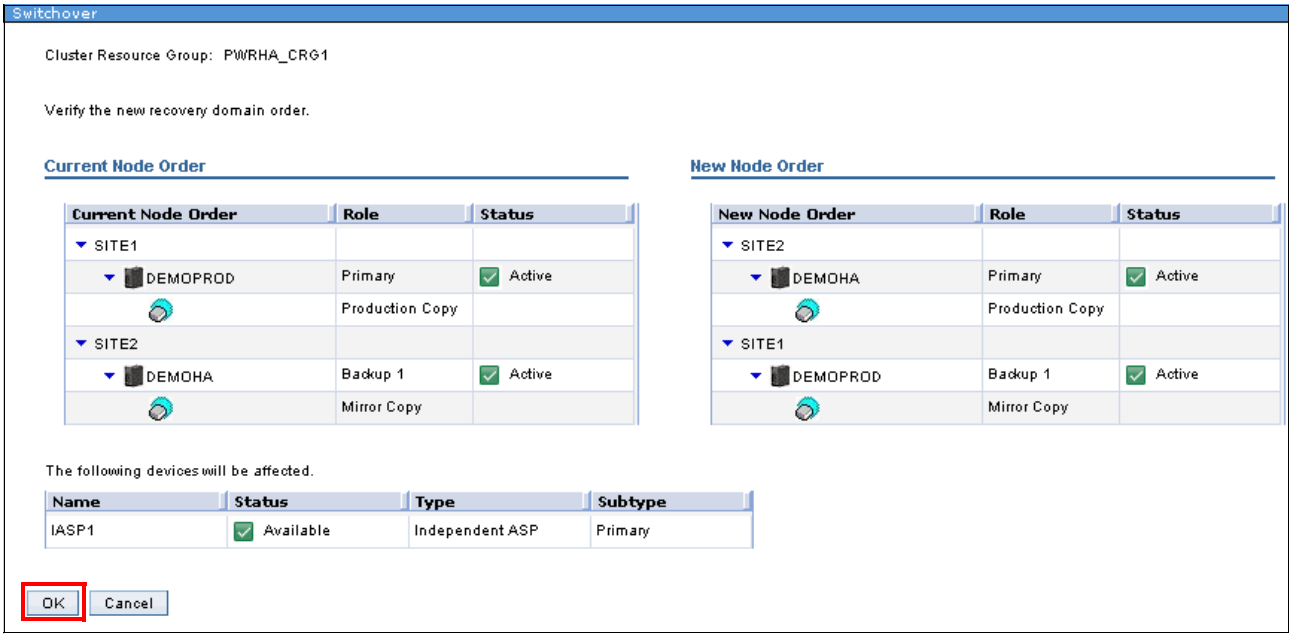


Figure 13-63 Metro Mirror switchover confirmation

Note: Even if an item is not in the appropriate status, the wizard gives you the ability to bypass its warning by clicking **OK** (Figure 13-64 on page 325) and tries to perform the switchover if you request it. However it might fail, in which case you will have to manually fix it.

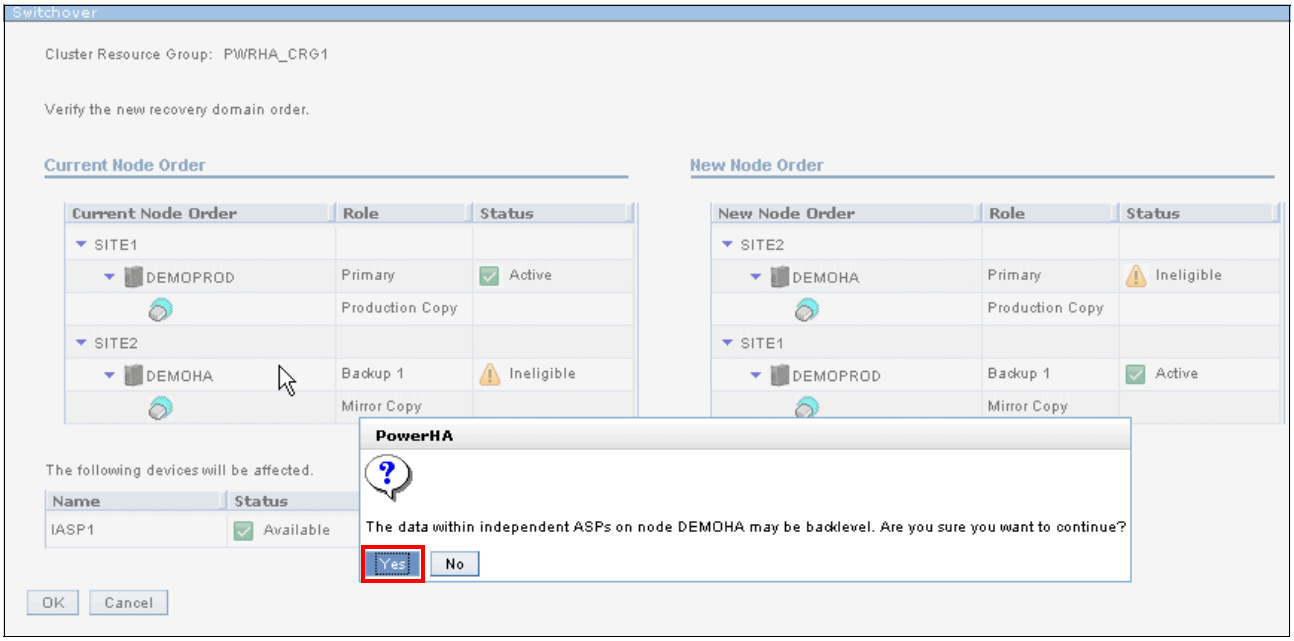


Figure 13-64 Metro Mirror confirm switchover

4. The wizard displays the switchover progress. There are seven steps (Figure 13-65):
   a. The first step ends all jobs using the IASP.
   b. The second step varies off the IASP on the production system.

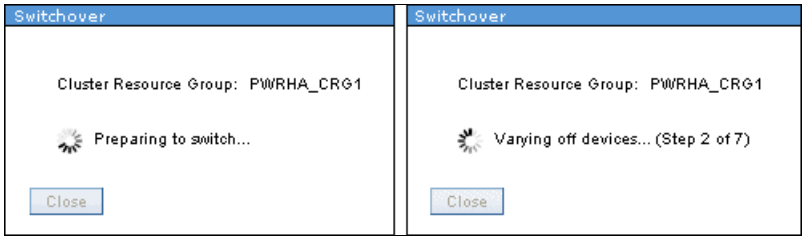


Figure 13-65 Metro Mirror switchover first steps

We might want to monitor the independent ASP vary status through a 5250 session using **DSPASPSTS** on the production system. Figure 13-66 shows the result. At this moment, the IASP is not yet varied off, but no job can access it anymore.

```
                          Display ASP Vary Status

 ASP Device . . . . :   IASP1           Step . . . . . . . :    5 / 5
 ASP Number . . . . :   33              Current time . . . :   00:00:13
 ASP State  . . . . :   VARIED ON       Previous time  . . :   00:00:58

       Step                                                     Elapsed
time
      Cluster vary job submission
      Ending jobs using the ASP                                 00:00:00
      Waiting for jobs to end                                   00:00:04
      Image catalog synchronization
 >    Writing changes to disk                                   00:00:08









 Press Enter to continue


 F3=Exit    F5=Refresh    F12=Cancel    F19=End automatic refresh
```

Figure 13-66 Metro Mirror: Independent ASP varying off on preferred production

Figure 13-67 shows the next switchover steps.

c. The third step ends the takeover IP interface on the production system.

d. The fourth step performs DS8000 activity for the Metro Mirror relationships to be reversed. **`failoverpprc`** is issued on the backup DS8000 for the target volumes to become source suspended, followed by **`failbackpprc`** for the volumes to be synchronized back from the backup DS8000 to the production DS8000.

e. The fifth steps ensures that the cluster resource group is started.

f. The sixth steps varies on the independent ASP on the backup system.

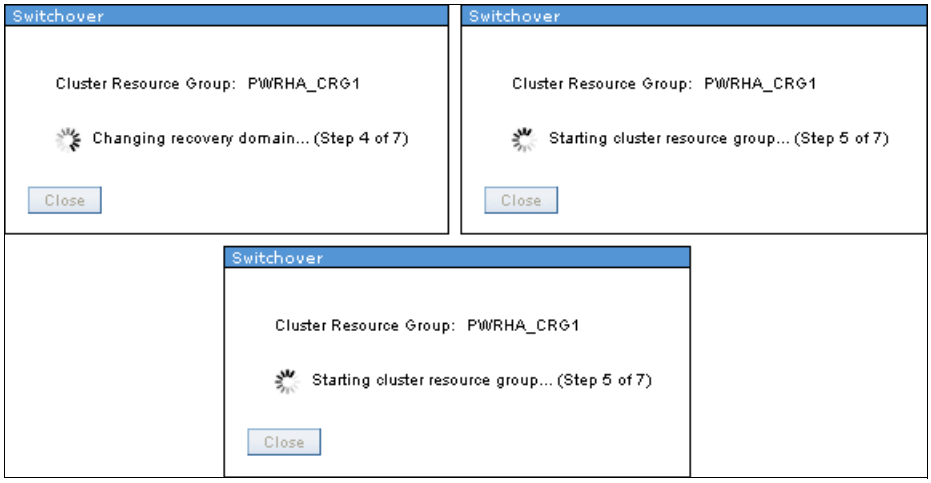


Figure 13-67 Metro Mirror switchover almost complete

We might want to monitor the independent ASP vary status through a 5250 session using **DSPASPSTS** on the backup system. Figure 13-68 shows the result. At this moment, the IASP is not yet available. The database cross-reference file merge is in progress.

```
                          Display ASP Vary Status

 ASP Device . . . . :   IASP1           Step . . . . . . . :   30 / 34
 ASP Number . . . . :   33              Current time . . . :   00:00:07
 ASP State  . . . . :   ACTIVE          Previous time  . . :   00:01:47

       Step                                                      Elapsed
time
      UID/GID mismatch correction                               00:00:00
      Database access path recovery                             00:00:01
 >    Database cross-reference file merge                       00:00:01
      SPOOL initialization
      Image catalog synchronization
      Command analyzer recovery
      Catalog validation




 Press Enter to continue


 F3=Exit   F5=Refresh   F12=Cancel   F19=End automatic refresh
```

Figure 13-68 Metro Mirror, Independent ASP varying on preferred backup

g. After the seventh step has completed, which is starting the takeover IP address on the backup system, the overall switchover is complete (Figure 13-69).

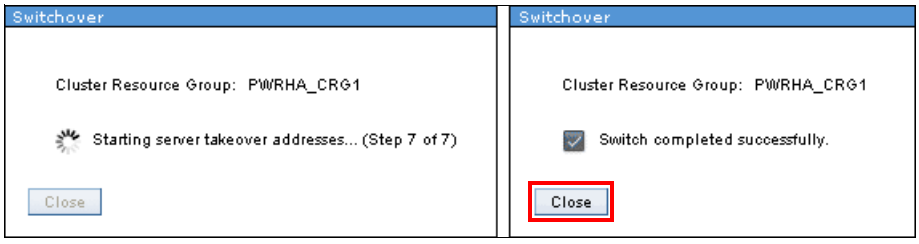


Figure 13-69 Metro Mirror switchover completed

5. After switchover completion, when coming back to the Cluster Resource Group panel (Figure 13-70), a refresh is needed to get the appropriate status. Click **Refresh**.

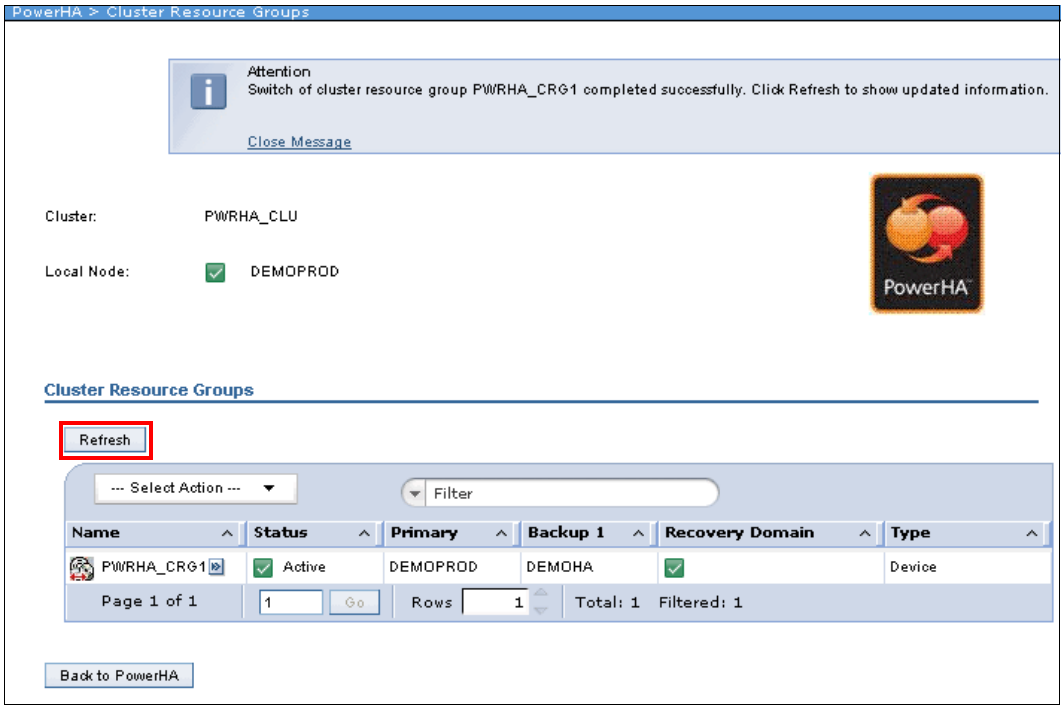


Figure 13-70 Cluster Resource Group panel after completed Metro Mirror switchover

6. We now have the correct status for the cluster resource group that we just switched over from production to backup, from the DEMOPROD to the DEMOHA node (Figure 13-71).

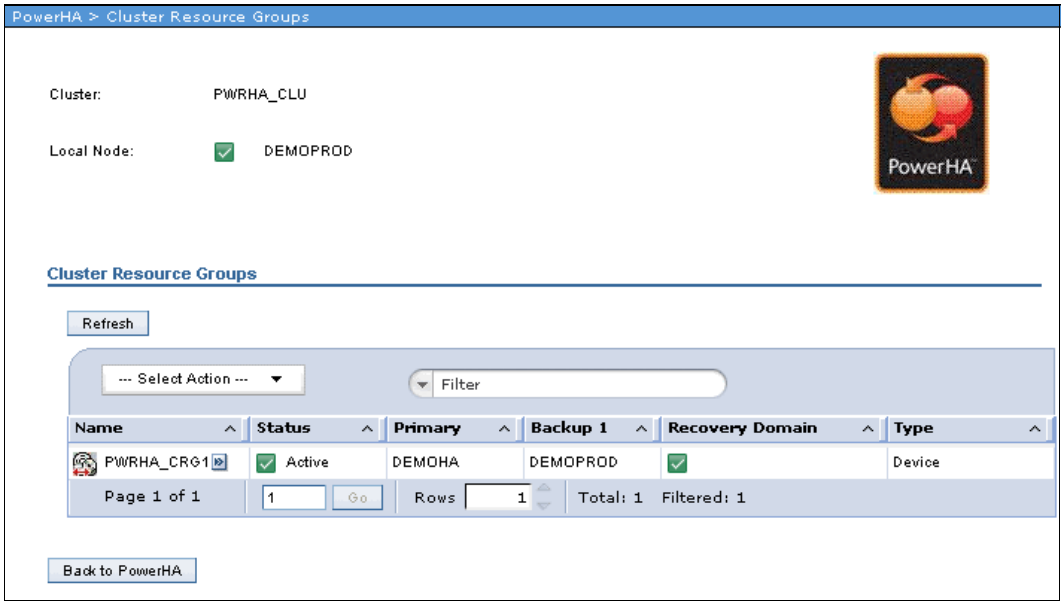


Figure 13-71 Metro Mirror cluster resource group status after switchover

After the switchover is complete, we can compare the situation to the one before the switchover. DEMOPROD is the current first backup system, and DEMOHA the current production system.

Figure 13-72 and Figure 13-73 on page 331 show the IP address configuration on both DEMOPROD and DEMOHA systems. On DEMOHA, takeover IP address 10.0.0.1 is active. On DEMOPROD, this IP address is now inactive.

```
                          Work with TCP/IP Interfaces
                                                                System:   DEMOPROD
Type options, press Enter.
  1=Add   2=Change   4=Remove   5=Display   9=Start   10=End

     Internet         Subnet           Interface     Alias
Opt  Address          Mask             Status        Name

     9.5.168.129      255.255.255.0    Active        *NONE
     9.5.168.133      255.255.255.0    Failed        *NONE
     9.5.168.134      255.255.255.0    Failed        *NONE
     10.0.0.1         255.255.255.0    Inactive      POWERHA_TAKEOVER_IP
     10.10.10.1       255.255.255.0    Active        POWERHA
     127.0.0.1        255.0.0.0        Active        LOCALHOST
     192.168.86.10    255.255.255.0    Active        *NONE
     192.168.87.10    255.255.255.0    Active        *NONE




                                                                          Bottom
F3=Exit      F5=Refresh   F6=Print list   F11=Display line information
F12=Cancel   F17=Top      F18=Bottom
```

Figure 13-72 Metro Mirror: IP configuration on preferred production after switchover

```
                          Work with TCP/IP Interfaces
                                                                 System:   DEMOHA
Type options, press Enter.
  1=Add   2=Change   4=Remove   5=Display   9=Start   10=End

     Internet         Subnet           Interface     Alias
Opt  Address          Mask             Status        Name

     9.5.168.130      255.255.255.0    Active        *NONE
     9.5.168.133      255.255.255.0    Active        *NONE
     9.5.168.134      255.255.255.0    Active        *NONE
     10.0.0.1         255.255.255.0    Active        POWERHA_TAKEOVER_IP
     10.10.10.2       255.255.255.0    Active        POWERHA
     127.0.0.1        255.0.0.0        Active        LOCALHOST
     192.168.86.11    255.255.255.0    Active        *NONE
     192.168.87.11    255.255.255.0    Active        *NONE




                                                                        Bottom
F3=Exit      F5=Refresh   F6=Print list   F11=Display line information
F12=Cancel   F17=Top      F18=Bottom
```

Figure 13-73 Metro Mirror: IP configuration on preferred backup after switchover

Figure 13-74 shows the current situation for DS volume members of the Metro Mirror relationships. With **lspprc** on the current backup DS8000 *75AY031*, all volumes are *target full duplex*, which is the normal status for target volumes. On the current production DS8000 *75AY032*, all volumes are *full duplex*, which is the normal status for source volumes.

```
dscli> lspprc -dev IBM.2107-75AY031 6000-61FF
Date/Time: September 29, 2011 9:50:13 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
ID        State              Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================================
6000:6000 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6001:6001 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6002:6002 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6100:6100 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
6101:6101 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
6102:6102 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid

dscli> lspprc -dev IBM.2107-75AY032 6000-61FF
Date/Time: September 29, 2011 9:53:23 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
ID        State       Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================================
6000:6000 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6001:6001 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6002:6002 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6100:6100 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6101:6101 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6102:6102 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
```

Figure 13-74 Metro Mirror volumes status after switchover

When it is scheduled to perform a switchback, all the operations detailed above are to be done, but on the other way. Switchback and switch invoke exactly the same operations.

Using CL commands for a Global Mirror planned switchover

Note: In contrast to the GUI, the CL command will not give you any possibility to switch over or switch back if there is one incorrect item.

Before activating the switch, we can check whether everything is fine from a clustering state perspective and that nothing will prevent the switchover from succeeding:

1. The first items that we check are the cluster consistency and the cluster resource group status. Use **WRKCLU**, then option 9 (Work with cluster resource groups) (Figure 13-75).

```
                              Work with Cluster
                                                          System:   DEMOPROD
Cluster  . . . . . . . . . . . . . . . . :   PWRHA_CLU

Select one of the following:

     1. Display cluster information
     2. Display cluster configuration information

     6. Work with cluster nodes
     7. Work with device domains
     8. Work with administrative domains
     9. Work with cluster resource groups
    10. Work with ASP copy descriptions

    20. Dump cluster trace



Selection or command
===> 9

F1=Help   F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel
```

Figure 13-75 Global Mirror switchback WRKCLU command panel

As shown in Figure 13-76, “Consistent information in cluster” must be Yes, and CRG status must be Active. To continue checking, select option 6 for Recovery domain against the cluster resource group, in our case PWRHA_CRG2.

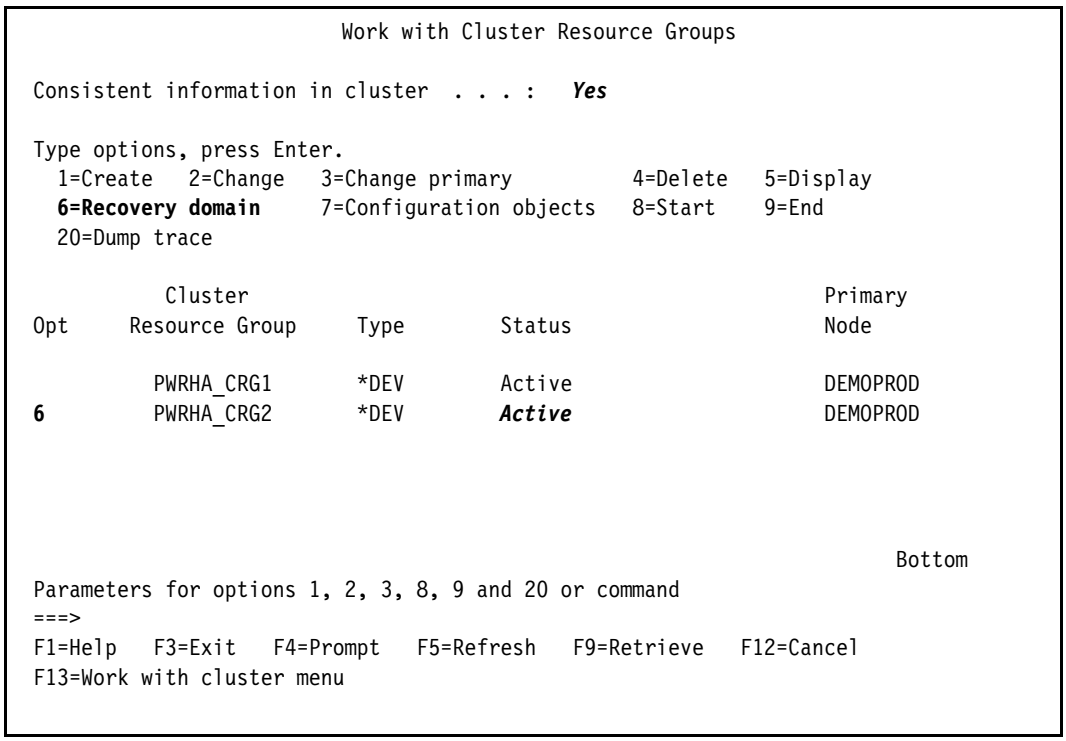


Figure 13-76 Global Mirror switch over Cluster Resource Group panel

2. As shown in Figure 13-77, the recovery domain status of both nodes must be active. We can also see their current roles. In our case, DEMOPROD is the production node and DEMOHA is the first backup. After the switch over, DEMOHA will be the production node and DEMOPROD the first backup node.

```
                            Work with Recovery Domain

Cluster resource group . . . . . :   PWRHA_CRG2
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add node   4=Remove node   5=Display more details   20=Dump trace

                                    Current        Preferred      Site
Opt     Node         Status         Node Role      Node Role      Name

        DEMOHA       Active         *BACKUP  1     *BACKUP  1     SITE2
        DEMOPROD     Active         *PRIMARY       *PRIMARY       SITE1

                                                                        Bottom
Parameters for options 1 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 13-77 Global Mirror switch version Recovery Domain panel

3. The CL command **DSPCRGINF** summarizes various information and statuses. In our case we run the following command:

   ```
   DSPCRGINF CLUSTER(*) CRG(PWRHA_CRG2)
   ```

   As shown in Figure 13-78 and Figure 13-79 on page 336, we must find the following information:

   – Consistent information in cluster is set to Yes.
   – Cluster resource group status is set to Active.
   – Both Recovery Domain nodes status is set to Active.
   – Both nodes roles are set to the expected values:
     • Primary
     • Backup

```
                            Display CRG Information

Cluster  . . . . . . . . . . . . :   PWRHA_CLU
Cluster resource group . . . . . :   PWRHA_CRG2
Reporting node . . . . . . . . . :   DEMOHA
Consistent information in cluster:   Yes

Cluster resource group type  . . . :   *DEV
Cluster resource group status  . . :   Active
Previous CRG status  . . . . . . . :   Switchover Pending
Exit program . . . . . . . . . . . :   *NONE
  Library  . . . . . . . . . . . . :     *NONE
Exit program job name  . . . . . . :   *NONE
Exit program format  . . . . . . . :   *NONE
Exit program data  . . . . . . . . :   *NONE

User profile . . . . . . . . . . . :   *NONE
Text . . . . . . . . . . . . . . . :   DS8000 Cluster Resource Group

                                                                      More...
Cluster  . . . . . . . . . . . . :   PWRHA_CLU
Cluster resource group . . . . . :   PWRHA_CRG2
Reporting node . . . . . . . . . :   DEMOHA
Consistent information in cluster:   Yes

Distribute information queue . . . :   *NONE
  Library  . . . . . . . . . . . . :     *NONE
Failover message queue . . . . . . :   *NONE
  Library  . . . . . . . . . . . . :     *NONE
Failover wait time . . . . . . . . :   *NOWAIT
Failover default action  . . . . . :   *PROCEED
CRG extended attribute . . . . . . :   *NONE
Application identifier . . . . . . :   *NONE
                                                                       Bottom
F1=Help   F3=Exit   F5=Refresh   F12=Cancel   Enter=Continue
```

Figure 13-78 Global Mirror switchover DSPCRGINF command first and second panel

```
                              Display CRG Information

Cluster  . . . . . . . . . . . . :   PWRHA_CLU
Cluster resource group . . . . . :   PWRHA_CRG2
Reporting node . . . . . . . . . :   DEMOHA
Consistent information in cluster:   Yes

                         Recovery Domain Information

                            Current        Preferred      Site
Node         Status         Node Role      Node Role      Name
DEMOPROD     Active         *PRIMARY       *PRIMARY       SITE1
DEMOHA       Active         *BACKUP  1     *BACKUP  1     SITE2









                                                                      Bottom
Number of recovery domain nodes  :   2


F1=Help   F3=Exit   F5=Refresh   F12=Cancel   Enter=Continue
```

Figure 13-79 Global Mirror switchover DSPCRGINF command fifth panel

4. After checking the cluster resources status, if you proceed with a switchover, end all applications using the IASP, as they will automatically be varied off by PowerHA when switching. There are two ways to initiate an IASP switchover. As shown in Figure 13-80, you can choose option 3 (Change primary against the cluster resource group), in our case PWRHA_CRG2. Or you can use **CHGCRGPRI CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG2)**. The same command runs under the covers when selecting option 3.

```
                          Work with Cluster Resource Groups

Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Create   2=Change   3=Change primary          4=Delete   5=Display
  6=Recovery domain     7=Configuration objects   8=Start    9=End
  20=Dump trace

           Cluster                                                 Primary
Opt     Resource Group     Type        Status                     Node

          PWRHA_CRG1       *DEV        Active                     DEMOPROD
3         PWRHA_CRG2       *DEV        Active                     DEMOPROD




                                                                          Bottom
Parameters for options 1, 2, 3, 8, 9 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 13-80 Global Mirror switch over from Cluster Resource Group panel

5. After the switchover is complete, the following message is shown:

   ```
   Cluster Resource Services API QcstInitiateSwitchOver completed
   ```

   There is no progress indicator when running the command, but it performs the same steps as described in “Using the PowerHA GUI for a Metro Mirror planned switchover” on page 321.

6. You can verify that the recovery domain nodes are in the expected active state and that their role is not their preferred one anymore. In our case, DEMOHA is the current production node and DEMOPROD is currently the first backup node (Figure 13-81).

```
                              Work with Recovery Domain

Cluster resource group . . . . . . . . . . :   PWRHA_CRG2
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add node   4=Remove node   5=Display more details   20=Dump trace

                                    Current        Preferred      Site
Opt     Node         Status         Node Role      Node Role      Name

        DEMOHA       Active         *PRIMARY       *BACKUP  1     SITE2
        DEMOPROD     Active         *BACKUP  1     *PRIMARY       SITE1









                                                                        Bottom
Parameters for options 1 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 13-81 Global Mirror switchover Cluster Resource Group panel

During the switchover, on both DS8000s the following actions are submitted by PowerHA during step 4 (Changing Recovery Domain) reported by the GUI:

a. Check the Global Copy status on the preferred source DS8000.

b. Check the Global Mirror LSS status on the preferred source DS8000.

c. Pause the Global Mirror session on the preferred source DS8000.

d. Remove volumes from the Global Mirror session on the preferred source DS8000.

e. Resume the Global Mirror session on the preferred source DS8000.

f. Fail over the Global Copy relationships on the preferred target DS8000.

g. Fast reverse restore FlashCopy relationships on the preferred target DS8000 to put consistent data on newly Global Copy source volumes.

   The persistent option is not set in the DS8000 command, which means that the FlashCopy relationships no longer exist at the end of reverseflash.

h. Fail back Global Copy relationships on the preferred target DS8000 to synchronize the current production volumes with the current backup volumes on the preferred source DS8000.

i. Create the FlashCopy relationships on the current target DS8000 to build a consistency group.

j. Update the Global Mirror session on the current source DS8000 to include the current source volumes.

k. Start the Global Mirror session on the current source DS8000.

Using CL commands for an asymmetrical Global Mirror switchback after a planned outage

In this section we focus on describing how to switch back to the preferred production node in an asymmetrical Global Mirror configuration. Manual steps are needed on the DS storage server to re-establish Global Mirror in the original direction from the preferred source to the preferred target DS8000.

The main difference is that D volumes do not exist when compared to our scenario setup in Figure 13-56 on page 313. Therefore, when creating the ASP copy description for the preferred source IASP, we do not specify the consistency group volumes (Example 13-20).

Example 13-20 Source ASP copy description creation command for an asymmetrical Global Mirror

```
ADDASPCPYD ASPCPY(IASP2_GM1) ASPDEV(IASP2) CRG(PWRHA_CRG2) LOCATION(*DEFAULT)
   SITE(SITE1) STGHOST('qlpar' 'password' ('9.5.168.32'))
   LUN('IBM.2107-13ABGAA' ('0200-0201' '0300-0301') ())
```

For a switchover to the preferred backup site in an asymmetrical Global Mirror configuration, PowerHA cannot reverse Global Mirror because there have been no FlashCopy consistency group volumes configured on the preferred source DS8000. Therefore, at the end of the switchover the Global Copy relationship volume state on the preferred backup DS8000, which is the current production DS8000, is source suspended. They are not members of the Global Mirror session on the preferred backup DS8000 and the FlashCopy relationships no longer exist.

After the switchover, if you do not wait for the next occurrence of the QYASSTEHCK checking job (which checks that replication is running properly), and although all PowerHA configuration checks (such as cluster node is active, cluster resource group is active, nodes in recovery domain are active) show that everything is correct for running a switchback with **CHGCRGPRI**, you are not allowed to do so (Figure 13-82).

```
Command Entry                          DEMOHA
                                                        Request level:   7
Previous commands and messages:
   > CHGCRGPRI CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG2)
     Cluster resource group exit program QYASSTEHCK in library QHASM on node
       DEMOHA failed.
     Error invoking exit programs on node DEMOHA.
     Nodes in cluster resource group PWRHA_CRG2 at mirror copy or target site
       are not eligible to become a primary node.
     Cluster resource group exit program QYASSTEHCK in library QHASM on node
       DEMOPROD failed.
     Error invoking exit programs on node DEMOPROD.
     A switch over can not be done for cluster resource group PWRHA_CRG2.
     A switch over can not be done for cluster resource group PWRHA_CRG2.
     Cluster Resource Services API QcstInitiateSwitchOver completed.
     Primary node of cluster resource group PWRHA_CRG2 not changed.
Type command, press Enter.
===> CHGCRGPRI CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG2)



F3=Exit   F4=Prompt   F9=Retrieve   F10=Include detailed messages
F11=Display full      F12=Cancel    F13=Information Assistant   F24=More keys
```

Figure 13-82 Asymmetrical Global Mirror switchback fails

The reason for PowerHA preventing a switchback to the preferred source DS8000 in an asymmetrical Global Mirror configuration is that we miss the consistency group volumes, which are intended to ensure consistency within a Global Mirror relationship. Without this consistency group volume, we cannot ensure that consistency exists, at least when source independent ASP is varied on, or Global Copy has tracks not yet written on target. After the command fails, the preferred production node status changes from Active to Ineligible (Figure 13-83).

```
                          Work with Recovery Domain

Cluster resource group . . . . . :   PWRHA_CRG2
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add node   4=Remove node   5=Display more details   20=Dump trace

                                    Current        Preferred      Site
Opt     Node         Status         Node Role      Node Role      Name

        DEMOHA       Active         *PRIMARY       *BACKUP  1     SITE2
        DEMOPROD     Ineligible     *BACKUP  1     *PRIMARY       SITE1









                                                                        Bottom
Parameters for options 1 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 13-83 Asymmetrical Global Mirror node ineligible status

Figure 13-84 shows details about the CPDBB0D message ID.

```
Additional Message Information

Message ID . . . . . :   CPDBB0D       Severity . . . . . :   30
Message type . . . . . :   Diagnostic
Date sent  . . . . . :   10/04/11      Time sent  . . . . . :   14:03:04

Message . . . . :   Nodes in cluster resource group PWRHA_CRG2 at mirror copy
  or target site are not eligible to become a primary node.
Cause . . . . . :   The backup nodes in cluster resource group PWRHA_CRG2 at
  mirror copy or target site are not eligible to become a primary node, for
  one or more of the following reasons:
     1 -- One or more auxiliary storage pools are missing.
     2 -- Cross site mirroring has not been configured for auxiliary storage
  pools in cluster resource group PWRHA_CRG2.
     3 -- Cross site mirroring is being synchronized.
     4 -- One or more mirror copies or targets of auxiliary storage pools are
  not usable and offline.
     5 -- One or more mirror copies or targets of auxiliary storage pools are
  not usable and cross site mirroring is suspended.
     6 -- Cross site mirroring is suspended for one or more mirror copies or
  targets of auxiliary storage pools.
Recovery  . . . :   Recovery actions for each reason code are:
     1 -- All auxiliary storage pools must be present. Determine the cause for
  the missing ones and fix the problem.
     2 -- Configure cross site mirroring for auxiliary storage pools.
     3 -- Wait for the synchronization to complete.
     4, 5 and 6 -- Resume cross site mirroring if it is suspended. Vary on the
  auxiliary storage pool and wait for the synchronization to complete.

Press Enter to continue.

F3=Exit   F6=Print   F9=Display message details   F12=Cancel
F21=Select assistance level
```

Figure 13-84 Asymmetrical Global Mirror switchback error detail

At the same time, the following message ID HAI2001 is sent to the cluster resource group or cluster message queue (Figure 13-85):

```
Cross-site mirroring (XSM) for ASP IASP2 is not active.
```

```
                        Additional Message Information

Message ID . . . . . :   HAI2001       Severity . . . . . :   10
Message type . . . . . :   Information
Date sent  . . . . . :   10/04/11      Time sent  . . . . . :   14:03:04

Message . . . . :   Cross-site mirroring (XSM) for ASP IASP2 is not active.
Cause . . . . . :   The cross-site mirroring copy state for auxiliary storage
  pool (ASP) IASP2 is not active.  The copy state is 12.  The source
  clustering node is DEMOHA and the target clustering node is DEMOPROD.
    The copy states are:
    -1 -- The ASP copy state is not known.
    11 -- The ASP copy is being synchronized.
    12 -- The ASP copy is suspended.
Recovery  . . . :   Recovery actions for the copy states are:
    -1 -- Verify that cross site mirroring is correctly configured and correct
  any configuration errors.
    11 -- Wait for the synchronization to complete.
    12 -- Perform a resume or reattach operation on the ASP.

Press Enter to continue.

F3=Exit   F6=Print   F9=Display message details   F12=Cancel
F21=Select assistance level
```

Figure 13-85 Asymmetrical Global Mirror switchback

The message suggests that you either perform a resume or a reattach operation on the ASP. In our case, the mirroring is not only suspended, but it is also in a failover status. This means that we need to run a reattach operation for the ASP session. However, we also need to establish the Global Mirror session in the proper direction again. This is the reason why in this special case, we have to run the following operations:

1. Make sure that independent ASPs are varied off on both sides.
2. On the current production DS, perform **failbackpprc**. This synchronizes data from the current production DS to the current backup DS (Example 13-21).

Example 13-21 Asymmetrical Global Mirror switch back, synchronize preferred DS with current DS

```
dscli> failbackpprc -dev IBM.2107-1375210 -remotedev IBM.2107-13ABGAA -type gcp 0200-0201:0200-0201
0300-0301:0300-0301
Date/Time: October 4, 2011 3:48:19 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-1375210
CMUC00197I failbackpprc: Remote Mirror and Copy pair 0200:0200 successfully failed back.
CMUC00197I failbackpprc: Remote Mirror and Copy pair 0201:0201 successfully failed back.
CMUC00197I failbackpprc: Remote Mirror and Copy pair 0300:0300 successfully failed back.
CMUC00197I failbackpprc: Remote Mirror and Copy pair 0301:0301 successfully failed back.
dscli> lspprc 0200-03FF
Date/Time: October 4, 2011 3:48:23 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-1375210
ID        State        Reason Type        SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================
0200:0200 Copy Pending -      Global Copy 02        60             Disabled      True
0201:0201 Copy Pending -      Global Copy 02        60             Disabled      True
```

```
0300:0300 Copy Pending -      Global Copy 03        60             Disabled      True
0301:0301 Copy Pending -      Global Copy 03        60             Disabled      True
```

3. On the current production node make sure that the replication is finished, as indicated by a synchronization progress of 100 in the **DSPASPSSN** command panel (Figure 13-86).

```
                              Display ASP Session                     DEMOHA
                                                              10/04/11  15:48:52
Session . . . . . . . . . . . . . . . . . . :   IASP2_GM
  Type  . . . . . . . . . . . . . . . . . . :     *GLOBALMIR
  Synchronization progress  . . . . . . . . :      100




                             Copy Descriptions

ASP
device         Name                 Role          State      Node
IASP2          IASP2_GM2          SOURCE        VARYOFF      DEMOHA
IASP2          IASP2_GM1          TARGET       RESUMING      DEMOPROD


                                                                        Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel   F19=Automatic refresh
```

Figure 13-86 Asymmetrical Global Mirror switchback, synchronizing back current to preferred production DS8000

4. Make the preferred source DS8000 the source of the Global Copy replication. On the preferred source DS8000, run the pair of **failoverpprc**/**failbackpprc** commands (Example 13-22).

Example 13-22 Asymmetrical Global Mirror switchback, making the preferred source DS8000 the current source

```
dscli> lspprc 0200-03FF
Date/Time: October 4, 2011 4:02:03 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
ID        State              Reason Type        SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================================
0200:0200 Target Copy Pending -      Global Copy 02        unknown        Disabled      Invalid
0201:0201 Target Copy Pending -      Global Copy 02        unknown        Disabled      Invalid
0300:0300 Target Copy Pending -      Global Copy 03        unknown        Disabled      Invalid
0301:0301 Target Copy Pending -      Global Copy 03        unknown        Disabled      Invalid
dscli> failoverpprc -dev IBM.2107-13ABGAA -remotedev IBM.2107-1375210 -type gcp 0200-0201:0200-0201
0300-0301:0300-0301
Date/Time: October 4, 2011 4:09:18 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
CMUC00196I failoverpprc: Remote Mirror and Copy pair 0200:0200 successfully reversed.
CMUC00196I failoverpprc: Remote Mirror and Copy pair 0201:0201 successfully reversed.
CMUC00196I failoverpprc: Remote Mirror and Copy pair 0300:0300 successfully reversed.
CMUC00196I failoverpprc: Remote Mirror and Copy pair 0301:0301 successfully reversed.
dscli> failbackpprc -dev IBM.2107-13ABGAA -remotedev IBM.2107-1375210 -type gcp 0200-0201:0200-0201
0300-0301:0300-0301
Date/Time: October 4, 2011 4:09:30 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
CMUC00197I failbackpprc: Remote Mirror and Copy pair 0200:0200 successfully failed back.
```

```
CMUC00197I failbackpprc: Remote Mirror and Copy pair 0201:0201 successfully failed back.
CMUC00197I failbackpprc: Remote Mirror and Copy pair 0300:0300 successfully failed back.
CMUC00197I failbackpprc: Remote Mirror and Copy pair 0301:0301 successfully failed back.
dscli> lspprc 0200-03FF
Date/Time: October 4, 2011 4:09:33 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
ID        State        Reason Type        SourceLSS Timeout (secs) Critical Mode First Pass Status
==========================================================================================================
0200:0200 Copy Pending -      Global Copy 02        60             Disabled      True
0201:0201 Copy Pending -      Global Copy 02        60             Disabled      True
0300:0300 Copy Pending -      Global Copy 03        60             Disabled      True
0301:0301 Copy Pending -      Global Copy 03        60             Disabled      True
```

5. The Global Mirror configuration also needs to be built again on the preferred source DS8000. Refer to the following sections:
   - “Making a Global Mirror session” on page 315

     The sessions should already exist. Just replace the **`mksession`** commands with **`chsession`** ones.
   - “Creating the FlashCopy relationships on the backup site” on page 314
   - “Starting Global Mirror session” on page 316.

   Note: You might need to remove the Global Mirror session using the **`rmgmir`** DS8000 command before being allowed to create it again.

6. Reverse IBM i cluster node roles:
   a. End the cluster resource group.
   b. Change the current role for the current backup node to the primary and vice versa with **`CHGCRG`** (Example 13-23).

      Example 13-23 Asymmetrical Global Mirror failback: Change CRG role command

      ```
      CHGCRG CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG2) CRGTYPE(*DEV) RCYDMNACN(*CHGCUR)
         RCYDMN(  (DEMOPROD *PRIMARY *SAME SITE1 *SAME *SAME)
                  (DEMOHA *BACKUP 1 SITE2 *SAME *SAME))
      ```

   c. Restart the cluster resource group.
   d. Wait for the backup node status to change from Ineligible to Active. This is done by the PowerHA healthcheck job QYASSTEHCK, which runs every hour.

After completing these steps you have re-established Global Mirror in its original direction from the preferred source to the preferred target DS8000.

13.2.2 Using CL commands for DS8000 LUN-level switching

To perform a planned LUN-level switching, follow these steps:

Note: Before starting any switchover make sure that no jobs are using the IASP anymore.

1. Make sure that “Consistent information in cluster” is Yes. Set the CRG status (in our case PWRHA_CRG3) to Active (Figure 13-87) by using **WRKCLU**, then option 9 (Work with cluster resource groups).

```
                        Work with Cluster Resource Groups

Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Create   2=Change   3=Change primary          4=Delete   5=Display
  6=Recovery domain     7=Configuration objects   8=Start    9=End
  20=Dump trace

           Cluster                                                Primary
Opt     Resource Group     Type        Status                     Node

          PWRHA_CRG3       *DEV        Active                     DEMOFC



                                                                          Bottom
Parameters for options 1, 2, 3, 8, 9 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 13-87 Cluster Resource Group status

2. Using option 6 (Recovery Domain) on the Work with Cluster Resource Groups panel, check that the recovery domain status of both nodes (DEMOFC as the production node and DEMOHA as the backup node) are active and that their roles are the ones that you expect (Figure 13-88).

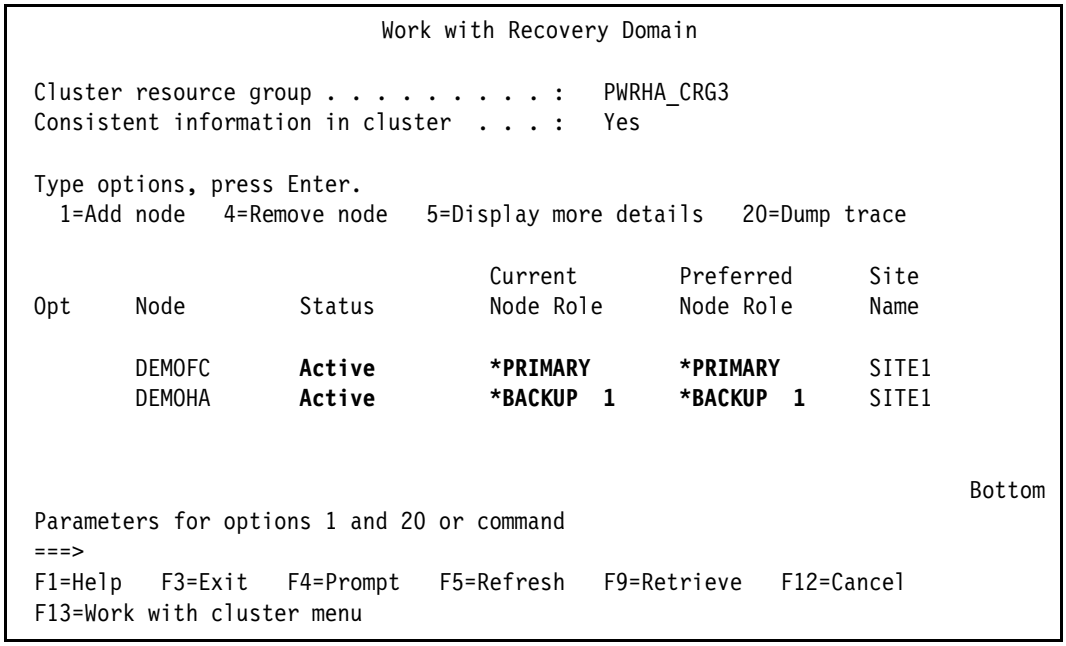


Figure 13-88 Node status in recovery domain

Alternatively, use **DSPCRGINF** to summarize the information and status of the CRG.

3. Use **CHGCRGPRI** to switch the IASP from the primary node to the backup node (Example 13-24).

Example 13-24 CHGCRGPRI command for switching the node

```
CHGCRGPRI CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG3)
```

Also, you can switch the IASP by using option 3 (Change primary) on the Work with Cluster Resource Group panel (Figure 13-87 on page 346).

While switching with the CL command there is no process indicator. Use **DSPASPSTS** to check the status of IASP on another session from the production and backup nodes (Figure 13-89).

```
                          Display ASP Vary Status

ASP Device . . . . :   IASP_LUN          Step . . . . . . . :   30 / 34
ASP Number . . . . :   71                Current time . . . :   00:00:07
ASP State  . . . . :   ACTIVE            Previous time  . . :   00:01:47

      Step                                                    Elapsed time
      UID/GID mismatch correction                               00:00:00
      Database access path recovery                             00:00:01
>     Database cross-reference file merge                       00:00:01
      SPOOL initialization
      Image catalog synchronization
      Command analyzer recovery
      Catalog validation


Press Enter to continue

F3=Exit    F5=Refresh    F12=Cancel    F19=End automatic refresh
```

Figure 13-89 IASP status on the backup node during switching

4. After the switchover is complete, the following message is shown on the bottom of panel and you can also see the change of the primary node for CRG PWRHA_CR3 from DEMOFC to DEMOHA on the Work with Cluster Resource Groups panel (Figure 13-90)

```
Cluster Resource Services API QcstInitiateSwitchOver completed
```

```
                      Work with Cluster Resource Groups

Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Create   2=Change   3=Change primary          4=Delete   5=Display
  6=Recovery domain     7=Configuration objects   8=Start    9=End
  20=Dump trace

           Cluster                                                Primary
Opt     Resource Group     Type        Status                     Node

          PWRHA_CRG3       *DEV        Active                     DEMOHA



                                                                        Bottom
Parameters for options 1, 2, 3, 8, 9 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
Cluster Resource Services API QcstInitiateSwitchOver completed.              +
```

Figure 13-90 CRG status after switching

You can verify that the Recovery Domain nodes are in the expected active status and that their role is the preferred backup status. In our case, DEMOHA is the production node and DEMOFC is the backup node (Figure 13-91).

```
                          Work with Recovery Domain

Cluster resource group . . . . . :   PWRHA_CRG3
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add node   4=Remove node   5=Display more details   20=Dump trace

                                  Current        Preferred      Site
Opt     Node         Status       Node Role      Node Role      Name

        DEMOFC       Active       *BACKUP  1     *PRIMARY       SITE1
        DEMOHA       Active       *PRIMARY       *BACKUP  1     SITE1



                                                                    Bottom
Parameters for options 1 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 13-91 Node status in recovery domain after switching

On DS8000 you can also verify that the host connection for the backup node is automatically assigned to the volume group that was assigned to the production node (Example 13-25).

Example 13-25 Host connection on DS8000 after switching

```
dscli> lshostconnect 2 5
Date/Time: September 28, 2011 4:31:36 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
Name            ID   WWPN             HostType Profile              portgrp volgrpID ESSIOport
==========================================================================================
demofc_powerha 0002 10000000C9523E9D iSeries  IBM iSeries - OS/400       0 -        all
DEMOHA_FC      0005 10000000C94122A2 iSeries  IBM iSeries - OS/400       0 V48      all
```

5. Verify whether the IASP has been varied on (that is, that it is in available status) at the backup node using the following command:

   ```
   WRKCFGSTS CFGTYPE(*DEV) CFGD(*ASP)
   ```

   If the configuration object online parameter in the device CRG is set to *ONLINE, PowerHA automatically varies on the IASP on node at a switch over.

13.2.3 Failing over and back for an unplanned outage

In this section we describe the actions being performed for an unplanned failover and switchback in a Metro Mirror or Global Mirror environment and a LUN-level switching environment.

Metro Mirror and Global Mirror failover and failback

Regarding unplanned outages, we distinguish three cases that apply to both Metro Mirror and Global Mirror configurations:

- A failover for a primary node failure with a panic message

  In this case, the production node is able to send a specific message to the backup node stating that the production node is being terminated in an abnormal way.

- A failover for a primary node without the panic message

  In this case, the backup node might receive primary node status information from the HMC or VIOS.

- A primary node partition condition

  In this case, the backup node does not find any way to determine the primary node status.

Failover events and actions you will take for each of these cases are detailed here:

- To simulate a primary node failure *with* a panic message, we force the primary node to go into a restricted stated using **ENDSBS SBS(*ALL) OPTION(*IMMED)**. (Using **PWRDWNSYS** would lead to the same results.) The primary node will be able to send a panic message to the backup node. After receiving this message, if the backup node is again able to establish a heartbeat confirmation, the backup node invokes the automatic failover operations.

  As shown in Figure 13-92, an inquiry the following message is logged in the defined cluster message queue on the backup node (in our case QSYSOPR):

  `CPABB02 "Cluster resource groups are failing over to node DEMOHA. (G C)"`

  By using the cluster FLVDFTACN parameter default setting *PROCEED, the failover continues automatically after expiration of the cluster failover time (the FLVWAITTIM parameter, in our case *NOMAX). Also, the following informational message is logged, indicating the name of the CRG that is being failed over:

  `CPIBB18 "Cluster resource group PWRHA_CRG1 is failing over from node DEMOPROD to node DEMO."`

  Replying G for go initiates the failover procedure.

```
                              Display Messages
                                                          System:   DEMOHA
Queue . . . . . :   QSYSOPR               Program . . . . :   *DSPMSG
  Library . . . :     QSYS                   Library . . . :
Severity  . . . :   99                     Delivery  . . . :   *HOLD

Type reply (if required), press Enter.
  Cluster resource group PWRHA_CRG1 is failing over from node DEMOPROD to
    node DEMOHA.
  Cluster resource groups are failing over to node DEMOHA.  (G C)
    Reply . . .




Bottom
 F3=Exit             F11=Remove a message               F12=Cancel
 F13=Remove all      F16=Remove all except unanswered    F24=More keys
```

Figure 13-92 Failover inquiry message in the cluster message queue

Note: If using the cluster default parameter settings FLVDFTACT=*PROCEED and FLVWAITTIM=*NOWAIT, an automatic failover occurs instantly. Whereas, when using FLVWAITTIM=*NOMAX, the cluster waits forever for the user to reply on the cluster message queue's CPABB02 inquiry message to cancel or proceed with the automatic failover.

Metro Mirror (or Global Mirror) IASP volume relationships are failed over to the secondary. That is, the Metro Mirror (or Global Mirror) target volumes become suspended source volumes (Figure 13-93). In our case, on backup DS8000 75AY032, volumes 6000 - 6002 and 6100 - 6102 are suspended.

```
dscli> lspprc -dev IBM.2107-75AY032 6000-61FF
Date/Time: September 30, 2011 9:38:57 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
ID        State     Reason      Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================================
6000:6000 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6001:6001 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6002:6002 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6100:6100 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid
6101:6101 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid
6102:6102 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid
```

Figure 13-93 Suspended source volumes

The IASP (in our case IASP1) is varied on for the node at the secondary site in our example, as we used the CRG configuration object *ONLINE setting. Also, the takeover IP address is started on the secondary site.

Cluster node roles are changed so that the node at the secondary site becomes the primary and the node at the failed primary site becomes the backup (Figure 13-94), which is the recovery domain for our cluster resource group PWRHA_CRG1. In our case, DEMOHA becomes the primary node and DEMOPROD becomes the first backup node with an *inactive* status.

```
                          Work with Recovery Domain

 Cluster resource group . . . . . . . . . . :   PWRHA_CRG1
 Consistent information in cluster  . . . :   Yes

 Type options, press Enter.
   1=Add node   4=Remove node   5=Display more details   20=Dump trace

                                    Current        Preferred      Site
 Opt     Node         Status        Node Role      Node Role      Name

         DEMOHA       Active        *PRIMARY       *BACKUP  1     SITE2
         DEMOPROD     Inactive      *BACKUP  1     *PRIMARY       SITE1




Bottom
 Parameters for options 1 and 20 or command
 ===>
 F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
 F13=Work with cluster menu
```

Figure 13-94 Metro Mirror failover recovery domain

When we are ready to fail back, the following steps needs to be done:

a. Make sure that the IASP on the preferred primary node is varied off.

b. Make sure that the cluster node on the preferred primary node is started. If it is not, use **STRCLUNOD**. In our case:

   ```
   STRCLUNOD CLUSTER(PWRHA_CLU) NODE(DEMOPROD)
   ```

   The status of the preferred primary node (in our case DEMOPROD) within the recovery domain becomes Ineligible (Figure 13-95).

```
                          Work with Recovery Domain

Cluster resource group . . . . . . . . . . :   PWRHA_CRG1
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add node   4=Remove node   5=Display more details   20=Dump trace

                                    Current        Preferred      Site
Opt     Node         Status         Node Role      Node Role      Name

        DEMOHA       Active         *PRIMARY       *BACKUP  1     SITE2
        DEMOPROD     Ineligible     *BACKUP  1     *PRIMARY       SITE1

                                                                        Bottom
Parameters for options 1 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 13-95 Metro Mirror failback recovery domain

c. As shown in Figure 13-96, the current status of the ASP session (in our case IASP1_MM) can be displayed with **DSPASPSSN**. The replication status is *suspended* and the preferred primary role is *detached*.

```
                          Display ASP Session                    DEMOPROD
                                                          09/30/11  10:47:41
Session . . . . . . . . . . . . . . . . . . . :   IASP1_MM
  Type  . . . . . . . . . . . . . . . . . . . :     *METROMIR




                            Copy Descriptions

ASP
device         Name                 Role          State      Node
IASP1          IASP1_MM2          SOURCE        UNKNOWN      DEMOHA
IASP1          IASP1_MM1        DETACHED      SUSPENDED      DEMOPROD


                                                                    Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel   F19=Automatic refresh
```

Figure 13-96 Metro Mirror failback session status

d. Establish back the replication from the preferred backup to the preferred production DS.

   i. For Metro Mirror or symmetrical Global Mirror, re-attach the detached ASP session on the *preferred primary node*. That is, start the Metro Mirror (or symmetrical Global Mirror) replication from the preferred secondary DS8000 to the preferred primary DS8000, with the following command:

   ```
   CHGASPSSN SSN(IASP1_MM) OPTION(*REATTACH) ASPCPY((IASP1_MM1 IASP1_MM2))
   ```

   By default, this command requires an answer to the following message in the QSYSOPR message queue:

   ```
   HAA2000 "Reattach of ASP session IASP1_MM was requested. (C G)"
   ```

The second level of the message provides all the necessary information (Figure 13-97).

```
                        Additional Message Information

Message ID . . . . . :   HAA2000       Severity . . . . . :   99
Message type . . . . . :   Inquiry
Date sent  . . . . . :   09/30/11      Time sent  . . . . . :   11:03:16

Message . . . . :   Reattach of ASP session IASP1_MM was requested. (C G)
Cause . . . . . :   A request was made to reattach auxiliary storage pool
  (ASP) session IASP1_MM.  This operation will start replication from cluster
  node DEMOHA to cluster node DEMOPROD.
Recovery  . . . :   Do one of the following:
     -- Type G to continue the reattach.
     -- Type C to cancel the reattach.
Possible choices for replying to message . . . . . :
  C -- Processing is terminated.
  G -- Processing is continued.

                                                                        Bottom
Type reply below, then press Enter.
Reply  . . . .

F3=Exit   F6=Print   F9=Display message details   F12=Cancel
F21=Select assistance level
```

Figure 13-97 Metro Mirror failback re-attach confirmation

ii. For asymmetrical Global Mirror, start the replication from the preferred secondary DS8000 to the preferred primary DS8000. This must be done with **failbackpprc** on the preferred secondary DS8000 (Example 13-21 on page 343).

e. After establishing back the replication from the preferred backup to the preferred production DS, we can see the ASP session status on the preferred backup node with **DSPASPSSN** as show in Figure 13-98 for Metro Mirror or as shown in Figure 13-99 for Global Mirror. We can compare with the PPRC relationships on both the preferred backup DS8000 and the preferred production DS8000 as shown in Figure 13-100 on page 357 for Metro Mirror and as shown in Figure 13-101 on page 357 for Global Mirror.

```
Display ASP Session                     DEMOHA
                                                        09/30/11  11:10:45
Session . . . . . :   IASP1_MM
  Type  . . . . . :     *METROMIR

                            Copy Descriptions

ASP
device         Name                Role          State     Node
IASP1          IASP1_MM2         SOURCE      AVAILABLE     DEMOHA
IASP1          IASP1_MM1         TARGET         ACTIVE     DEMOPROD



                                                                    Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel   F19=Automatic refresh
```

Figure 13-98 Metro Mirror failback ASP session status

```
Display ASP Session                     DEMOHA
                                                        10/05/11  10:32:42
Session . . . . . :   IASP2_GM
  Type  . . . . . :     *GLOBALMIR
  Synchronization progress  . . . . . :      100







                            Copy Descriptions

ASP
device         Name                Role          State     Node
IASP2          IASP2_GM2         SOURCE      AVAILABLE     DEMOHA
IASP2          IASP2_GM1         TARGET       RESUMING     DEMOPROD



                                                                    Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel   F19=Automatic refresh
```

Figure 13-99 Global Mirror failback ASP session status

```
dscli> lspprc -dev IBM.2107-75AY032 6000-61FF
Date/Time: September 30, 2011 11:16:55 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
ID        State       Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================
6000:6000 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6001:6001 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6002:6002 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6100:6100 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6101:6101 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6102:6102 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid

dscli> lspprc -dev IBM.2107-75AY031 6000-61FF
Date/Time: September 30, 2011 11:17:00 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
ID        State              Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=======================================================================================================
6000:6000 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6001:6001 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6002:6002 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6100:6100 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
6101:6101 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
6102:6102 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
```

Figure 13-100 Metro Mirror failback PPRC relationships status

```
dscli> lspprc -dev IBM.2107-1375210 0200-0300
Date/Time: October 5, 2011 10:37:29 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-1375210
ID        State        Reason Type        SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================
0200:0200 Copy Pending -      Global Copy 02        60             Disabled      True
0201:0201 Copy Pending -      Global Copy 02        60             Disabled      True
0300:0300 Copy Pending -      Global Copy 03        60             Disabled      True
0301:0301 Copy Pending -      Global Copy 03        60             Disabled      True
dscli> lspprc -dev IBM.2107-13ABGAA 0200-0300
Date/Time: October 5, 2011 10:38:15 AM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-13ABGAA
ID        State               Reason Type        SourceLSS Timeout (secs) Critical Mode First Pass Status
=======================================================================================================
0200:0200 Target Copy Pending -      Global Copy 02        unknown        Disabled      Invalid
0201:0201 Target Copy Pending -      Global Copy 02        unknown        Disabled      Invalid
0300:0300 Target Copy Pending -      Global Copy 03        unknown        Disabled      Invalid
0301:0301 Target Copy Pending -      Global Copy 03        unknown        Disabled      Invalid
```

Figure 13-101 Global Mirror failback PPRC relationships status

Now we can perform a switchback to the preferred primary node, as described in section 13.2.1, “Switchover and switchback for a Metro Mirror or Global Mirror planned outage” on page 320.

- To simulate a primary node failure *without* a panic message, we force an immediate power off from the HMC on primary node partition. This action simulates a server failure. In this case, the primary node is not able to send a panic message to the backup node.

If using the advanced node failure detection feature, as we do in our example f(or which the configuration steps are described in 11.2, “Setting up cluster monitors” on page 208), the HMC CIMOM server is able to send information about partition failures to the registered IBM i cluster nodes. The preferred production node status becomes *failed* (Figure 13-102). (DEMOPROD is the preferred production node in our example.) Therefore, the backup node is able to invoke automatic failover operations just like our first case above. It sends to the cluster message queue the same inquiry message CPABB02, with the same considerations about wait time, then it performs the same failover operations.

```
                              Work with Cluster Nodes

Local node . . . . . . . . . . . . . . . . :   DEMOHA
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add   2=Change   4=Remove   5=Display more details   6=Work with monitors
  8=Start   9=End    20=Dump trace

Opt     Node         Status         Device Domain

        DEMOFC       Active          PWRHA_DMN
        DEMOHA       Active          PWRHA_DMN
        DEMOPROD     Failed          PWRHA_DMN




                                                                          Bottom
Parameters for options 1, 2, 9 and 20 or command
===>
F1=Help   F3=Exit     F4=Prompt    F5=Refresh   F9=Retrieve
F11=Order by status   F12=Cancel   F13=Work with cluster menu
```

Figure 13-102 Metro Mirror failover preferred production node status

After bringing back the preferred production node, he steps to fail back are the same as in our first case above.

If you are not using the advanced node failure detection feature, the backup node is unable to know the real production node status, and it puts the production node in the *partition* status within the recovery domain. We describe this behavior in the third case, below.

- To simulate a cluster *partition* condition, we break the heartbeat IP connection by ending the heartbeat IP interface on the primary node.

  If the issue is either a real production node issue, or, for example, a network issue that prevents any user from connecting to the production node, you might decide to perform a failover.

  The node status for the primary cluster node needs to be changed from the partition to failed with the following command:

  ```
  CHGCLUNODE CLUSTER(PWRHA_CLU) NODE(DEMOPROD) OPTION(*CHGSTS)
  ```

The command handles only these failover events:

– It changes the cluster node status from partition to failed and the cluster resource group to inactive, and it changes the cluster resource group node role (Figure 13-103).

```
Work with Recovery Domain

Cluster resource group . . . . . :   PWRHA_CRG1
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add node   4=Remove node   5=Display more details   20=Dump trace

                                    Current        Preferred      Site
Opt     Node         Status         Node Role      Node Role      Name

        DEMOHA       Active         *PRIMARY       *BACKUP  1     SITE2
        DEMOPROD     Inactive       *BACKUP  1     *PRIMARY       SITE1

                                                                        Bottom
Parameters for options 1 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 13-103 Metro Mirror failback: Partition status changed to inactive

– It performs the failover actions on the DS8000 (Figure 13-104).

```
dscli> lspprc -dev IBM.2107-75AY032 6000-61FF
Date/Time: September 30, 2011 2:34:39 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
ID        State     Reason      Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================================
6000:6000 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6001:6001 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6002:6002 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6100:6100 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid
6101:6101 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid
6102:6102 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid
```

Figure 13-104 Metro Mirror failback: Preferred backup volumes status

The remaining actions, such as varying on the independent ASP and starting the takeover IP address, must be done manually.

After recovery of the preferred primary node, the steps to fail back are the same as for our first case above.

When heartbeat communication is no longer possible, the backup node sets both the production cluster node (Figure 13-105) and the cluster resource group (Figure 13-106 on page 361) status to *partition*, and sends the following informational message to the cluster message queue (Figure 13-107 on page 361):

CPFBB4F "Automatic fail over not started for cluster resource group PWRHA_CRG1 in cluster PWRHA_CLU. "

```
                              Work with Cluster Nodes

Local node . . . . . . . . . . . . . . . . :   DEMOHA
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add   2=Change   4=Remove   5=Display more details   6=Work with monitors
  8=Start   9=End    20=Dump trace

Opt     Node         Status         Device Domain

        DEMOFC       Active
        DEMOHA       Active          PWRHA_DMN
        DEMOPROD     Partition       PWRHA_DMN




                                                                          Bottom
Parameters for options 1, 2, 9 and 20 or command
===>
F1=Help   F3=Exit     F4=Prompt    F5=Refresh   F9=Retrieve
F11=Order by status   F12=Cancel   F13=Work with cluster menu
```

Figure 13-105 Metro Mirror failback cluster node partition status

```
                              Work with Recovery Domain

Cluster resource group . . . . . :   PWRHA_CRG1
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add node   4=Remove node   5=Display more details   20=Dump trace

                                    Current        Preferred      Site
Opt     Node         Status         Node Role      Node Role      Name

        DEMOHA       Active         *BACKUP  1     *BACKUP  1     SITE2
        DEMOPROD     Partition      *PRIMARY       *PRIMARY       SITE1









                                                                        Bottom
Parameters for options 1 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 13-106 Metro Mirror failback CRG partition status

```
                           Additional Message Information

Message ID . . . . . :   CPFBB4F       Severity . . . . . :   40
Message type . . . . . :   Information
Date sent  . . . . . :   09/30/11      Time sent  . . . . . :   13:46:09

Message . . . . :   Automatic fail over not started for cluster resource group
  PWRHA_CRG1 in cluster PWRHA_CLU.
Cause . . . . . :   A failure has occurred on primary node DEMOPROD or with
  the job associated with cluster resource group PWRHA_CRG1 on node DEMOPROD.
  Cluster Resource Services is unable to determine the state of the resources
  being managed by cluster resource group PWRHA_CRG1 and is unable to perform
  automatic fail over. The type of failure is reason code 1. Possible reason
  codes are:
    1 -- The cluster has become partitioned because the failure appears to be
  a communication failure.
    2 -- Either there is no backup node defined in the cluster resource group
  or all backup nodes are not active.  If this is a resilient device cluster
                                                                        More...
Press Enter to continue.

F3=Exit   F6=Print   F9=Display message details   F12=Cancel
F21=Select assistance level
```

Figure 13-107 Metro Mirror failback Heartbeat lost informational message

If the issue is really only about heartbeat communication and does not impact production activity, and if it can be fixed quickly, after it is fixed, after a 15-minute interval (the default value), the cluster status will automatically change back to active.

DS8000 LUN-level switching failover and back

A system failure or other major outage at the production node might require an unplanned switch of an IASP. This is handled in the same way as a planned switch in 13.2.2, “Using CL commands for DS8000 LUN-level switching” on page 346. However, when the IASP is varied on, there are added delay factors due to the same abnormal IPL considerations for rebuilding database access paths that are encountered during a system IPL. Consider using systems-managed access-path protection (SMAPP) and setting it to the shortest rebuild time possible. For more information, see in 15.5, “Best practices for reducing IASP vary on times” on page 409.

13.2.4 Detaching and reattaching a remote copy ASP session

If you want to get access to the IASP on your backup node, you have to *detach* the IASP session associated with that IASP first. Reasons to do this might be short-term testing on the backup node or major application changes on the production node that you do not want to propagate to the backup system before doing final testing on the production side. This scenario applies to Metro Mirror and both asymmetrical and symmetrical Global Mirror configurations.

Note: To get a consistent status of your application data with DS8000 Remote Copy Services, the IASP needs to be varied off before being detached.

Caution: While the ASP session is detached, it is not possible to perform a switch.

Before starting the operations, make sure that everything is correct for the cluster, cluster resource groups, and recovery domain, as described in previous sections:

1. Vary off the independent ASP (in our case IASP1) on the production node using the following command:

   ```
   VRYCFG CFGOBJ(IASP1) CFGTYPE(*DEV) STATUS(*OFF)
   ```

2. Detach the ASP session, using the following command:

   ```
   CHGASPSSN SSN(IASP1_MM) OPTION(*DETACH)
   ```

   Note: The DETACH option is required to run on the backup node for a Global Mirror configuration and on the production node for a Metro Mirror configuration.

The result of detaching an ASP session is that both source and target volumes are *suspended source* so that volumes are available for host I/O (Figure 13-108).

```
dscli> lspprc -dev IBM.2107-75AY031 6000-61FF
Date/Time: September 30, 2011 3:34:12 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
ID        State     Reason      Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
==========================================================================================================
6000:6000 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6001:6001 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6002:6002 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6100:6100 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid
6101:6101 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid
6102:6102 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid

dscli> lspprc -dev IBM.2107-75AY032 6000-61FF
Date/Time: September 30, 2011 3:34:17 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
ID        State     Reason      Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
==========================================================================================================
6000:6000 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6001:6001 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6002:6002 Suspended Host Source Metro Mirror 60        60             Disabled      Invalid
6100:6100 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid
6101:6101 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid
6102:6102 Suspended Host Source Metro Mirror 61        60             Disabled      Invalid
```

Figure 13-108 PPRC relationship status after detaching session

3. Vary on the independent ASP (in our case IASP1) on the production node using the following command:

   ```
   VRYCFG CFGOBJ(IASP1) CFGTYPE(*DEV) STATUS(*ON)
   ```

4. After the detach of the ASP session you can also vary on the independent ASP on the backup node. As expected and shown on Figure 13-109, the DEMOHA node is in *ineligible* status, meaning that it cannot be used for switchovers or failovers.

```
                          Work with Recovery Domain

Cluster resource group . . . . . . . . . . :   PWRHA_CRG1
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add node   4=Remove node   5=Display more details   20=Dump trace

                                    Current        Preferred      Site
Opt     Node         Status         Node Role      Node Role      Name

        DEMOHA       Ineligible     *BACKUP  1     *BACKUP  1     SITE2
        DEMOPROD     Active         *PRIMARY       *PRIMARY       SITE1









                                                                        Bottom
Parameters for options 1 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 13-109 Backup mode ineligible

5. When you are ready to re-establish replication in the original direction from the preferred primary to the preferred backup node, the next step is to vary off the independent ASP on the backup node and re-attach the ASP session on the backup node using the following command:

```
CHGASPSSN SSN(IASP1_MM) OPTION(*REATTACH)
```

Proper PPRC relationships are restored from the production node to the backup node (Figure 13-110), and the recovery domain is active again (Figure 13-111).

```
dscli> lspprc -dev IBM.2107-75AY031 6000-61FF
Date/Time: September 30, 2011 4:00:02 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
ID        State       Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================================
6000:6000 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6001:6001 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6002:6002 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6100:6100 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6101:6101 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6102:6102 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
dscli> lspprc -dev IBM.2107-75AY032 6000-61FF
Date/Time: September 30, 2011 4:00:05 PM CDT IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
ID        State              Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================================
6000:6000 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6001:6001 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6002:6002 Target Full Duplex -      Metro Mirror 60        unknown        Disabled      Invalid
6100:6100 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
6101:6101 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
6102:6102 Target Full Duplex -      Metro Mirror 61        unknown        Disabled      Invalid
```

Figure 13-110 PPRC relationships after re-attach

```
                          Work with Recovery Domain

Cluster resource group . . . . . :   PWRHA_CRG1
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add node   4=Remove node   5=Display more details   20=Dump trace

                                    Current        Preferred      Site
Opt     Node          Status        Node Role      Node Role      Name

        DEMOHA        Active        *BACKUP  1     *BACKUP  1     SITE2
        DEMOPROD      Active        *PRIMARY       *PRIMARY       SITE1









                                                                          Bottom
Parameters for options 1 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 13-111 Recovery Domain active back

In the example above, the REATTACH option is run on the backup node, which means that production-independent ASP data will copied to the backup independent ASP. This direction is still consistent with the cluster resource group status.

Switching while detached

For specific cases, you might need to copy current backup independent ASP data to the current and preferred production independent ASP and perform a role swap, leading the currently backup node to become the production node. To perform such actions, run the following steps:

1. Make sure that independent ASPs are varied off on both sides.
2. Change the role of each cluster resource group node with **CHGCRG** (Example 13-26). The node that was previously the primary one becomes the backup one, and the node that was previously the backup one becomes the production one. This command requires the cluster resource group to be inactive.

Example 13-26 Changing cluster resource group node role

```
CHGCRG CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG1) CRGTYPE(*DEV) RCYDMNACN(*CHGCUR)
   RCYDMN(  (DEMOHA *PRIMARY *SAME *SAME *SAME *SAME)
            (DEMOPROD *BACKUP 1 *SAME *SAME *SAME))
```

3. You can now run the REATTACH option in the proper way, which is from the current production site to the current backup site (that is, on the current backup node) (Example 13-27) with **CHGASPSSN**.

Note: Do not use **CHGASPSSN** for asymmetrical Global Mirror. Instead use the DS8000 **failbackpprc** command. Refer to specific operations for asymmetrical Global Mirror.

Example 13-27 Reattach ASP session

```
CHGASPSSN SSN(IASP1_MM) OPTION(*REATTACH)
```

13.2.5 Managing FlashCopy

In this section we describe the management of DS8000 FlashCopy with IBM PowerHA SystemMirror for i including FlashCopy reverse, incremental FlashCopy, and detaching and reattaching a FlashCopy ASP session.

Reversing FlashCopy

Note: To make the FlashCopy reversible you need to start the session with the options FLASHTYPE(*COPY) and PERSISTENT(*YES).

Reverse operation of the FlashCopy causes the source volume data to be overwritten by target volume data. This operation might be helpful if you want to remove the changes done to the source IASP since the FlashCopy was started. In this example we use the same environment as in “Environment overview” on page 295. The only changes we made was removing system names in the ASP copy description, so the IASP target copy will not be accessed by any other system. To do this we used the following command:

```
CHGASPCPYD ASPCPY(ASPCPYD2) LOCATION(*NONE)
```

In addition, we used target disks not included in any volume group on the D8000.

Now we need to start the FlashCopy session in the normal way:

1. Vary off the IASP or quiesce the database activity for the IASP on the production system.
2. Start the FlashCopy session with the following command:

   ```
   STRASPSSN SSN(SESFC1) TYPE(*FLASHCOPY) ASPCPY((ASPCPYD1 ASPCPYD2))
   FLASHTYPE(*COPY) PERSISTENT(*YES)
   ```

   Figure 13-112 show the status of the session.

```
                                Display ASP Session                     DEMOHA
                                                          10/05/11  18:16:12
Session . . . . . . . . . . . . . . . . . . :   SESFC1
  Type  . . . . . . . . . . . . . . . . . . :     *FLASHCOPY
  Persistent  . . . . . . . . . . . . . . . :     *NO
  FlashCopy type  . . . . . . . . . . . . . :     *COPY
  Number sectors copied . . . . . . . . . . :     3450496
  Number sectors remaining to be copied . . :     63658368


                              Copy Descriptions

ASP
device         Name                 Role          State     Node
IASP3          ASPCPYD1           SOURCE      AVAILABLE     DEMOHA
IASP3          ASPCPYD2           TARGET        UNKNOWN     *NONE



                                                                      Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel   F19=Automatic refresh
```

Figure 13-112 ASP session for FlashCopy reverse

3. Reverse the FlashCopy session.

   When you want to remove the changes done to the source IASP3 and bring it to the state that it was in when the FlashCopy was started, take the following steps:

   a. Vary off the IASP3.

   b. On the source system use the following command:

      ```
      CHGASPSSN SSN(SESFC1) OPTION(*REVERSE).
      ```

   c. Vary on the IASP3.

   After this procedure is done you should see in the IASP3 the data that was there when the FlashCopy was started. When the FlashCopy session finishes copying the data in the reverse direction you might want to end the FlashCopy or reverse it back so that you have another point-in-time copy for the recovery point.

> **Note:** It is impossible to reverse a FlashCopy when its source is in a Metro Mirror or a Global Mirror relationship.

Figure 13-113 shows FlashCopy after the reverse.

```
                          Display ASP Session                    DEMOHA
                                                          10/05/11  22:18:05
Session . . . . . . . . . . . . . . . . . . :   SESFC1
  Type  . . . . . . . . . . . . . . . . . . :     *FLASHCOPY
  Persistent  . . . . . . . . . . . . . . . :     *YES
  FlashCopy type  . . . . . . . . . . . . . :     *COPY
  Number sectors copied . . . . . . . . . . :     67108864
  Number sectors remaining to be copied . . :     0


                          Copy Descriptions

ASP
device         Name                 Role          State     Node
IASP3          ASPCPYD2           SOURCE        UNKNOWN     *NONE
IASP3          ASPCPYD1           TARGET      AVAILABLE     DEMOHA



                                                                   Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel   F19=Automatic refresh
```

Figure 13-113 FlashCopy after reverse

Incremental FlashCopy

When you have a FlashCopy target that is used for a longer period of time (for example, as a source for your data warehouse), you might want to have the FlashCopy session active all the time and do the incremental updates of the target IASP. To do this you need to configure the ASP copy descriptions pointing to your source and target volumes and allowing the IASPs to be used by the source and target nodes. The ASP copy descriptions that we use for this example is the same as in “Environment overview” on page 295. To do a incremental FlashCopy we need to take the following steps:

1. Quiesce the source IASP with the following command:

   ```
   CHGASPACT ASPDEV(IASP3) OPTION(*SUSPEND) SSPTIMO(30)
   ```

2. Start the FlashCopy session with the PERSIST(*YES) option, as we do here:

   ```
   STRASPSSN SSN(SESFC1) TYPE(*FLASHCOPY) ASPCPY((ASPCPYD1 ASPCPYD2))
   FLASHTYPE(*NOCOPY) PERSISTENT(*YES)
   ```

3. Resume the source IASP operation with the following command:

   ```
   CHGASPACT ASPDEV(IASP3) OPTION(*RESUME)
   ```

4. Vary on the target IASP on the target node (DEMOFC) and use it.

Any time that you want to update the content of the target IASP take the following steps:

1. Quiesce the source IASP with the following command:

   ```
   CHGASPACT ASPDEV(IASP3) OPTION(*SUSPEND) SSPTIMO(30)
   ```

2. Do the FlashCopy increment operation with the following command:

   ```
   CHGASPSSN SSN(SESFC1) OPTION(*INCR)
   ```

3. Resume the source IASP operation with the following command:

   ```
   CHGASPACT ASPDEV(IASP3) OPTION(*RESUME).
   ```

4. Vary on the target IASP on the target node (DEMOFC) and use it.

Detaching and reattaching FlashCopy

To detach the FlashCopy session you need to run following commands on the target system:

```
VRYCFG CFGOBJ(IASP3) CFGTYPE(*DEV) STATUS(*OFF)
CHGASPSSN SSN(SESFC1) OPTION(*DETACH) TRACK(*YES)
```

Specifying TRACK(*YES) enables the DS8000 to track the changes that are done to the source volumes, and allows the reattach operation to be done faster. When the session is detached, the target volumes are not usable. In this state you can do the system maintenance of the target. In the DS8000, the FlashCopy session is not present.

When you want to bring the session back, you need to issue the following command:

```
CHGASPSSN SSN(SESFC1) OPTION(*REATTACH)
```

14

Configuring and managing CSVC/V7000 Copy Services

In this chapter we provide you with step-by-step instructions for setting up a SVC/V7000 environment using PowerHA SystemMirror together with Metro Mirror, Global Mirror, and FlashCopy.

In addition, each section contains recommendations for daily operation in the respective environment.

14.1 SVC/V7000 Copy Services

In this section, we provide you with step-by-step setup instructions for an SVC/V7000 environment using PowerHA System Mirror for i together with Metro Mirror, Global Mirror, and FlashCopy. In addition, we provide recommendations for the daily operation in the respective environment.

Environment overview

For this scenario, we need to create several cluster items. Cluster, cluster monitors, IASP, and administrative domain setup are covered in Chapter 11, “Creating a PowerHA base environment” on page 199. Creating a PoweHA base environment is common to all the scenarios.

These are the specific cluster items for this scenario:

- Cluster resource group
  - Device domain
- ASP copy descriptions and sessions

Table 14-1 and Figure 14-1 on page 373 describe and show our environment.

Table 14-1 Settings for the SVC/V7000 scenarios

	Preferred primary	**Preferred backup**	**FlashCopy**
System name	CTCIHA9V	CTCIHA9W	CTCIHA9X
Cluster name	PWRHA_CLU		
Cluster resource group	SVC_MM_CRG / SVC_GM_CRG		
Device domain	PWRHA_DMN		
Administrative domain	PWRHA_CAD		
IASP name, number	IASP1, 144 / IASP2, 145	IASP1, 144 / IASP2, 145	
Cluster resource group site name	SITE1	SITE2	
Takeover IP	10.0.0.1		
Heartbeat cluster IP	10.10.10.1	10.10.10.2	10.10.10.3
Management access IP	9.5.167.53	9.5.167.54	9.5.167.60
SVC IP address	9.5.168.218	9.5.168.220	9.5.168.218 or 9.5.168.220
ssh key file location	/QIBM/UserData/HASM/hads/.ssh/id_rsa		
Volumes IDs[a]	15-18 / 19-22	0-3 / 4-7	23-26 or 8-11
SCSI adapter resource name	DC07 / DC09	DC08 / DC09	DC07 or DC01

a. Volumes IDs do not need to be the same on the source and target of remote copy relationships.

Figure 14-1 shows a graphical overview of the setup used in the SVC scenarios.

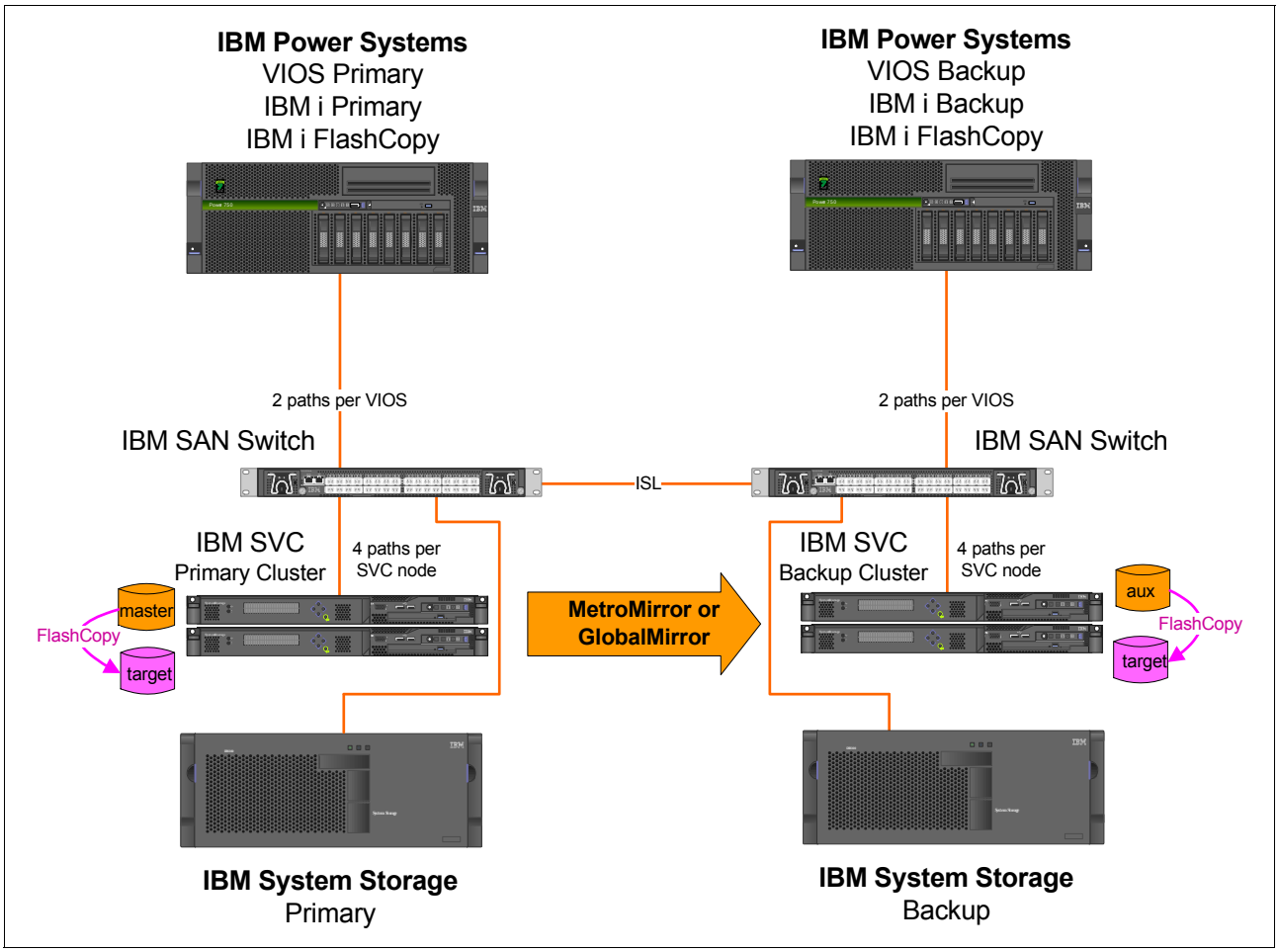


Figure 14-1 Overview of SVC setup for scenarios

14.1.1 Setting up an IBM i SVC/V7000 Copy Services environment

In this section we discuss Setting up an IBM i SVC/V7000 Copy Services environment.

Preparing for SSH connection between IBM i and SVC/V7000

Communication between PowerHA and SVC or V7000 is done using SSH. Therefore, you must create SSH key pairs and attach the SSH public key to a user on the SVC or V7000. The corresponding private key file to be used is specified in the creation of the ASP copy descriptions. It has to be distributed to all nodes in the cluster.

Generation of the SSH key pair is done on IBM i from QSHELL. Example 14-1 shows the command to do this and a list of the files created.

Example 14-1 Generation of ssh keys on IBM i

```
> cd /QIBM/UserData/HASM/hads/.ssh/
   $
> ssh-keygen -t rsa -f id_rsa -N ''
  Generating public/private rsa key pair.
  Your identification has been saved in id_rsa.
  Your public key has been saved in id_rsa.pub.
```

```
  The key fingerprint is:
  76:49:e1:9e:04:0a:c5:e2:68:3a:d0:6b:0b:4b:e2:2e powerha@CTCIHA9V.rchland.ibm.com
  $
> ls -la
  total 64
  drwxrwsrwx    2 powerha  0             8192 Sep 21 17:02 .
  drwx--Sr-x    3 qsys     0             8192 Sep 21 15:19 ..
  -rw-------    1 powerha  0             1679 Sep 21 17:02 id_rsa
  -rw-r--r--    1 powerha  0              414 Sep 21 17:02 id_rsa.pub
  $
```

Note: The ssh key pair generated here and used by PowerHA is in OpenSSH key format. It cannot be used by PuTTY, as PuTTY expects SSH2 keys.

You then have to import the `id_rsa.pub` file as a key into a user on your SVCs. This user must have administrator role to perform the functions used by PowerHA. Transfer the file to your PC and import it to the SVC user (Figure 14-2).

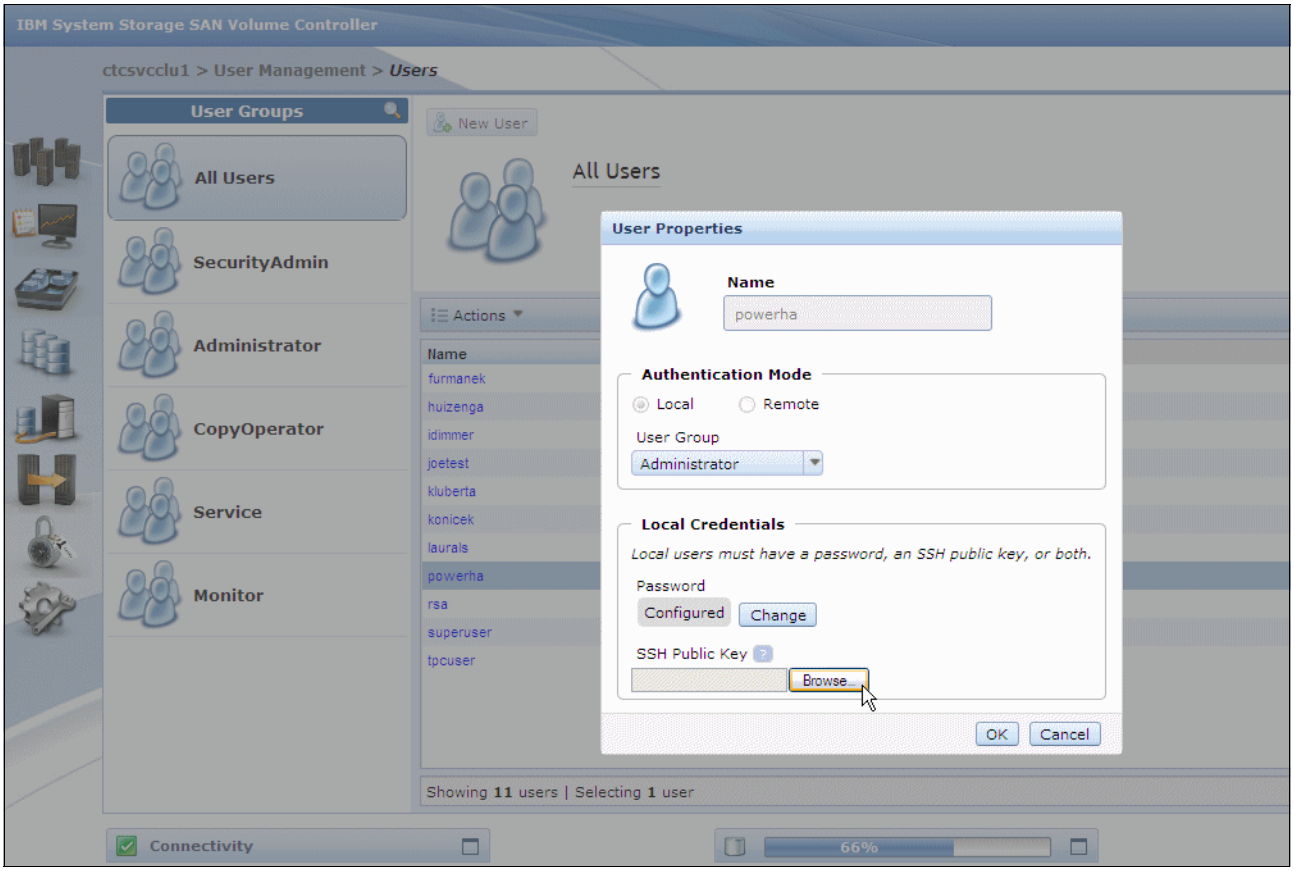


Figure 14-2 Import SSH public key to SVC user

Make sure to distribute the `id_rsa` file to all nodes in your cluster to the same directory.

Initializing IBM i disk units on the backup nodes

Prior to setting up Copy Services on SVC/V7000, you have to initialize and format the read-protected DPHxxx disk units to become usable for the IASP on the IBM i *backup* nodes. This can be done in SST by choosing option 3 (Working with disk units), then selecting option

3 (Work with disk unit recovery), then selecting option 2 (Disk unit problem recovery procedure), and finally selecting option 1 (Initialize and format disk unit). Failing to do so can result in IASP disk units not showing up properly after a switchover/failover to the secondary system.

14.1.2 Configuring IBM i SVC/V7000 remote Copy Services

This section describes the configuration of IBM PowerHA SystemMirror for i IASP replication using SVC/V7000 Metro Mirror or Global Mirror.

Setting up SVC/V7000 remote Copy Services

Take the following steps to set up SVC/V7000 remote Copy Services:

1. To set up a remote Copy Services partnership between two SVC clusters, both clusters need to be zoned in the SAN switch configuration such that they can see each other. This can be verified using **`svcinfo lsclustercandidate`**, as shown for each SVC cluster in Example 14-2.

Example 14-2 svcinfo lsclustercandidate

```
IBM_2145:ctcsvcclu1:admin>svcinfo lsclustercandidate
id               configured name
0000020065FFFFFE no         ctcsvcclu2

IBM_2145:ctcsvcclu2:admin>svcinfo lsclustercandidate
id               configured name
0000020065005ADE no         ctcsvcclu1
```

2. After the SVC inter-cluster zoning has been configured, we create SVC cluster partnerships between the two clusters using **`svctask mkpartnership`** on each cluster (Example 14-3). Note that when the partnership has been configured only on one cluster it is in a partially_configured state.

Example 14-3 svctask mkpartnership

```
IBM_2145:ctcsvcclu1:admin>svctask mkpartnership -bandwidth 200 ctcsvcclu2

IBM_2145:ctcsvcclu2:admin>svctask mkpartnership -bandwidth 200 ctcsvcclu1

IBM_2145:ctcsvcclu1:admin>svcinfo lscluster
id               name       location partnership      bandwidth id_alias
0000020065005ADE ctcsvcclu1 local
0000020065005ADE
0000020065FFFFFE ctcsvcclu2 remote   fully_configured 200
0000020065FFFFFE
```

3. Before creating the remote copy volume relationships we create a remote copy consistency group on each SVC cluster, which is required by PowerHA (Example 14-4).

Example 14-4 svctask mkrcconsistgrp

```
IBM_2145:ctcsvcclu1:admin>svctask mkrcconsistgrp -cluster ctcsvcclu2 -name IASP1_MM
RC Consistency Group, id [0], successfully created

IBM_2145:ctcsvcclu1:admin>svcinfo lsrcconsistgrp
id name     master_cluster_id master_cluster_name aux_cluster_id   aux_cluster_name primary state relationship_count copy_type
0  IASP1_MM 0000020065005ADE  ctcsvcclu1          0000020065FFFFFE ctcsvcclu2              empty 0                 empty_group
```

4. After creating the remote copy consistency group we create the remote copy volume relationships for our IBM i IASP volumes:
   - CTCIHA9V_MM_V*x* for the IBM i production node
   - CTCIHA9W_MM_V*x* for the backup node (Example 14-5)

Note: `svctask mkrcrelationship` creates Metro Mirror remote copy relationships by default. If Global Mirror relationships are desired, add the -global parameter.

Example 14-5 Creating SVC Metro Mirror relationships

```
IBM_2145:ctcsvcclu1:admin>svcinfo lsvdisk -filtervalue name=CTCIHA9V_MM*
id name       IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name capacity type   FC_id FC_name RC_id RC_name vdisk_UID
fc_map_count copy_count fast_write_state se_copy_count
15 CTCIHA9V_MM_V0 0           io_grp0       online 1            PowerHA        20.00GB  striped
600507680194016B7800000000000012 0            1          empty            0
16 CTCIHA9V_MM_V1 0           io_grp0       online 1            PowerHA        20.00GB  striped
600507680194016B7800000000000013 0            1          empty            0
17 CTCIHA9V_MM_V2 0           io_grp0       online 1            PowerHA        20.00GB  striped
600507680194016B7800000000000014 0            1          empty            0
18 CTCIHA9V_MM_V3 0           io_grp0       online 1            PowerHA        20.00GB  striped
600507680194016B7800000000000015 0            1          empty            0

IBM_2145:ctcsvcclu2:admin>svcinfo lsvdisk -filtervalue name=CTCIHA9W_MM*
id name       IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name capacity type   FC_id FC_name RC_id RC_name vdisk_UID
fc_map_count copy_count fast_write_state se_copy_count
0  CTCIHA9W_MM_V0 0           io_grp0       online 0            PowerHA        20.00GB  striped
600507680197FFFFF800000000000000 0            1          empty            0
1  CTCIHA9W_MM_V1 0           io_grp0       online 0            PowerHA        20.00GB  striped
600507680197FFFFF800000000000001 0            1          empty            0
2  CTCIHA9W_MM_V2 0           io_grp0       online 0            PowerHA        20.00GB  striped
600507680197FFFFF800000000000002 0            1          empty            0
3  CTCIHA9W_MM_V3 0           io_grp0       online 0            PowerHA        20.00GB  striped
600507680197FFFFF800000000000003 0            1          empty            0

IBM_2145:ctcsvcclu1:admin>svctask mkrcrelationship -master CTCIHA9V_MM_V0 -aux CTCIHA9W_MM_V0 -cluster ctcsvcclu2 -consistgrp IASP1_MM
-name CTCIHA9_MM_V0
RC Relationship, id [15], successfully created
IBM_2145:ctcsvcclu1:admin>svctask mkrcrelationship -master CTCIHA9V_MM_V1 -aux CTCIHA9W_MM_V1 -cluster ctcsvcclu2 -consistgrp IASP1_MM
-name CTCIHA9_MM_V1
RC Relationship, id [16], successfully created
IBM_2145:ctcsvcclu1:admin>svctask mkrcrelationship -master CTCIHA9V_MM_V2 -aux CTCIHA9W_MM_V2 -cluster ctcsvcclu2 -consistgrp IASP1_MM
-name CTCIHA9_MM_V2
RC Relationship, id [17], successfully created
IBM_2145:ctcsvcclu1:admin>svctask mkrcrelationship -master CTCIHA9V_MM_V3 -aux CTCIHA9W_MM_V3 -cluster ctcsvcclu2 -consistgrp IASP1_MM
-name CTCIHA9_MM_V3
RC Relationship, id [18], successfully created

IBM_2145:ctcsvcclu1:admin>svcinfo lsrcrelationship -filtervalue name=CTCIHA9_MM*
id name          master_cluster_id master_cluster_name master_vdisk_id master_vdisk_name aux_cluster_id   aux_cluster_name aux_vdisk_id
aux_vdisk_name primary consistency_group_id consistency_group_name state               bg_copy_priority progress copy_type
15 CTCIHA9_MM_V0 0000020065005ADE  ctcsvcclu1          15              CTCIHA9V_MM_V0    0000020065FFFFFE ctcsvcclu2       0
CTCIHA9W_MM_V0 master  0                   IASP1_MM              inconsistent_stopped 50               0        metro
16 CTCIHA9_MM_V1 0000020065005ADE  ctcsvcclu1          16              CTCIHA9V_MM_V1    0000020065FFFFFE ctcsvcclu2       1
CTCIHA9W_MM_V1 master  0                   IASP1_MM              inconsistent_stopped 50               0        metro
17 CTCIHA9_MM_V2 0000020065005ADE  ctcsvcclu1          17              CTCIHA9V_MM_V2    0000020065FFFFFE ctcsvcclu2       2
CTCIHA9W_MM_V2 master  0                   IASP1_MM              inconsistent_stopped 50               0        metro
18 CTCIHA9_MM_V3 0000020065005ADE  ctcsvcclu1          18              CTCIHA9V_MM_V3    0000020065FFFFFE ctcsvcclu2       3
CTCIHA9W_MM_V3 master  0                   IASP1_MM              inconsistent_stopped 50               0        metro
```

5. After the remote copy relationships have been created and added to the consistency group we start the consistency group using **`svctask startrcconsistgrp`** (Example 14-6). Note that after starting the consistency group its status changes from inconsistent_stopped to inconsistent_copying.

Example 14-6 Starting the SVC remote copy consistency group

```
IBM_2145:ctcsvcclu1:admin>svctask startrcconsistgrp IASP1_MM

IBM_2145:ctcsvcclu1:admin>svcinfo lsrcconsistgrp IASP1_MM
id 0
name IASP1_MM
```

```
master_cluster_id 0000020065005ADE
master_cluster_name ctcsvcclu1
aux_cluster_id 0000020065FFFFFE
aux_cluster_name ctcsvcclu2
primary master
state inconsistent_copying
relationship_count 4
freeze_time
status
sync
copy_type metro
RC_rel_id 15
RC_rel_name CTCIHA9_MM_V0
RC_rel_id 16
RC_rel_name CTCIHA9_MM_V1
RC_rel_id 17
RC_rel_name CTCIHA9_MM_V2
RC_rel_id 18
RC_rel_name CTCIHA9_MM_V3
```

6. We can now observe the progress of the SVC remote copy background synchronization process using **svcinfo lsrcrelationship** (Example 14-7).

Note: As long as the SVC remote copy background synchronization progress has not completed (that is, it has not reached consistent_synchronized), the remote copy relationships should not be switched, as the data on the remote site is still inconsistent.

Example 14-7 Viewing the SVC remote copy relationship status

```
IBM_2145:ctcsvcclu1:admin>svcinfo lsrcrelationship CTCIHA9_MM_V0
id 15
name CTCIHA9_MM_V0
master_cluster_id 0000020065005ADE
master_cluster_name ctcsvcclu1
master_vdisk_id 15
master_vdisk_name CTCIHA9V_MM_V0
aux_cluster_id 0000020065FFFFFE
aux_cluster_name ctcsvcclu2
aux_vdisk_id 0
aux_vdisk_name CTCIHA9W_MM_V0
primary master
consistency_group_id 0
consistency_group_name IASP1_MM
state inconsistent_copying
bg_copy_priority 50
progress 39
freeze_time
status online
sync
copy_type metro
```

Setting up PowerHA for SVC/V7000 remote Copy Services

Assuming that you have already created a cluster environment, an administrative domain (if needed), and an IASP on your production site, here are the steps required to set up the PowerHA configuration for SVC/V7000 Metro Mirror:

1. On the backup system, create a device description for the IASP with the same IASP name that you use on your production site:

   ```
   CRTDEVASP DEVD(IASP1) RSRCNAME(IASP1) RDB(*GEN)
   ```

   If you choose a different RDB name than the IASP name on the production site, make sure to specify the same value on the backup site.

2. Create the cluster resource group (CRG) using the command shown in Figure 14-3.

```
                    Create Cluster Resource Group (CRTCRG)

Type choices, press Enter.

Cluster  . . . . . . . . . . . . > PWRHA_CLU     Name
Cluster resource group . . . . . > SVC_MM_CRG    Name
Cluster resource group type  . . > *DEV          *DATA, *APP, *DEV, *PEER
CRG exit program . . . . . . . . > *NONE         Name, *NONE
  Library  . . . . . . . . . . .                 Name
User profile . . . . . . . . . . > *NONE          Name, *NONE

Recovery domain node list:
  Node identifier  . . . . . . . > CTCIHA9V      Name
  Node role  . . . . . . . . . . > *PRIMARY      *CRGTYPE, *PRIMARY...
  Backup sequence number . . . .   *LAST         1-127, *LAST
  Site name  . . . . . . . . . . > SITE1         Name, *NONE
  Data port IP address . . . . .   *NONE

  Node identifier  . . . . . . . > CTCIHA9W      Name
  Node role  . . . . . . . . . . > *BACKUP       *CRGTYPE, *PRIMARY...
  Backup sequence number . . . .   *LAST         1-127, *LAST
  Site name  . . . . . . . . . . > SITE2         Name, *NONE
  Data port IP address . . . . .   *NONE

Exit program format name . . . .   EXTP0100      EXTP0100, EXTP0101...
Exit program data  . . . . . . .   *NONE
Distribute info user queue . . .   *NONE         Name, *NONE
  Library  . . . . . . . . . . .                 Name
Job  . . . . . . . . . . . . . .   *JOBD          Name, *JOBD, *CRG

Configuration object list:
  Configuration object . . . . . > IASP1         Name, *NONE
  Configuration object type  . .   *DEVD        *DEVD, *CTLD, *LIND, *NWSD
  Configuration object online  . > *ONLINE       *OFFLINE, *ONLINE,
  Server takeover IP address . . > '10.0.0.1'

Text description . . . . . . . .   *BLANK
Failover message queue . . . . .   *NONE         Name, *NONE
  Library  . . . . . . . . . . .                 Name


F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 14-3 Create cluster resource group for SVC/V7000 Metro Mirror

3. Start the CRG using the following command:

```
STRCRG CLUSTER(PWRHA_CLU) CRG(SVC_MM_CRG)
```

4. Add the ASP copy descriptions using **ADDSVCCPYD** (Figure 14-4). This has to be done for the IASP configuration on the production site and for the IASP configuration on the backup site, so create two ASP copy descriptions.

```
                Add SVC ASP Copy Description (ADDSVCCPYD)

 Type choices, press Enter.

 ASP copy . . . . . . . . . . . . > SVC_MM_S      Name
 ASP device . . . . . . . . . . . > IASP1         Name
 Cluster resource group . . . . . > SVC_MM_CRG    Name, *NONE
 Cluster resource group site  . . > SITE1         Name, *NONE
 Node identifier  . . . . . . . . > *CRG          Name, *CRG, *NONE
 Storage host:
   User name  . . . . . . . . . . > admin
   Secure shell key file  . . . . > '/QIBM/UserData/HASM/hads/.ssh/id_rsa'

   Internet address . . . . . . . > '9.5.168.218'

 Virtual disk range:
   Range start  . . . . . . . . . > 15           0-8191
   Range end  . . . . . . . . . . > 18           0-8191
              + for more values


Bottom
 F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
 F24=More keys
```

Figure 14-4 Add SVC ASP Copy Description

Unless you are on SVC/V7000 6.2 or later, the user name that you specify here has to be admin. This profile has no relationship to the user that is used on the SVC. The actual user is chosen based on the ssh key file pair only. With SVC6.2 you can either specify admin as the user or the name of the SVC/V7000 user that the ssh keyfile actually belongs to. The virtual disk range is the SVC volume IDs of the disks in your IASP.

5. Start the ASP session using **STRSVCSSN** (Figure 14-5).

```
                      Start SVC Session (STRSVCSSN)

Type choices, press Enter.

Session  . . . . . . . . . . . . > SVC_MM        Name
Session type . . . . . . . . . . > *METROMIR     *METROMIR, *GLOBALMIR...
Cluster resource group . . . . . > SVC_MM_CRG    Name
Switchover reverse replication     *YES          *YES, *NO
Failover reverse replication . .   *YES          *YES, *NO

Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 14-5 Start SVC Session

For the session type you must specify whether the underlying remote copy function used in the SVC setup is Metro Mirror (specify session type as *METROMIR) or Global Mirror (specify session type as *GLOBALMIR).

The two parameters, Switchover reverse replication and Failover reverse replication, determine whether reverse replication in the SVC environment should be started automatically after a switchover or failover has occurred or whether the SVC session should stay in suspended mode.

Notice that you do not have to explicitly specify the ASP copy descriptions here. They are chosen by PowerHA from the CRG name provided in the session description and the information about IASP name and site name in the CRG.

Your IBM PowerHA SystemMirror i SVC/V7000 remote copy configuration is now done and your IASP is highly available for planned and unplanned site switches, as described in 14.2, “Managing IBM i SVC/V7000 Copy Services” on page 386.

14.1.3 Configuring IBM i SVC/V7000 FlashCopy

PowerHA SystemMirror for i allows you to create a FlashCopy of an IASP in a non-replicated IASP environment but also from a replicated IASP with Metro Mirror or Global Mirror. This allows a point-in-time copy to be taken of either the primary or the secondary site of a remote replication environment.

To create a FlashCopy point-in-time copy for an IASP in the IBM i cluster device domain, perform the following steps:

1. Create ASP copy descriptions for both the FlashCopy source and the target volumes.
2. Start an ASP session with type *FLASHCOPY, which links the copy descriptions and creates the FlashCopy relationships on the SVC/V7000 storage system.

Next we describe these steps for an example scenario of configuring FlashCopy on a Metro Mirror secondary site based on our SVC/V7000 environment (as shown in 14.1, “SVC/V7000 Copy Services” on page 372), taking the FlashCopy from the Metro Mirror secondary volumes.

Because we have not added our IBM i partition CTCIHA9X to the cluster and device domain, we must add it as the third node to the cluster and device domain first before being able to create an ASP copy description referencing it (Example 14-8). We use the IBM i partition as the FlashCopy partition accessing the FlashCopy target volumes for doing a backup to physical tape.

Example 14-8 Adding the FlashCopy node into the cluster and device domain

```
ADDCLUNODE CLUSTER(PWRHA_CLU) NODE(CTCIHA9X ('10.10.10.3'))
ADDDEVDMNE CLUSTER(PWRHA_CLU) DEVDMN(PWRHA_DMN) NODE(CTCIHA9X)
```

We also need to create a device description for the IASP on the FlashCopy target node (CTCIHA9X in our example) using **CRTDEVASP** (Example 14-9). Make sure to specify the same database name here that you specified when creating your original IASP.

Example 14-9 Creating the ASP device description on the FlashCopy target node

```
CRTDEVASP DEVD(IASP1) RSRCNAME(IASP1)
```

Creating ASP copy descriptions for FlashCopy

To create an ASP copy description for FlashCopy:

1. As we already have an ASP copy description for the Metro Mirror volumes on the secondary site, we do not need to create a new one for FlashCopy. However, we can use the existing ASP copy description from the Metro Mirror volumes to describe the FlashCopy source volumes. If you do not have an existing ASP copy description already from a remote copy relationship, use **ADDSVCCPYD** to add a corresponding SVC/V7000 copy description for the FlashCopy source volumes as shown in the following step.
2. We create an ASP copy description for the FlashCopy target volumes (Figure 14-6) by specifying information about the IASP being replicated and the details of the virtual disks (VDisks) within the SVC/V7000 that will be used as the FlashCopy target volumes for the IASP.

```
                  Add SVC ASP Copy Description (ADDSVCCPYD)

 Type choices, press Enter.

 ASP copy . . . . . .   SVC_FLC_T     Name
 ASP device . . . . . .   IASP1         Name
 Cluster resource group . . . . .   *NONE         Name, *NONE
 Cluster resource group site  . .   *NONE         Name, *NONE
 Node identifier  . . . . . .   CTCIHA9W      Name, *CRG, *NONE
 Storage host:
   User name  . . . . . .   admin
   Secure shell key file  . . . .   '/QIBM/UserData/HASM/hads/.ssh/id_rsa'

   Internet address . . . . . .   '9.5.168.220'

 Virtual disk range:
   Range start  . . . . . .   8             0-8191
   Range end  . . . . . .   11            0-8191
               + for more values


Bottom
 F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
 F24=More keys
```

Figure 14-6 Add SVC/V7000 ASP Copy Description for FlashCopy target volumes

Note: For the target copy description that is to be used for FlashCopy, the cluster resource group and cluster resource group site must be *NONE. The node identifier must be set to the cluster node name that will own the target copy of the independent ASP.

Creating an ASP session for FlashCopy

To create an ASP session for FlashCopy, follow these steps:

1. After having created the two ASP copy descriptions for the FlashCopy source and target volumes we are ready to create a FlashCopy for the IASP. However, before taking the FlashCopy by starting an ASP session for FlashCopy (Figure 14-7), the IASP either needs to be quiesced (using **CHGASPACT** to have as much modified data from main memory flushed to disk for reaching a consistent database state) or varied off.

```
                        Start SVC Session (STRSVCSSN)

Type choices, press Enter.

Session  . . . . . . . . . . . . > SVC_FLC       Name
Session type . . . . . . . . . . > *FLASHCOPY    *METROMIR, *GLOBALMIR...
ASP copy:
  Preferred source . . . . . . .   SVC_MM_T      Name
  Preferred target . . . . . . .   SVC_FLC_T     Name
              + for more values
Incremental flash  . . . . . . .   *NO           *NO, *YES
Copy rate  . . . . . . . . . . .   0             0-100
Cleaning rate  . . . . . . . . .   0             0-100
Grain size . . . . . . . . . . .   256           256, 64
Consistency group  . . . . . . .   *NEW

                                                                       Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 14-7 Starting an SVC/V7000 FlashCopy session

Note: IBM PowerHA SystemMirror for i automatically creates the FlashCopy mappings within the SVC/V7000 when starting an ASP session for type *FLASHCOPY. That is, the FlashCopy mappings between the VDisks must not be created on the SVC/V7000 by a user before **STRSVCSSN** is executed.

STRSVCSSN allows the user to also create an incremental FlashCopy relationship and specify a FlashCopy background copy rate and grain size. PowerHA requires the FlashCopy relationships to be included in a consistency group that is by default newly created by PowerHA on the SVC/V7000 when starting a FlashCopy session. Alternatively, the user can specify the name for an existing FlashCopy consistency group. For further information about these parameters see 7.4, “FlashCopy” on page 124. Information about managing a FlashCopy environment with SVC/V7000 can be found in 14.2, “Managing IBM i SVC/V7000 Copy Services” on page 386.

Example 14-10 shows a CL script to be run from the FlashCopy target node for automating a FlashCopy backup, including quiescing the IASP on production node CTCIHA9V prior to starting the FlashCopy session, varying on the IASP on the FlashCopy target node CTCIHA9X for doing the backup to tape before varying off the IASP on the FlashCopy node, and removing the FlashCopy session again.

Example 14-10 CHGASPACT run from the FlashCopy target node for quiescing an IASP

```
PGM
RUNRMTCMD  CMD('CHGASPACT ASPDEV(IASP1) +
             OPTION(*SUSPEND) SSPTIMO(30)') +
             RMTLOCNAME(CTCIHA9V *IP) RMTUSER(POWERHA) +
             RMTPWD(REDBOOK)
STRSVCSSN  SSN(SVC_FLC) TYPE(*FLASHCOPY) +
             ASPCPY((SVC_MM_T SVC_FLC_T))
RUNRMTCMD  CMD('CHGASPACT ASPDEV(IASP1) +
             OPTION(*RESUME)') RMTLOCNAME(CTCIHA9V +
             *IP) RMTUSER(POWERHA) RMTPWD(REDBOOK)
VRYCFG     CFGOBJ(IASP1) CFGTYPE(*DEV) STATUS(*ON)
/* INSERT CALL OF YOUR BACKUP PROGRAMS HERE */
VRYCFG     CFGOBJ(IASP1) CFGTYPE(*DEV) STATUS(*OFF)
ENDSVCSSN  SSN(SVC_FLC)
ENDPGM
```

Notes: **`STRSVCSSN`** and **`ENDSVCSSN`** must be executed from the IBM i cluster node that owns the target copy of the IASP.

The *DETACH and *REATTACH operations are not supported for a SVC/V7000 FlashCopy session.

Displaying an ASP session for FlashCopy

Using the DSPSVCSSN SSN(SVC_FLC), we can display our newly created ASP session for FlashCopy (Figure 14-8). Note that the ASP status for the IASP FlashCopy source node is always UNKNOWN, as the FlashCopy target node cannot determine the ASP state for the FlashCopy source node. For the FlashCopy target node, the ASP status shows AVAILABLE as we varied on the IASP on our target node CTCIHA9X.

```
                          Display SVC Session                    CTCIHA9X
                                                              09/27/11  17:40:26
Session . . . . . . . . . . . . . . . . . . . :   SVC_FLC
Type  . . . . . . . . . . . . . . . . . . . . :   *FLASHCOPY

Incremental flash . . . . . . . . . . . . . . :   *NO
Copy rate . . . . . . . . . . . . . . . . . . :   0
Cleaning rate . . . . . . . . . . . . . . . . :   0
Grain size (KB) . . . . . . . . . . . . . . . :   256
Consistency group . . . . . . . . . . . . . . :   fccstgrp2
                                                                        More...
Storage cluster name  . . . . . . . . . . . . :   ctcsvcclu2
                                                                         Bottom
                              Copy Descriptions

ASP         ASP copy              ASP         Replication
device      name        Role      Status         state     Node
IASP1       SVC_MM_T    SOURCE    UNKNOWN     ACTIVE       CTCIHA9W
            SVC_FLC_T   TARGET    AVAILABLE               CTCIHA9X

                                                                        Bottom
                              Copy Descriptions

ASP                               Copy
device      Role      Node      progress  Storage state
IASP1       SOURCE    CTCIHA9W    0       Copying
            TARGET    CTCIHA9X
Press Enter to continue

F3=Exit   F5=Refresh   F11=View 2   F12=Cancel
```

Figure 14-8 Displaying an ASP session for FlashCopy using DSPSVCSSN

The PowerHA SystemMirror for i log file `/QIBM/UserData/HASM/hads/xsm.log` shows the actions performed by PowerHA for starting the ASP session including the actual SVC/V7000 CLI commands that it executed for creating the FlashCopy consistency group and mappings (Example 14-11).

Example 14-11 /QIBM/UserData/HASM/hads/xsm.log

```
09/27/2011 17:31:34 95OF29B47F165OO1 <INFO> : YaspPluginSession : start : Start of plugin session start for ASP session: SVC_FLC
09/27/2011 17:31:34 95OF29B47F165OO1 <INFO> : YaspSession : deleteIAspObjects : Start of delete IASP Objects (Perm: 0) : 144
09/27/2011 17:31:34 95OF29B47F165OO1 <INFO> : YaspSession : deleteIAspObjects : delete IASP objects succeeded
09/27/2011 17:31:34 95OF29B47F165OO1 <INFO> : YaspSession : iapEnlistReject : Start of IASP enlist reject: 5
144
09/27/2011 17:31:34 95OF29B47F165OO1 <INFO> : YaspSession : iapEnlistReject : iaspEnlistReject succeeded
09/27/2011 17:31:34 95OF29B4CA3B3OO1 <INFO> : YaspSVCSession : startFlashCopy : Start of FlashCopy start for session SVC_FLC
09/27/2011 17:31:35 95OF29B4CA3B3OO1 /QOpenSys/usr/bin/ssh -i "/QIBM/UserData/HASM/hads/.ssh/id_rsa" -o UserKnownHostsFile=/dev/null -o
StrictHostKeyChecking=no admin@9.5.168.22O "svctask mkfcconsistgrp"

FlashCopy Consistency Group, id [2], successfully created
```

```
09/27/2011 17:31:35 950F29B4CA3B3001 /QOpenSys/usr/bin/ssh -i "/QIBM/UserData/HASM/hads/.ssh/id_rsa" -o UserKnownHostsFile=/dev/null -o
StrictHostKeyChecking=no admin@9.5.168.220 "svcinfo lsfcconsistgrp -delim \"&\" -nohdr -filtervalue \"id=2\""

2&fccstgrp2&empty

09/27/2011 17:31:36 950F29B4CA3B3001 /QOpenSys/usr/bin/ssh -i "/QIBM/UserData/HASM/hads/.ssh/id_rsa" -o UserKnownHostsFile=/dev/null -o
StrictHostKeyChecking=no admin@9.5.168.220 "svctask mkfcmap -source 0 -target 8 -consistgrp fccstgrp2 -copyrate 0 -grainsize 256
-cleanrate 0"

FlashCopy Mapping, id [0], successfully created

09/27/2011 17:31:36 950F29B4CA3B3001 /QOpenSys/usr/bin/ssh -i "/QIBM/UserData/HASM/hads/.ssh/id_rsa" -o UserKnownHostsFile=/dev/null -o
StrictHostKeyChecking=no admin@9.5.168.220 "svctask mkfcmap -source 1 -target 9 -consistgrp fccstgrp2 -copyrate 0 -grainsize 256
-cleanrate 0"

FlashCopy Mapping, id [1], successfully created

09/27/2011 17:31:37 950F29B4CA3B3001 /QOpenSys/usr/bin/ssh -i "/QIBM/UserData/HASM/hads/.ssh/id_rsa" -o UserKnownHostsFile=/dev/null -o
StrictHostKeyChecking=no admin@9.5.168.220 "svctask mkfcmap -source 2 -target 10 -consistgrp fccstgrp2 -copyrate 0 -grainsize 256
-cleanrate 0"

FlashCopy Mapping, id [2], successfully created

09/27/2011 17:31:38 950F29B4CA3B3001 /QOpenSys/usr/bin/ssh -i "/QIBM/UserData/HASM/hads/.ssh/id_rsa" -o UserKnownHostsFile=/dev/null -o
StrictHostKeyChecking=no admin@9.5.168.220 "svctask mkfcmap -source 3 -target 11 -consistgrp fccstgrp2 -copyrate 0 -grainsize 256
-cleanrate 0"

FlashCopy Mapping, id [3], successfully created

09/27/2011 17:31:38 950F29B4CA3B3001 /QOpenSys/usr/bin/ssh -i "/QIBM/UserData/HASM/hads/.ssh/id_rsa" -o UserKnownHostsFile=/dev/null -o
StrictHostKeyChecking=no admin@9.5.168.220 "svctask startfcconsistgrp -prep fccstgrp2"

09/27/2011 17:31:39 950F29B4CA3B3001 <INFO> : YaspSVCSession : startFlashCopy : FlashCopy session SVC_FLC    started successfully
09/27/2011 17:31:39 950F29B4CA3B3001 <INFO> : yaspSVCActionPgm : doSessionAction : Session action completed with return code: 1
09/27/2011 17:31:39 950F29B47F165001 <INFO> : YaspSession : iapEnlistReject : Start of IASP enlist reject: 6
144
09/27/2011 17:31:39 950F29B47F165001 <INFO> : YaspSession : iapEnlistReject : iaspEnlistReject succeeded
09/27/2011 17:31:39 950F29B47F165001 <INFO> : YaspSession : waitForUnitsToEnlist : Start of wait for units to enlist for session: SVC_FLC
09/27/2011 17:31:47 950F29B47F165001 <INFO> : YaspSession : waitForUnitsToEnlist : Disk units for all ASPs in session: SVC_FLC    have
enlisted
09/27/2011 17:31:47 950F29B47F165001 <INFO> : YaspSession : resetMultipath : Start of reset multipath: 144
09/27/2011 17:31:47 950F29B47F165001 <INFO> : YaspSession : resetMultipath : Reset multipath succeeded
09/27/2011 17:31:47 950F29B47F165001 <INFO> : YaspPluginSession : start : Plugin session start for ASP Session: SVC_FLC    completed
successfully.
```

14.2 Managing IBM i SVC/V7000 Copy Services

In this section we discuss managing IBM i SVC/V7000 Copy Services.

14.2.1 Displaying and changing a remote copy ASP session

An existing SVC/V7000 remote copy ASP session can be displayed using **DSPSVCSSN**, as shown for our synchronized Metro Mirror IASP environment in Figure 14-9.

```
                           Display SVC Session                    CTCIHA9V
                                                                09/29/11  17:12:11
 Session . . . . . . . . . . . . . . . . . . . :   SVC_MM
 Type  . . . . . . . . . . . . . . . . . . . . :   *METROMIR

 Switchover reverse replication  . . . . . . . :   *YES
 Failover reverse replication  . . . . . . . . :   *NO
 Consistency group . . . . . . . . . . . . . . :   IASP1_MM
 Source storage cluster name . . . . . . . . . :   ctcsvcclu1
 Target storage cluster name . . . . . . . . . :   ctcsvcclu2
                                                                          More...
 CRG name  . . . . . . . . . . . . . . . . . . :   SVC_MM_CRG
 Source site name  . . . . . . . . . . . . . . :   SITE1
 Target site name  . . . . . . . . . . . . . . :   SITE2

                                                                           Bottom
                              Copy Descriptions

 ASP         ASP copy              ASP         Replication
 device      name        Role      Status         state     Node
 IASP1       SVC_MM_S    SOURCE    AVAILABLE   ACTIVE       CTCIHA9V
             SVC_MM_T    TARGET    UNKNOWN                  CTCIHA9W

                                                                           Bottom
ASP                               Copy
 device      Role      Node      progress  Storage state
 IASP1       SOURCE    CTCIHA9V    100     Consistent sync
             TARGET    CTCIHA9W

                                                                           Bottom
 Press Enter to continue

 F3=Exit   F5=Refresh   F11=View 2   F12=Cancel
```

Figure 14-9 Displaying an SVC/V7000 Metro Mirror ASP session

Using **CHGSVCSSN ... OPTION(*CHGATTR)**, only the switchover and failover reverse replication parameters can be changed.

14.2.2 Suspending a remote copy ASP session

A remote copy ASP session for Metro Mirror or Global Mirror can be suspended to pause replication, as for performing disruptive maintenance actions for the remote site storage system. The remote copy secondary volumes remain inaccessible for the host.

An SVC/V7000 remote copy ASP session can be suspended from the IBM i cluster node that owns the primary or secondary copy of the IASP, using **CHGSVCSSN ... OPTION(*SUSPEND)**, which puts the SVC/7000 remote copy relationships in consistent_stopped state.

To resume replication with synchronization of tracked changes from the primary volumes to the secondary volumes use **CHGSVCSSN ... OPTION(*RESUME)**.

The backup node becomes *ineligible* while the remote copy replication is suspended or resuming (Figure 14-10). That is, the CRG cannot be switched or failed over before the remote copy status becomes consistent_synchronized again.

```
                          Work with Recovery Domain

Cluster resource group . . . . . . . . . . :   SVC_MM_CRG
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add node   4=Remove node   5=Display more details   20=Dump trace

                                    Current        Preferred      Site
Opt     Node         Status         Node Role      Node Role      Name

        CTCIHA9V     Active         *PRIMARY       *PRIMARY       SITE1
        CTCIHA9W     Ineligible     *BACKUP  1     *BACKUP  1     SITE2



                                                                        Bottom
Parameters for options 1 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 14-10 Work with Recovery Domain panel after suspending an ASP session

A suspended SVC/V7000 Metro Mirror or Global Mirror ASP session (because data consistency is ensured by using SVC/V7000 remote copy consistency groups) can also be *detached*, which puts the remote copy relationship in *idling* state to allow access of the secondary copy of the IASP from the IBM i backup node.

14.2.3 Detaching and reattaching a remote copy ASP session

If you want to get access to the IASP on your backup node, you have to *detach* the IASP session associated with that IASP first. Reasons to do this might be short-term testing on the backup node or major application changes on the production node that you do not want to propagate to the backup system before doing final testing on the production side. To get a consistent status of your application data, issue **CHGASPACT** on your production site first (Example 14-12).

Example 14-12 CHGASPACT for quiescing an IASP

```
CHGASPACT ASPDEV(IASP1) OPTION(*SUSPEND) SSPTIMO(30)
```

You can then detach the session from your backup node using the command shown in Example 14-13.

Example 14-13 CHGSVCSSN for detaching an IASP

```
CHGSVCSSN SSN(SVC_GM) OPTION(*DETACH)
```

After the detach is finished, the SVC/V7000 remote copy relationships are in *idling* status and you can resume access to your IASP on the production system using **CHGASPACT OPTION(*RESUME)**. The IASP can be varied on on the backup node and can be used there. By default, changes are tracked on the source side and on the target side so that a full resynchronization is not required after a *reattach*.

The reattach again has to be done from the backup node. Any changes made to the secondary copy IASP while detached will be overwritten by the data from the primary ASP copy. Vary off the IASP on the backup node and issue the command shown in Example 14-14.

Example 14-14 CHGSVCSSN for reattaching an IASP

```
CHGSVCSSN SSN(SVC_GM) OPTION(*REATTACH) SNDINQMSG(*YES)
```

By default, a message in the QSYSOPR message queue tells you which node will be acting as the primary node after the reattach and waits for confirmation before starting remote copy services on the SVC/V7000 again. If you want to avoid this behavior, you can specify the SNDINQMSG(*NO) parameter.

14.2.4 Planned switchover

Planned switchovers are usually done for maintenance purposes on the production system. For example, you might need to install PTFs that require an IPL on your production system, and you want to keep your users working during this timeframe.These are the steps for a planned switchover:

1. Using **DSPSVCSSN**, make sure that remote copy services on your SVC/V7000 environment work correctly and that data is in sync between production and backup. Figure 14-11 gives an example of an ASP session for SVC Metro Mirror replication, where Metro Mirror is in a consistent status. Notice that the replication status is reported as *active*, meaning that the target IASP is current with the source IASP.

```
                          Display SVC Session                    CTCIHA9V

Session . . . . . . . . . . . . . . . . . . . :   SVC_MM
Type  . . . . . . . . . . . . . . . . . . . . :   *METROMIR

Switchover reverse replication  . . . . . . . :   *YES
Failover reverse replication  . . . . . . . . :   *YES
Consistency group . . . . . . . . . . . . . . :   IASP1_MM
Source storage cluster name . . . . . . . . . :   ctcsvcclu1
Target storage cluster name . . . . . . . . . :   ctcsvcclu2
CRG name  . . . . . . . . . . . . . . . . . . :   SVC_MM_CRG
Source site name  . . . . . . . . . . . . . . :   SITE1
Target site name  . . . . . . . . . . . . . . :   SITE2


                             Copy Descriptions

ASP         ASP copy              ASP         Replication
device      name        Role      Status         state     Node
IASP1       SVC_MM_S    SOURCE    AVAILABLE   ACTIVE       CTCIHA9V
            SVC_MM_T    TARGET    UNKNOWN                  CTCIHA9W

ASP                               Copy
device      Role      Node      progress  Storage state
IASP1       SOURCE    CTCIHA9V    100     Consistent sync
            TARGET    CTCIHA9W
```

Figure 14-11 Displaying an ASP session for SVC Metro Mirror replication

2. End all applications using the IASP on the production site.
3. Use **CHGCRGPRI** to switch from the current production node to the next node in the recovery domain.

CHGCRGPRI performs the following tasks:

- It ends the switchable IP interface on the production site if one is defined.
- It varies off the IASP on the production site.
- In the cluster, it promotes the first backup node to primary node and moves the old primary node to the last backup position.

- If switchover reverse replication is configured with the default value of *YES in the ASP session, then the remote copy direction is switched on the SVC/V7000.

  If switchover reverse replication is configured as *NO in the ASP session, then the remote copy is detached on the SVC/V7000. When you want to restart the replication, you have to issue **CHGSVCSSN <*session name*> OPTION(*REATTACH)** from the current backup system.
- It varies on the IASP on the backup site if the configuration object information in the CRG specifies *ONLINE for the IASP.
- It starts the switchable IP interface on the backup site if one is defined.

14.2.5 Unplanned failover

For an unplanned failover we can distinguish between the following failure scenarios from the cluster resource service perspective:

- Primary node failure triggering an automatic failover event
- Primary node failure without an automatic failover event
- Primary node partition status

Each scenario requires different failover/recovery actions, which we describe in further detail in the following sections.

Primary node failure triggering an automatic failover event

An unplanned automatic failover event is triggered for a switchable IASP in a cluster resource group for a primary node failure event detected by cluster resource services either by a panic message sent by the failing primary node, as for an action of ending a cluster node or powering down the system, or by a power state change event sent by the HMC CIMOM server for a partition failure to the registered IBM i cluster nodes when using advanced node failure detection, as described in 11.2, “Setting up cluster monitors” on page 208.

For an unplanned automatic failover event, a CPABB02 inquiry message is sent to the cluster or CRG message queue on the *backup* node if a failover message queue is defined for either the cluster or the CRG. If no failover message queue is defined, then the failover starts immediately without any message being posted.

With the default cluster parameters FLVDFTACT=*PROCEED and FLVWAITTIM=*NOWAIT, an automatic failover is triggered. Setting a *failover wait time* with the FLVWAITTIM cluster parameter, either specified with a duration in minutes or *NOMAX, allows the user to respond to the CPABB02 inquiry message to either proceed with or cancel the failover. The *failover default action* FLVDFTACT parameter setting determines IBM PowerHA SystemMirror for its behavior to either automatically proceed or cancel the failover processing after the specified failover wait time expired without getting a response for the inquiry message from the user. Note that in any case, the primary IASP is taken offline by IBM PowerHA SystemMirror for i for a failover event regardless of these cluster failover parameter settings.

When using the default ASP session setting parameter FLVRVSREPL=*NO for an SVC/V7000 remote copy ASP session, for a failover event the session is only detached (Figure 14-12) for our ASP session for SVC Metro Mirror after an unplanned failover event. Also, the remote copy relationships are not reversed, which allows the user to preserve the primary node IASP data for possible further failure analysis.

```
                            Display SVC Session                    CTCIHA9W
                                                                09/28/11  18:23:47
 Session . . . . . . . . . . . . . . . . . . . :   SVC_MM
 Type  . . . . . . . . . . . . . . . . . . . . :   *METROMIR

 Switchover reverse replication  . . . . . . . :   *YES
 Failover reverse replication  . . . . . . . . :   *NO
 Consistency group . . . . . . . . . . . . . . :   IASP1_MM
 Source storage cluster name . . . . . . . . . :   ctcsvcclu2
 Target storage cluster name . . . . . . . . . :   ctcsvcclu1
                                                                        More...
 CRG name  . . . . . . . . . . . . . . . . . . :   SVC_MM_CRG
 Source site name  . . . . . . . . . . . . . . :   SITE2
 Target site name  . . . . . . . . . . . . . . :   SITE1

                                                                       Bottom
                               Copy Descriptions

 ASP         ASP copy              ASP         Replication
 device      name        Role      Status         state     Node
 IASP1       SVC_MM_T    SOURCE    AVAILABLE   DETACHED     CTCIHA9W
             SVC_MM_S    TARGET    UNKNOWN                  CTCIHA9V

ASP                               Copy
 device      Role      Node      progress  Storage state
 IASP1       SOURCE    CTCIHA9W    100     Idling
             TARGET    CTCIHA9V


                                                                       Bottom
 Press Enter to continue

 F3=Exit   F5=Refresh   F11=View 2   F12=Cancel
```

Figure 14-12 Displaying an ASP session for SVC Metro Mirror after an unplanned failover event

After recovery of the preferred primary node, including restart of its cluster services, the detached ASP session should be re-attached to resume remote copy replication and make the IASP highly available again. There are options for re-attaching the ASP session:

- To resume remote copy data replication from the secondary site back to the primary use **CHGSVCSSN ... OPTION(*REATTACH)** on the preferred primary site node, which is the current backup node. A planned switchback to the preferred primary node can then be done using **CHGCRGPRI**.
- Alternatively, if instead the data updates performed on the secondary site while the primary site node was not available will be discarded, the direction of remote copy data replication can be changed from the preferred primary node to the preferred backup node by ending the CRG, changing the recovery domain primary/backup node roles via **CHGCRG** before reattaching the session via **CHGSVCSSN ... OPTION(*REATTACH)** to resume replication between the preferred primary and preferred backup node site.

Note: The re-attach operation always needs to be run on the node that is supposed to become the target for the remote copy replication.

Primary node failure without an automatic failover event

A primary node failure without an automatic failover event being triggered could be caused, for example, by an IASP or SYSBAS storage access loss.

In this case the user should carefully consider whether it is more appropriate to try to recover from the access loss condition before invoking a failover to the backup cluster node, which requires the following manual actions. If the primary node is still responsive, the ASP session can be *detached* as follows to stop the remote copy relationships by allowing access to the secondary volumes so that IASP can be varied on at the backup node:

```
CHGASPSSN SSN(<ASP_session>) OPTION(*DETACH)
```

The *DETACH operation for a failed IASP implies that the node roles still need to be changed via **ENDCRG** and **CHGCUR RCYDMNACT(*CHGCUR)**.

Otherwise, if the primary node is no longer responsive, a cluster partition condition and failover can be enforced from the backup node as follows:

```
CHGCLUNODE CLUSTER(<cluster_name>) NODE(<primary_node_name>) OPTION(*CHGSTS)
```

Note: If **CHGCLUNODE** first fails with message CPFBB89, retry the command a short time later after the cluster changed to a partition condition.

A **CHGCLUNODE** triggers a cluster failover without a failover inquiry message and still requires the user to vary on the IASP on the new primary node and start the takeover IP interface.

After recovery of the preferred primary node, including a restart of its cluster services and of the device CRG, the ASP session should be reattached on the preferred primary node, which is the current backup node to resume replication. A switchback to the preferred primary node can then be done using **CHGCRGPRI**.

Primary node partition status

A primary node partition status can be triggered by a sudden primary node failure or cluster communication error. Using advanced node failure detection, as described in 11.2, “Setting up cluster monitors” on page 208, can help to avoid cluster partition conditions for many cases of sudden node failures, which the HMC or flexible service processor (FSP) are able to detect by sending corresponding CIM server power state events to the registered IBM i cluster nodes to invoke an automatic failover. However, there are still cases like cluster heartbeat communication errors or sudden system power outages that lead to a cluster node partition condition. The user should determine the reason for the partition condition of a cluster node to decide whether to fix it or declare the node in partition status as failed to invoke a cluster failover to a backup node.

A cluster partition condition with a failed cluster node indicated by message ID CPFBB20 requires manual actions to invoke a cluster failover, for example, by setting the primary cluster node from *partition* to *failed* status using the following CL command:

```
CHGCLUNODE CLUSTER(<cluster_name>) NODE(<primary_node_name>) OPTION(*CHGSTS)
```

After this **CHGCLUNODE** command, cluster node roles are changed and remote copy relationships are stopped so that the IASP can be manually varied on at the new primary node on the backup site.

After recovery of the preferred primary node, including restart of its cluster services and of the device CRG, a switchback to the preferred primary node can be done using **CHGCRGPRI**.

14.2.6 Displaying and changing a FlashCopy ASP session

An existing SVC/V7000 FlashCopy ASP session can be displayed using **DSPSVCSSN**, as shown in Figure 14-13 for our FlashCopy no-copy environment after varying on the IASP on the FlashCopy target node CTCIHA9X.

```
                           Display SVC Session                    CTCIHA9X
                                                                09/29/11  18:27:08
 Session . . . . . :   SVC_FLC
 Type  . . . . . :   *FLASHCOPY

 Incremental flash . . . . . :   *NO
 Copy rate . . . . . :   0
 Cleaning rate . . . . . :   0
 Grain size (KB) . . . . . :   256
 Consistency group . . . . . :   fccstgrp1
                                                                        More...
 Storage cluster name  . . . . . :   ctcsvcclu2

                                                                        Bottom
                             Copy Descriptions

 ASP         ASP copy              ASP         Replication
 device      name        Role      Status         state     Node
 IASP1       SVC_MM_T    SOURCE    UNKNOWN     ACTIVE       CTCIHA9W
             SVC_FLC_T   TARGET    AVAILABLE               CTCIHA9X

                                                                        Bottom
ASP                              Copy
 device      Role      Node      progress  Storage state
 IASP1       SOURCE    CTCIHA9W    0       Copying
             TARGET    CTCIHA9X

                                                                        Bottom
 Press Enter to continue

 F3=Exit   F5=Refresh   F11=View 2   F12=Cancel
```

Figure 14-13 Displaying an SVC/V7000 FlashCopy ASP session

When using FlashCopy with the background copy (that is, copy rate > 0), the background copy progress is reported in the *copy progress* information with a *storage state* of copying when the background copy has not finished yet or *copied* when the background copy has completed, as also indicated by a copy progress of 100.

Using **CHGSVCSSN ... OPTION(*CHGATTR)**, only the copy rate and cleaning rate parameters of a FlashCopy ASP session can be changed. This operation can be executed from either the IBM i cluster node that owns the source copy or the target copy of the IASP.

14.2.7 Reversing a FlashCopy ASP session

FlashCopy reverse for SVC/7000 is currently not supported by PowerHA SystemMirror for i and is planned to be released with a future PTF.

14.2.8 Using incremental FlashCopy

Incremental FlashCopy copies data from the source to the target volumes that have been modified since the initial creation of the FlashCopy or the last time an increment operation was performed.

Using incremental FlashCopy requires that the ASP session for the *initial* FlashCopy is created with the incremental flash option set to *YES using **`STRSVCSSN ... TYPE(*FLASHCOPY) INCR(*YES)`**. A background copy rate greater than 0 should be specified when starting the ASP session for the initial FlashCopy because any subsequent incremental FlashCopy operations can only be performed using **`CHGSVCSSN`** with the *INCR option after the initial background copy has finished.

As with any other FlashCopy operation, the FlashCopy incremental operation must also be executed from the IBM i cluster node that owns the target copy of the independent IASP.

Figure 14-14 shows an existing FlashCopy incremental session displayed using **DSPSVCSSN**.

```
                          Display SVC Session                    CTCIHA9X
                                                          10/03/11  12:02:16
Session . . . . . . . . . . . . . . . . . . :   SVC_FLC
Type  . . . . . . . . . . . . . . . . . . . :   *FLASHCOPY

Incremental flash . . . . . . . . . . . . . :   *YES
Copy rate . . . . . . . . . . . . . . . . . :   100
Cleaning rate . . . . . . . . . . . . . . . :   0
Grain size (KB) . . . . . . . . . . . . . . :   256
Consistency group . . . . . . . . . . . . . :   fccstgrp0
                                                                    More...
Storage cluster name  . . . . . . . . . . . :   ctcsvcclu2

                                                                    Bottom
                             Copy Descriptions

ASP         ASP copy              ASP         Replication
device      name        Role      Status         state     Node
IASP1       SVC_MM_T    SOURCE    UNKNOWN     ACTIVE       CTCIHA9W
            SVC_FLC_T   TARGET    AVAILABLE               CTCIHA9X

                                                                    Bottom
ASP                               Copy
device      Role      Node      progress  Storage state
IASP1       SOURCE    CTCIHA9W    100     Idle or copied
            TARGET    CTCIHA9X

                                                                    Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F11=View 1   F12=Cancel
```

Figure 14-14 Display SVC Session panel for an incremental FlashCopy session

If the background copy has not been enabled or finished, the incremental FlashCopy operation is rejected with the following CMMVC5907E message:

```
The FlashCopy mapping or consistency group was not started because the mapping or
consistency group is already in the copying state.
```

In this case the user should either wait for an existing background copy process to complete, which can be checked from the copy progress information displayed by **DSPSVCSSN**, or enable the background copy by using **CHGSVCSSN ... OPTION(*CHGATTR) CPYRATE(...)**.

15

Best practices

In this chapter we describe best practices that can help you efficiently use IBM PowerHA SystemMirror for i.

For further details about PowerHA functions and issues, you might also want to consult the “IBM Software Knowledge Base” section related to high availability at the following URL:

http://www-912.ibm.com/s_dir/slkbase.NSF/wHighAv?OpenView&view=wHighAv

15.1 Clustering configuration

In this section we discuss clustering configuration.

15.1.1 PowerHA license consideration

The 5770HAS Licensed Program Product is not tier based, but processor based. This means that on each node member of a cluster, you need a license key related to the number of processors using the product on every partition of the same server.

To avoid product commands from becoming unusable in case of processor activation related to temporary Capacity On Demand feature usage, make sure to apply PTF SI41735.

15.1.2 Requirements

When looking at the cluster properties with the GUI, there is an option to check the requirements. Make sure to use this option and apply all of them. As shown in Figure 11-10 on page 207, these are the most common requirements to be applied:

- System value QRETSVRSEC must be set to 1. This value is necessary for the administrative domain to retrieve user profiles passwords.

  Note: Considerations apply regarding this system value, which are described in the knowledge base document number 518525704 Admin Domain: What to Consider For User Profiles When Implementing, at the following URL:

  `http://www-912.ibm.com/s_dir/slkbase.NSF/cd034e7d72c8b65c862570130058d69e/92eb712dc067c1558625757b006a2a55?OpenDocument`

- The QGPL/QBATCH job queue entry for QSYS/QBATCH must have *NOMAX or a value greater than 1. This value is necessary because most of the cluster jobs are submitted to this job queue, and they must not be prevented from running by user jobs.
- QIBM_PWRDWNSYS_CONFIRM and QIBM_ENDSYS_CONFIRM environment variables should be set to *YES.

15.1.3 Independent ASP

For a device recovery configuration, independent ASPs are the main building block for implementing a hardware-based replication solution with PowerHA. Therefore, most of the best practices that you can use for the independent ASP can apply for clustering. See the overview in Chapter 2, “Implementing an independent auxiliary storage pool” on page 13, and *IBM i 6.1 Independent ASPs: A Guide to Quick Implementation of Independent ASPs*, SG24-7811, for more information.

Caution: Any IASP existing on a system before adding this system to a device domain must be deleted.

15.1.4 Communications

The *inetd* service must be started on all cluster nodes.

The network attribute Allow Add to Cluster (ALWADDCLU) must have the value *ANY.

The cluster makes use of several types of IP interfaces:

- Heartbeat interfaces

  These are used to ensure efficient communication between all cluster nodes. They do not need high bandwidth, as they exchange only a small amount of information.

  To avoid situations in which the users can no longer connect to the node, and at the same time nodes can no longer exchange heartbeat information (as described in “Metro Mirror and Global Mirror failover and failback” on page 350), you might want to use separate network devices for both user access and heartbeat access.

  Conversely, to protect heartbeat exchanges you can also activate redundancy by using virtual IP addressing or using the two available IP interfaces at the cluster definition level. Of course, physical links must use separate network devices.

- Data port interfaces

  These interfaces are used for data replication in case of geographic mirroring implementation. For bandwidth and redundancy purposes, you can configure up to four IP addresses. These IPs should use separate Ethernet adapters and network devices.

  We recommend that you do not mix data port traffic and that users access traffics to avoid performance degradation for the replication. Heartbeat interfaces and data port interfaces can share the same IP addresses.

- Communication with HMC or VIOS

When using advanced node failure detection, the HMC or VIOS should be able to send appropriate information to cluster nodes in case of partition failure. There is no specific configuration step for this item, but the name or IP address of the HMC or VIOS should be available. Therefore, this communication takes the IP routes available when needed. You might want to make sure that no network device failure could have an impact on this communication.

  If there is a firewall between the HMC or VIOS and the cluster nodes, make sure that the flow is authorized. Advanced node failure detection makes use of the CIMOM services, which listen on TCP 5989 (unencrypted communication) and 5990 (ssl encrypted communication) ports. Flow must be authorized in both ways.

- Communication with DS8000

  When using DS8000 replication, the cluster nodes need to establish IP communication with the DS8000 to check the replication status and start the appropriate actions, such as failover or failback. There is no specific configuration step for this item, but the name or IP address of the DS8000 should be available. Therefore, this communication will take the IP routes available when needed. You might want to make sure that no network devices failure can have an impact on this communication.

  If there is a firewall between the cluster nodes and the DS8000, make sure that the flow is authorized. DS8000 listens on the TCP 1750 port. The flow must be authorized from the nodes to the DS8000.

15.1.5 Failover wait time and default action

The failover wait time and default action settings determine the cluster behavior in case of a possible failover need detected by the backup node. Make sure that you understand how they work together to avoid unexpected events:

- Wait time defines the time to wait for a reply to the failover inquiry message sent to the cluster message queue:
  - THe *NOWAIT value means that failover proceeds immediately without any confirmation.
  - The *NOMAX value means that there is no limit for the cluster waiting on a reply for the message. If nobody answers, failover does not proceed.
  - “a number of minutes” means that the cluster will wait this number of minutes before proceeding as specified by the default action.
- The default action defines the action at the end of wait time if specified in minutes.
  - The *PROCEED value means to proceed with the failover.
  - The *CANCEL value means that failover occurs.

15.1.6 Administrative domain

One of the most common errors is related to a UID/GID user profile mismatch between the source and target nodes. As detailed in 15.5.2, “Synchronizing UIDs/GIDs” on page 409, it might have a huge impact when switching over or failing over. Using the cluster administrative domain is a good way to maintain consistency regarding this parameter. However, the administrative domain only takes care of resources that you have requested it to monitor. By itself, it is not able to take care of new resources. For example, when a new user profile is created, it is not added to the monitored resources list. A workaround is to use the QIBM_QSY_CRT_PROFILE exit point, which is activated when a user profile is created. You can write a CL program and register it against the exit point. This program should add the newly created user profile to the administrative domain with **ADDCADMRE**. Note that the same kind of considerations apply when you want to delete a user profile. It must be removed from the administrative domain monitored resources with RMVCADMRE before being deleted. The QIBM_QSY_DLT_PROFILE exit point can be used to automate the process. Care must be taken when deleting a user profile that owns objects. You need to decide what to do with the owned objects.

15.2 Journaling

It is difficult to achieve both Recovery Time Objective (RPO) and Recovery Point Objective (RTO) with an hardware replication solution such as IBM PowerHA SystemMirror for i without looking seriously to a journaling setup. This solution relies on disk replication. Therefore, it copies, from source disks to target disks, only data that exists on disks. Everything that is still in main memory and that has not been yet written to disk cannot be replicated. It is not a matter for *planned* switches, for which you will use varying off the IASP, which flushes memory content to disks. It becomes a matter for *unplanned* switches, which can occur at any time, for any reason, and for which we have to apply techniques to prevent loosing data and to reduce recovery time.

Let us see an example. Assume that an application makes use of a data area to keep track of the next customer number of a database file to record registration information for the new customer to be assigned and of an IFS file to capture a scanned image of the customer’s

signature. Three separate objects are used to store data. At the end of the transaction that enrolls this new customer, everything is consistent, for those three objects in *main memory*. But there is no assurance that the related *disk image* has received the updates. Most likely, they have not been received yet. It might happen that the disk update order will be different from the memory updates order, depending on main memory flushing algorithm.

This lack of consistency on disks drives clearly affects the ability to reach planned RPO. Some objects, like database files, might also need time-consuming internal consistency checks and recovery at IASP vary-on time. Detecting and correcting them affects planned RTO also.

Using journal is the way to go for achieving a consistent state and reducing recovery times as much as possible. Journal protection should be enabled for all critical recovery point objectives, data areas, database files, data queues, and IFS files. Using journal does not require an update to the application code. Starting, ending, and managing journal operations are to be done outside the application. The best protection for data consistency is achieved by changing the application to take care of transaction consistency with commitment control. The application decides when, from a functional standpoint, any transaction is finished and then it performs a commit. If an outage occurs during commitment control cycles, the system ensures, at IASP vary-on time, that data is consistent regarding the transactions. It undoes any update that is not committed.

Journaling has never been mandatory with the IBM i operating system. Therefore, numerous users have never tried to implement it, even when using hardware replication solutions. Note that software replication solutions are based on journaling, so considerations differ for them.

Note: It is a user responsibility to take care of journaling. IBM PowerHA SystemMirror for i has no option to help you with it.

In the past, there were two main obstacles to using journaling:

- Performance impact
- Management effort

15.2.1 Journal performance impact

Performance impact still exists and always will because the way that journaling works implies using more hardware resources, specifically increased disk writes operations. The objective of journaling is to write to disk, in the journal receiver object, on a synchronous manner, all updates to journaled objects before these updates become effective on disks.

However, in IBM i release after release, major improvements have been done on this subject:

- When using internal drives for IASP, write cache performance is a key factor for journal performance. For more information about this topic, refer to *Journaling - Configuring for your Fair Share of Write Cache*, TIPS0653, at:

  `http://publib-b.boulder.ibm.com/abstracts/tips0653.html`

- If you decide to use a private ASP for receivers, do not skimp on the quantity of disk arms available for use by the journal. For more information about this topic, refer to *Journaling - User ASPs Versus the System ASP*, TIPS0602, at:

  `http://publib-b.boulder.ibm.com/abstracts/tips0602.html`

  Note: Do not forget that a private (user) ASP can be created inside an IASP as a secondary ASP type and coexist with a primary type ASP.

- If you do not intend to use journal receiver entries for application purposes, or if you want to do so and are ready for API programming, consider minimizing receiver entries data. Journal parameters *receiver size options*, *minimize entry specific data*, and *fixed length data* play a role in this optimization step. For example, if you are using those journals only for recovery purposes, do you really need the entire database file record content in each record entry? For more information about this topic, refer to *Journaling: How to View and More Easily Audit Minimized Journal Entries on the IBM System i Platform*, TIPS0626, at:

  http://www.redbooks.ibm.com/abstracts/tips0626.html

- Depending on application needs, you cannot start journaling for work files that are used only to help data processing and do not contain any critical data.

- The number of disks arms used by a journal receiver is determined through the journal threshold parameter. Using as many disks as possible is a good point for performance improvement. For more information about this topic, refer to *Journaling - How Can It Contribute to Disk Usage Skew?*, TIPS0603, and *Journaling: Unraveling the mysteries of sporadic growth of Journal receivers*, TIPS0652, at:

  http://www.redbooks.ibm.com/abstracts/tips0603.html
  http://www.redbooks.ibm.com/abstracts/tips0652.html

  Note: PTF *MF51614* for IBM i 7.1 apply result is that *journal receivers are spread across all the disks* just like any other object type, via normal storage management technique. The journal threshold is no longer used to determine the number of disks required for journaling.

- Make sure that your journal is employing the most recent defaults. Many journals created prior to release IBM i 5.4 might still be locked into old settings, and these old settings might be thwarting performance. One of the easiest ways to ensure that you remain in lock-step with the best journal settings is to let the system apply its own latest defaults. You can help ensure that such settings are employed by specifying the RCVSIZOPT(*SYSDFT) parameter on **CHGJRN**.

- Consider installing and enabling the journal caching feature if journal performance (especially during batch jobs) is a major concern in your shop. Make sure to understand possible impacts on missing receiver writes in case of an outage. For more information about this topic, refer to *Journal Caching: Understanding the Risk of Data Loss*, TIPS0627, at:

  http://www.redbooks.ibm.com/abstracts/tips0627.html

- The more actively being-modified objects (such as open files) that you have associated with a journal, the higher you might want to set your journal recovery count. Failure to do so slows your runtime performance and increases your housekeeping overhead on your production system without adding much benefit to your RTO. Increasing this value might make good sense, but do not get carried away. For more information about this topic, refer to *The Journal Recovery Count: Making It Count on IBM i5/OS*, TIPS0625, at:

  http://www.redbooks.ibm.com/abstracts/tips0625.html

- If you have physical files that employ a force-write-ratio (the FRCRATIO setting) and those same files are also journaled, disable the force-write-ratio. Using both a force-write-ratio and journal protection for the same file yields no extra recovery/survivability benefit and only slows down your application. It is like paying for two health insurance policies when one will suffice. The journal protection approach is the more efficient choice.

- If your applications tend to produce transactions that consist of fewer than 10 database changes, you might want to give serious consideration to use of *soft* commitment control. Just make sure to understand that the last committed transaction can not be written to the journal. For more information about this topic, refer to *Soft Commit: Worth a Try on IBM i5/OS V5R4*, TIPS0623, at:

  http://www.redbooks.ibm.com/abstracts/tips0623.html

- Both before and after images of each updated record can be written to the journal receiver depending on the way that the database file journaling is started. For recovery purposes, we only need the after image. When the IASP is varied on, storage database recovery applies after images if needed. Consider journaling only if after images are an option.

> **Note:** Commitment control automatically activates before images if needed for a database file involved in a transaction and for which only after images are journaled.

More in-depth treatments of a number of these best practices can be found in *Striving for Optimal Journal Performance on DB2 Universal Database for iSeries*, SG24-6286, at:

http://www.redbooks.ibm.com/abstracts/sg246286.html

There are also numerous technotes that are probably more recent than the publication above, which can be found on the IBM Redbooks publication website.

You might want to try this query URL to find information about Journaling and IBM i:

http://publib-b.boulder.ibm.com/Redbooks.nsf/searchdomain?SearchView&query=[subjects]=AS400+and+Journaling&SearchOrder=1&category=systemi

15.2.2 Journal management effort

It is now much easier than in the past to manage the journal on a library-wide basis. New commands exist, and certain restrictions existing in the previous releases have disappeared.

Automatic journaling for changed objects within a (set of) library

STRJRNLIB allows the system to automatically start journaling for any object in this library that receives the operation create, move, restore, or all of these (Figure 15-1).

```
Start Journal Library (STRJRNLIB)

Type choices, press Enter.

Library  . . . . . .                 Name, generic*
             + for more values
Journal  . . . . . .                 Name
  Library  . . . . . .     *LIBL     Name, *LIBL, *CURLIB
Inherit rules:
  Object type  . . . . . .   *ALL     *ALL, *FILE, *DTAARA, *DTAQ
  Operation  . . . . . .     *ALLOPR  *ALLOPR, *CREATE, *MOVE...
  Rule action  . . . . . .   *INCLUDE *INCLUDE, *OMIT
  Images . . . . . .         *OBJDFT  *OBJDFT, *AFTER, *BOTH
  Omit journal entry . . . . .   *OBJDFT  *OBJDFT, *NONE, *OPNCLO
  Remote journal filter  . . . .   *OBJDFT  *OBJDFT, *NO, *YES
  Name filter  . . . . . .   *ALL     Name, generic*, *ALL
             + for more values
Logging level  . . . . . .   *ERRORS  *ERRORS, *ALL


                                                                  Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 15-1 STRJRNLIB

By using this command, all objects created in, moved to, and restored in the library will get *automatic journaling start* to the same journal, with the same settings. There is no longer a need to take care of them. However, you have to take care of those objects that you do not want to journal. End journaling must be run for them.

Note: This command does not take care of existing objects in the library. It only applies to changes to the library.

A new panel exists through the **DSPLIBD** command to display the current settings for automatic journaling of a library. Figure 15-2 shows an example with a journaled library MYLIB.

```
Display Library Description

Library  . . . . . :   MYLIB         Type . . . . . :   TEST

Journaling information:
  Currently journaled  . . . . . :   YES
  Current or last journal  . . . . . :   MYJOURNAL
    Library  . . . . . :     MYLIB
  Journal images . . . . . :   *AFTER
  Omit journal entry . . . . . :   *NONE
  New objects inherit journaling . . . :   *YES
  Inherit rules overridden . . . . . :   NO
  Journaling last started date/time  . :   09/20/11  17:00:25
  Starting receiver for apply  . . . . :
    Library  . . . . . :
    Library ASP device . . . . . :
    Library ASP group  . . . . . :




                                                                    Bottom
F3=Exit   F10=Display inherit rules   F12=Cancel   Enter=Continue
```

Figure 15-2 Journaling information for MYLIB library

Considerations apply with another method to automatically start journaling, introduced in i5/OS V5R4, with QDFTJRN data area. Form more information about these considerations and about **STRJRNLIB**, refer to *Journaling at Object Creation with i5/OS V6R1M0*, TIPS0662, at:

http://www.redbooks.ibm.com/abstracts/tips0662.html

Start journaling for all or generic files or objects for one or more library

STRJRNPF, **STRJRNAP**, and **STRJRNOBJ** have been enhanced to allow the start of journaling for all or generic files or objects at one time. In the previous release, we had to write a program to build a list of objects, read this list, and start journaling for each entry of this list. For example, all database files included in a library, or a set of libraries, can be journaled with one command (Figure 15-3).

```
Start Journal Physical File (STRJRNPF)

Type choices, press Enter.

Physical file to be journaled  .                Name, generic*, *ALL
  Library  . . . . . . . . . . .     *LIBL      Name, *LIBL, *CURLIB
             + for more values
                                     *LIBL
Journal  . . . . . . . . . . . .                Name
  Library  . . . . . . . . . . .     *LIBL      Name, *LIBL, *CURLIB
Record images  . . . . . . . . .   *AFTER       *AFTER, *BOTH
Journal entries to be omitted  .   *NONE        *NONE, *OPNCLO
Logging level  . . . . . . . . .   *ERRORS      *ERRORS, *ALL

                                                                    Bottom
F3=Exit    F4=Prompt    F5=Refresh    F12=Cancel    F13=How to use this display
F24=More keys
```

Figure 15-3 STRJRNPF command

Therefore, for an existing library with four commands, it is possible to start journaling for all existing applicable objects and for all changes that apply in the future. Example 15-1 shows an example with library name MYLIB.

Example 15-1 Automatic journaling starting for MYLIB library

```
STRJRNPF FILE(MYLIB/*ALL) JRN(MYLIB/MYJOURNAL) IMAGES(*AFTER) OMTJRNE(*OPNCLO)
STRJRNOBJ OBJ(MYLIB/*ALL) OBJTYPE(*DTAARA) JRN(MYLIB/MYJOURNAL) IMAGES(*AFTER)
STRJRNOBJ OBJ(MYLIB/*ALL) OBJTYPE(*DTAQ) JRN(MYLIB/MYJOURNAL) IMAGES(*AFTER)
STRJRNLIB LIB(MYLIB) JRN(MYLIB/MYJOURNAL) INHRULES((*ALL *ALLOPR *INCLUDE *AFTER
*OPNCLO))
```

End journaling locking restriction lifted

We can now end journaling of a physical file or an object, even if the object has been opened by an application. There is no longer a need to shut down the application to perform this action. It still remains impossible to do so for files in the midst of a commitment control cycle. With this update, **ENDJRNPF** and **ENDJRNOBJ** now also support generic names and all objects.

Receiver name wrap

In the past, for a user-created journal, when the receiver name got its maximum sequence (for example, JRNRCV9999), the system was unable to create a new one, and no more

entries could be written through the journal. All application programs were waiting for an answer to an inquiry message in the QSYSOPR message queue requesting that the user does something. Change Journal (**CHGJRN**) command processing has been enhanced to support receiver name wrapping. This means that if your currently attached receiver name is at the maximum value, the system will now generate a new receiver with the receiver number value being wrapped (for example, JRNRCV0000). And if (often due to poor housekeeping practices) a journal receiver with the name JRNRCV0000 lingers, the system simply increments the numerical suffix to skip over potential lingering receivers from days gone by and continues.

Pseudo journal tool

The pseudo journal tool is a standalone set of software. Its purpose is to assist in estimating the quantity of journal traffic that will ensue that journal protection is enabled for a set of designated physical files. It is useful to answer these questions:

- How many journals should I configure?
- Will the total quantity of journal/disk traffic justify use of more than one journal?
- Does it make sense for me to configure the journal caching feature on my production system and, if I do so, how much benefit am I likely to gain for my particular applications?

The nice thing about the pseudo journal tool is that it not only helps answer these questions, it does so without having a high impact on your system as it performs the analysis. Better yet, it produces a customized analysis of projected additional disk traffic, tuned to your particular application and its database reference pattern.

More information regarding the pseudo journal tool along with software to download and a tutorial can be found on the database tools website:

http://www-03.ibm.com/systems/i/software/db2/journalperfutilities.html

15.3 Best practices for planned site switches

In this section we discuss best practices for planned site switches.

15.3.1 Regular tests

A high-availability solution has no value if it is not tested on a regular basis and if regular production activities are not conducted on both production and backup sites.

At a minimum, we recommend a switchover every six months. This way, you can run half a year on one site, switch over to the other site, and start again six months later. A six month switch frequency is usually enough to make sure that the hardware is ready to handle production workload when needed. Also make sure that all people are trained to handle an unplanned failover. Running production on a six-month time window on the same node makes you comfortable with events that occur with an higher frequency than the week or the month.

15.3.2 Check cluster and replication health

Each time hat a planned switchover is required, cluster and replication health must be checked.

Nodes included in the planned switchover must be active, the administrative domain must be active, the related cluster resource group must be active, and recovery domain nodes must be active.

Note: Cluster resource groups with an inactive status are not handled by the failover procedure.

Normally, if there is an issue with the replication, the recovery domain node is *ineligible* instead of active, which prevents you from switching. Using **DSPASPSSN** or **DSPSVCSSN** gives you the opportunity to double-check the replication status, while using **DSPCRGINF** allows you to check the recovery domain node status.

Messages are sent to the cluster message queue, which can be monitored to fix related issues.

An the following message is an example of such a message:

```
HAI2001 "Cross-site mirroring (XSM) for ASP &1 is not active."
```

Fixing this event allows the related recovery node status to change from ineligible to *active*.

15.3.3 Reverse replication

When running a planned switch, make sure to understand that before switching the node roles PowerHA reverses the replication. This means that all the updates that you will do on the new production node will be replicated to the new backup node. For an SVC/V7000 remote Copy Services environment, the switchover reverse replication parameter (SWTRVSREPL) of the ASP session allows you to determine whether reverse replication is to be started after the switchover. For a geographic mirroring or DS8000 remote Copy Services environment, a switchover without starting replication in the reverse direction can be achieved by detaching the ASP session and manually changing cluster node roles in the CRG using **CHGCRG**.

15.3.4 Ending applications using the IASP before a switchover

When a switchover is performed all jobs using the independent ASP are ended abnormally during its vary-off. Therefore, it is important to properly end all applications using the IASP before switching over to help reduce application recovery and vary-on times.

Make sure that the interactive job running **CHGCRGPRI** is not using the IASP anymore because though the command succeeds, the interactive job gets disconnected otherwise.

15.4 Best practices for unplanned site switches

As for planned switchovers, all cluster resources and ASP sessions must be active. With IBM i 7.1, new CL commands were introduced to better support CL programming for cluster automation management by retrieving status information for diverse clustering objects that you might want to query on a regular basis. These commands are:

- **RTVCLU**
- **RTVCRG**
- **RTVASPCPYD**
- **RTVASPSSN**
- **PRTCADMRE**

Note: Cluster resource groups with inactive status are not handled by a failover procedure.

When using a CRG or cluster message queue, the cluster parameters *Failover wait time* and *Failover default action* can be used to either set up an automatic cluster failover or require a user confirmation for the failover to proceed or be cancelled. If both a CRG and a cluster message queue have been configured, the cluster message queue takes precedence over the corresponding CRG settings.

Note: In any case, clustering varies off the IASP on the current production node first when it detects a failover condition before issuing the failover message.

A cluster *partition* condition, as described in 15.7, “Resolving a cluster partition condition” on page 412, requires the user to decide whether the reason for the partition condition should be fixed or whether a cluster failover should be invoked by changing the cluster node from the partition to *failed* status.

In contrast to a planned switch, after an unplanned switch, by default replication is not automatically started from the new production to the new backup node to still allow for failure analysis of the original production node data.

15.5 Best practices for reducing IASP vary on times

In this section we describe the main tips that you can follow to be able to reduce the IASP vary-on times.

15.5.1 Keeping as few DB files in SYSBAS as possible

The system disk pool and basic user disk pools (SYSBAS) should primarily contain operating system objects, licensed program libraries, and few-to-none user libraries. This structure yields the best possible protection and performance. Application data is isolated from unrelated faults and can also be processed independently of other system activity. IASP vary-on and switchover times are optimized with this structure. Expect longer IASP vary-on and switchover times if you have a large number of database objects residing in SYSBAS because additional processing is required to merge database cross-reference information into the disk pool group cross-reference table.

15.5.2 Synchronizing UIDs/GIDs

In a high-availability environment, a user profile is considered to be the same across systems if the profile names are the same. The name is the unique identifier in the cluster. However, a user profile also contains a user identification number (UID) and group identification number (GID).

This UID and GID are used when looking at object ownership. To reduce the amount of internal processing that occurs during a switchover, where the IASP is made unavailable on one system and then made available on a different system, the UID and GID values should be synchronized across the recovery domain of the device cluster resource group. If this is not the case, then each object owned by a user profile with a non-matching UID needs to be accessed, and the UID needs to be changed as part of the vary-on process of the IASP after

each switchover or failover. Synchronization of user profiles including IUD and GID can be accomplished by using the administrative domain support.

15.5.3 Access path rebuild

To reduce the amount of time that an independent ASP vary-on waits for access path rebuilds, consider having the rebuilds performed in the background after the vary-on has completed. This is determined by the RECOVER attribute of a logical file. If RECOVER(*IPL) is specified, the vary-on will wait for the rebuild to complete when one is necessary. If RECOVER(*AFTIPL) is specified, the vary-on does not wait and the AP is built in the background after the vary-on is complete.

Access paths are not available while they are being rebuilt. Therefore, if there are specific logical files that an application requires to be valid shortly after a vary-on is complete, the user should consider specifying RECOVER(*IPL) to avoid a situation in which jobs will try to use the them before they are valid.

Another option to consider is the journaling of access paths so that they are recovered from the journal during a vary-on and do not need to be rebuilt. Access paths are journaled implicitly by SMAPP, as discussed in 15.5.4, “System Managed Access Path Protection” on page 410. To ensure that specific, critical access paths are journaled, they should be explicitly journaled.

When the system rebuilds an access path after IPL time, you can use **EDTRBDAP** to modify rebuild sequences to select those access paths that need to be rebuilt now. But, by default, this command applies only to SYSBASE. To allow it to work with independent ASP access paths, do the following:

1. Enter the following command, where *YY* is the IASP number:

   ```
   CRTDTAARA DTAARA(QTEMP/QDBIASPEDT) TYPE(*DEC) LEN(4) VALUE(YY)
   ```

2. Enter `EDTRBDAP` when the iASP becomes ACTIVE or AVAILABLE. It might be necessary to invoke the command multiple times while the iASP is being varied on and is ACTIVE. A CPF325C message will be sent until the command is allowed to be used.
3. Enter the following command to delete the data area:

   ```
   DLTDTAARA DTAARA(QTEMP/QDBIASPEDT)
   ```

15.5.4 System Managed Access Path Protection

Analyze your System Managed Access Path Protection (SMAPP) setting and ensure that you are not locked into an outdated setting inherited from the last decade. A SMAPP setting that is too high can significantly increase your recovery duration and thereby cause you to miss your RTO. SMAPP is a form of behind-the-scenes journaling. If you see a SMAPP setting larger than, for example, 50 minutes, give serious consideration to lowering the value. (An original default setting nearly a decade ago was 150 minutes, and many shops that have not revisited this setting as hardware speeds have improved might still be operating with outdated settings. Continuing to do so can make the vary-on duration for an IASP exceed your RTO.)

SMAPP applies to the IASP vary-on (or IPL for the SYSBASE) step, which is responsible for rebuilding access paths in case of damages after an outage. For access paths (or indexes in the SQL world) dependant on large physical files (or tables in the SQL world), it might take a considerable amount of time, which can be several hours. To avoid rebuilding the access paths, SMAPP uses existing journals for access path updates behind-the-scene recording,

just like if the access paths were journaled. They are recorded so that IASP vary-on step can use them to update access paths in place of rebuilding them.

Access paths update recording occurs automatically at each ASP level (independent or not, and including SYSBASE) depending on the following parameters:

- Access path recovery time target for each ASP.
- Access path recovery estimate for each ASP. Each access path for which the estimated rebuild time is higher than the target will be protected by SMAPP.

SMAPP effect the overall system performance. The lower the target recovery time that you specify for access paths, the greater this effect can be. Typically, the effect is not noticeable, unless the processor is nearing its capacity.

In the Figure 15-4, you can see an example of an existing independent ASP installation.

There is no target for any Independent ASP. Therefore, the system, by itself, has started access path journaling to achieve the overall system target, which is set to 50 minutes. Using the F14 key, we can see which access paths are currently protected (Figure 15-4).

For this example, the user should review the targets to specify a better recovery time target, for example, by specifying *MIN, which means the lower possible recovery time.

```
Display Recovery for Access Paths                SYSTEMA
                                                          05/10/11  21:15:58
Estimated system access path recovery time  . . . :            44   Minutes
Total not eligible recovery time  . . . . . . . . :             0   Minutes
Total disk storage used . . . . . . . . . . . . . :       909,073   MB
% of disk storage used  . . . . . . . . . . . . . :         0,035
System access path recovery time  . . . . . . . . :            50
Include access paths  . . . . . . . . . . . . . . :          *ALL

            ------Access Path Recovery Time------  -----Disk Storage Used-----
ASP               Target          Estimated             Megabytes     ASP %
1                  *NONE                  1                34,611     0,009
IASP1              *NONE                  9               265,965     0,022
IASP2              *NONE                 33               608,497     0,059

                                                                         Bottom
F3=Exit   F5=Refresh   F12=Cancel    F13=Display not eligible access paths
F14=Display protected access paths   F15=Display unprotected access paths
```

Figure 15-4 DSPRCYAP command result

```
Display Protected Access Paths                SYSTEMA
                                                              05/10/11  21:21:06
                                                 Estimated
                                                 Recovery
File           Library        ASP            If Not Protected
OSASTD10       M3EPRD         IASP1              00:03:03
OSASTD90       M3EPRD         IASP1              00:02:57
OSASTD00       M3EPRD         IASP1              00:02:55
QADBIFLD       QSYS00033      IASP1              00:02:50
MITTRA30       MVXCDTA800     IASP2              00:02:00
MITTRA35       MVXCDTA800     IASP2              00:01:55
OODOCU00       MVXCDTA800     IASP2              00:01:53
MITTRAZ9       MVXCDTA800     IASP2              00:01:50
MITTRA50       MVXCDTA800     IASP2              00:01:49
QADBIFLD       QSYS00034      IASP2              00:01:49
MITTRA20       MVXCDTA800     IASP2              00:01:48
MITTRA70       MVXCDTA800     IASP2              00:01:48
MITTRA60       MVXCDTA800     IASP2              00:01:45
MITTRAU6       MVXCDTA800     IASP2              00:01:41
MITTRA90       MVXCDTA800     IASP2              00:01:41
MITTRA80       MVXCDTA800     IASP2              00:01:40
                                                                        More...
F3=Exit   F5=Refresh   F12=Cancel   F17=Top   F18=Bottom
```

Figure 15-5 Protected access paths

15.6 Switching mirroring while detached

When mirroring is detached, you still have the opportunity to switch, which means changing the node role. The usual **CHGCRGPRI** command cannot be used at this time because the backup node has become ineligible.

This is the appropriate way to do a switch while detached:

1. Change cluster resource group node roles with **CHGCRG**.
2. Reattach the ASP session from the current backup node to re-establish remote mirroring from the current production to the backup node.

Refer to “Switching while detached” on page 366.

15.7 Resolving a cluster partition condition

A cluster partition condition occurs when nodes in a cluster can no longer communicate with each other (that is, they no longer receive heartbeat information). Therefore, clustering cannot distinguish between a communication failure or a real node failure. You cannot switch to a node while it is in partition status.

To avoid some of these partition conditions, the HMC managing the production node server or the VIOS managing the production node partition can send information about the status of

the partition through the *advanced node failure detection* mechanism. When it is really in a not running status, the backup node receives the information and automatically starts a failover procedure. Refer to 11.2, “Setting up cluster monitors” on page 208, for more information.

Make sure that you analyze the cause of any partition condition when deciding to either resolve it or start a manual cluster failover. It could be due to a temporary network issue, which can be easily corrected. As soon as the temporary failure is corrected, on a 15-minute cycle the partition status is automatically solved and the node status changes from partition back to *active*.

If you really need to start a failover use **CHGCLUNODE** to change the node status from partition to *failed*.

See “Metro Mirror and Global Mirror failover and failback” on page 350, for examples of these scenarios.

15.8 IBM i hosting environments

IBM i hosting environments can be used to deploy either IBM i, AIX, or Linux partitions, or they can be used to provide storage for IBM Blade servers or IBM System x® servers attached to IBM i using iSCSI. In all these cases, storage is provided to the hosted environment using network server storage spaces and a network server description. As these network server storage spaces from an IBM i perspective are simply objects located in the IFS, it is possible to move these network server storage spaces into an IASP and therefore use IBM PowerHA SystemMirror for i to provide high availability for these hosted environments. This is not recommended for large IBM i, AIX, or Linux partitions, but is an option for smaller environments.

There are, however, a few additional steps that need to be considered if using a hosted environment:

- Although the network server storage space can reside in an IASP, the network server description and network server configuration objects needed for iSCSI cannot. Therefore, make sure to use the administrative domain to keep the network server descriptions synchronized between all nodes in your cluster.
- For hosted IBM i, AIX, or Linux partitions, virtual SCSI adapters have to be defined on the production system as well as on the backup system to connect a network storage space to the hosted LPAR.
- For a planned switch, make sure to first end all activities in your hosted environment. Then power down your hosted environment and end the network server description before issuing **CHGCRGPRI**.
- After a planned switch or a failover, vary on the network server description on the backup system and start your hosted environment on the backup system.
- The following white paper provides sample exit programs that can be used to automate the vary-off and vary-on processes for network server descriptions:

  http://www-03.ibm.com/systems/resources/systems_i_os_aix_pdf_linux_aix_xsm_final.pdf

Perform a switchover test before moving the environment to production to make sure that the setup does work on the backup site (that is, that you have done the necessary configuration steps on the backup site correctly).

15.9 Upgrading IBM i and PowerHA release

If your business can afford to power down your system, then you can upgrade both systems to the new version of the operating system at the same time.

If you need to run continuous operations and are confident with the release upgrade, you can do a so-called *rolling upgrade*. During this process (described in 15.9.1, “Performing a rolling upgrade in a clustering environment” on page 414) either the production or the backup system is available to the user. Be aware though that there are periods of time when you do not have a backup system that you could fail over to if your current production system fails.

If you like to test your applications on one system using the new release while keeping the application environment intact on the other system (so that in case of problems you can simply move back to using this environment with the old version of the operating system), do the release upgrade on the backup system after detaching the IASP.

15.9.1 Performing a rolling upgrade in a clustering environment

When doing a rolling upgrade you first have to upgrade your backup system to the new version of the operating system. The reason for this is that you can switch an independent auxiliary storage pool (iASP) from a lower release to a higher release, but not from a higher release down to a lower release, so you cannot switch back to lower release node after the IASP has been varied on at the upgraded node.

The general order of steps when performing IBM i and PowerHA release upgrades is as follows:

1. If you are using geographic mirroring, you can suspend it with tracking.
2. Upgrade IBM i and PowerHA on the current backup node and restart the backup cluster node.
3. If you are using geographic mirroring and suspended it, resume it to synchronize the IASP from the current production to the current backup node.
4. Switch from production to the upgraded backup node.

   Independent ASP database conversion from one release to another occurs when the IASP is first varied on.
5. If you are using geographic mirroring, you can suspend it with tracking.
6. Upgrade IBM i and PowerHA on the current backup node and restart the backup cluster node.
7. If you are using geographic mirroring and suspended it, resume it to synchronize the IASP from the current production to the current backup node.
8. When all cluster nodes are on the same release, update the PowerHA version with **CHGCLUVER**.

   Note: **CHGCLUVER** might need to be used twice to update the PowerHA version if you skipped a release at the upgrade, for example, from V5R4 to i 7.1.
9. Optional: Switch back to your preferred production node.
10. You might want to review your cluster administrative domain monitored resource entries for additional objects supported by the new release in a cluster administrative domain or IASP.

15.9.2 Performing release upgrade while retaining current production system

Using the “switching while detached” mechanism allows you to use the backup node for production activity after the release upgrade while preserving the production node data.

With this option, the following steps apply:

1. If you are using geographic mirroring, you can suspend it with tracking.
2. Upgrade IBM i and PowerHA on the current backup node and restart the backup cluster node.
3. If you are using geographic mirroring and suspended it, resume it to synchronize the IASP from the current production to the current backup node.
4. Perform a detach operation for your ASP session. At this time the independent ASP on both sides can be varied on.
5. Allow the users to connect to the upgraded node, which is still the backup node from a cluster standpoint. If problems exist, you can then route user connections to the normal production node, with data current at the time of the detach operation.
6. If there are no or minor problems on the backup node, you can decide to make this node the new production one in preparation for upgrading the previous production node. Use **`CHGCRG ... RCYDMNACN(*CHGCUR)`** to switch node roles.
7. Upgrade IBM i and PowerHA on the current backup node and restart clustering. Perform a reattach operation on the current backup node to synchronize current production data on the current backup node. Note that this option submits a full synchronization process for geographic mirroring as both IASP copies had been varied on.
8. When all cluster nodes run the same release, update the PowerHA version with **`CHGCLUVER`**.

   Note: **`CHGCLUVER`** might need to be used twice to update the PowerHA version by two if you skipped a release at the upgrade, for example, from V5R4 to i 7.1.

9. Optional: Switch back to your preferred production node.
10. You might want to review your cluster administrative domain monitored resource entries for additional objects supported by the new release in a cluster administrative domain or IASP.

15.10 Integration with BRMS

This section describes one way to use BRMS for saving to tapes operations in order to keep track of operations in the current BRMS production database, even they effectively run on a backup node. We use the possibility for BRMS to have its own system name distinct from the network attribute one. Refer to this Knowledge Base document for more information about BRMS system name:

`http://www-912.ibm.com/s_dir/SLKBase.nsf/a9b1a92c9292c2818625773800456abc/ff759abf1f796e8e8625715d00167c07?OpenDocument`

15.10.1 Normal BRMS usage

You might prefer running the save-to-tape operations on the backup node and keeping save information here (that is, in the backup BRMS database) even for production IASP copies saves. In this case, there is no difference from a standard BRMS implementation and, therefore, this section does not apply.

However, you will have to deal with questions like “What was the system name for the last backup tape for this file I need to restore now?” because saves to tapes can potentially run on production and backup nodes.

BRMS network configuration can help you. Make sure to share the media information at the library level. With this setting, to find the last backup for your file you can run the following command on each system of your production node, which can run the save to tapes:

```
WRKMEDIBRM LIB(MYLIB) ASP(MYASP) FROMSYS(MYSYS)
```

To restore the desired object (assume that the last save was done on the MYSYS system), set your job to use the appropriate IASP, then run the following command:

```
RSTOBJBRM OBJ(MYFILE) SAVLIB(MYLIB) DEV(*MEDCLS) SAVLVL(*CURRENT) FROMSYS(MYSYS)
```

There is another disadvantage with this usage. BRMS does not allow appending backups to tape owned by another BRMS system, and retention calculation is done at the BRMS system level. This means that you might need much more media than with the proposed implementation discussed later.

15.10.2 Production IASP copy save-to-tapes on backup nodes

If you want BRMS to run save-to-tape on a backup node for production IASP copies and to update the production BRMS database just like saves that were done on the PRODUCTION system, there are considerations to take care of when using a BRMS network, using a BRMS name, and sharing tapes and locations.

In any case, when using a BRMS network, we recommend using a dedicated IP addresses for BRMS.

Suppose that we have three systems in a BRMS network.

- PRODUCTION is the system name that runs production activities as a preferred role. The PRODUCTION name is longer than allowed (eight characters maximum), but this name is only for description purposes.
- BACKUP is the system name of the current and preferred backup node with an active ASP mirroring. BACKUP is also used for save-to-tape operations, for example, after detaching the ASP session to make the target ASP available. Or BACKUP can also be another node using a FlashCopy target IASP.
- OTHER is a third system.

In this specific case, when a save operation is scheduled to run on a system with a BRMS name different from the system name, other BRMS network members must be aware of the save operation. They need to know the current tape usage by the system running the save, in place of the usual system and they need to update the BRMS database of the system running the save.

When there is no save operation for the usual production system, the BRMS OTHER system must connect to the BRMS PRODUCTION system to be aware of real-availability tapes, and it must update the PRODUCTION BRMS database when it writes to available tapes.

Figure 15-6 shows the BRMS communication flow. The BRMS OTHER system communicates with the BRMS PRODUCTION system.

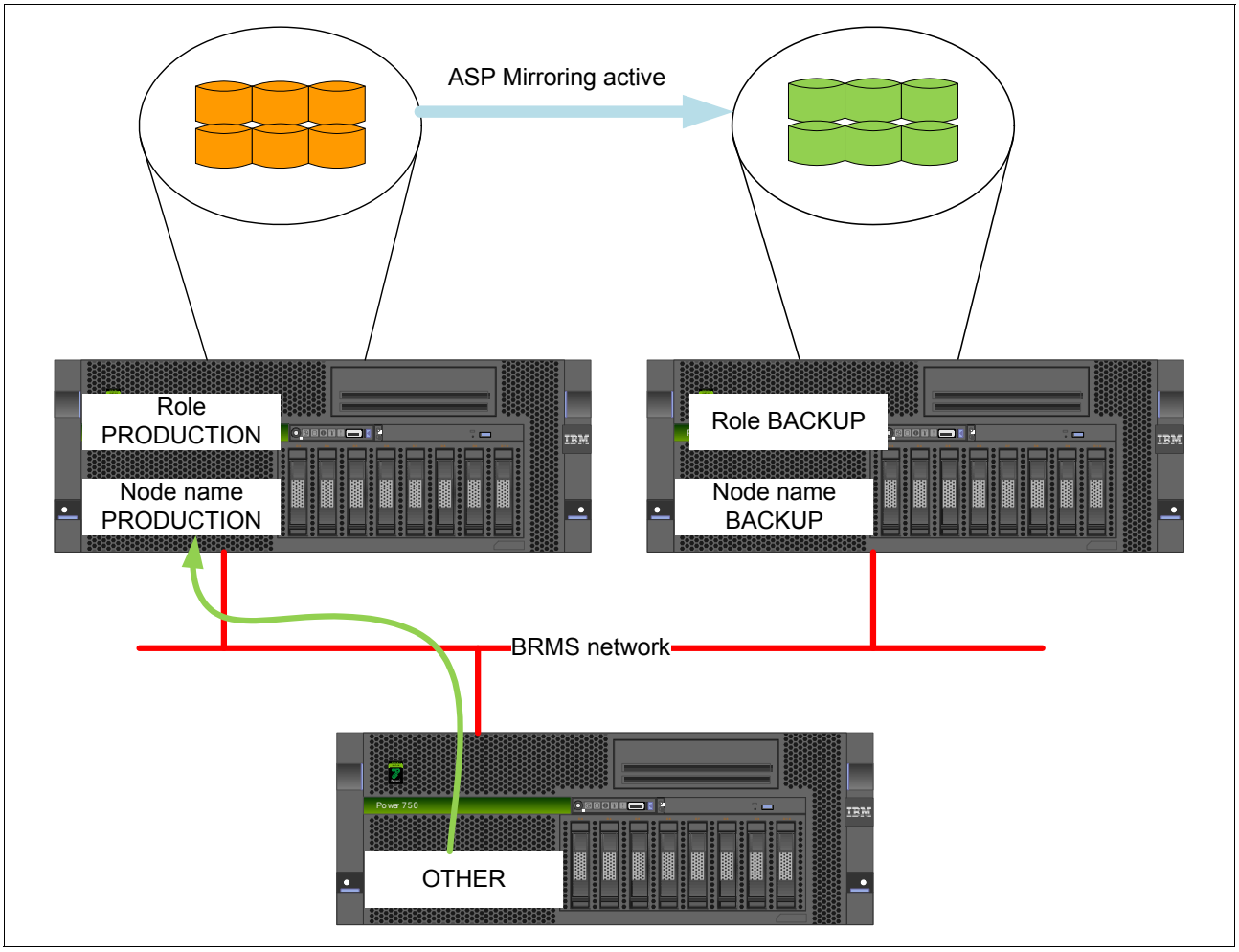


Figure 15-6 BRMS network when no save operation is running and roles are the preferred ones

When running a save-to-tape for PRODUCTION ASP copies, the BRMS OTHER system must no longer communicate with the BRMS PRODUCTION system but instead with the BACKUP system because it is now acting like it is PRODUCTION for the BRMS activity. As shown in Figure 15-7, the easiest way to perform this swap, without changing anything on the OTHER system, is to swap the BRMS PRODUCTION IP address on the BACKUP system.

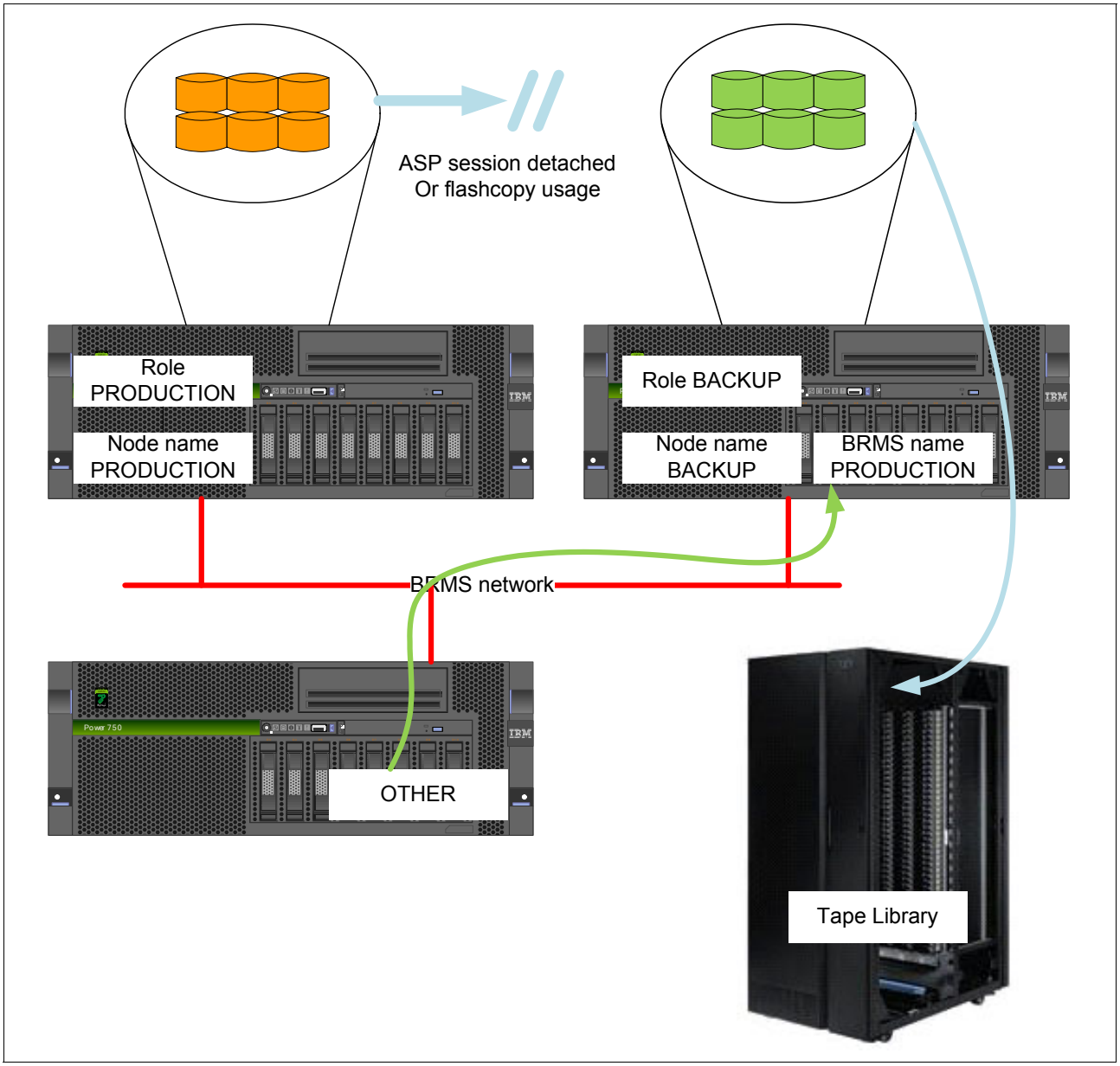


Figure 15-7 BRMS network when running a save-to-tape operation on the backup node and the roles are preferred ones

These are the steps to update the PRODUCTION BRMS database when running this ASP save on the BACKUP system:

1. Make independent ASP available on the BACKUP system by using the ASP session with the detach option or FlashCopy procedures.
2. On the PRODUCTION system, we need to send BRMS information to the BACKUP system:
   a. End the Q1ABRMNET subsystem.
   b. End the BRMS IP interface.
   c. Save the QUSRBRM library to a save file and send it to the BACKUP system.
3. Make sure to understand that, from this point, any update to the PRODUCTION BRMS database will be lost. This means that you should not run any save or media operation on PRODUCTION from this time on. The restore operation can be run, but related BRMS history will be lost.
4. On the BACKUP system, we need to save the current BRMS information:
   a. End the Q1ABRMNET subsystem.
   b. End the BACKUP BRMS IP interface.
   c. Save the QUSRBRM library to a save file.
   d. Clear the QUSRBRM library.
5. On the BACKUP system, we want the BRMS database to be the PRODUCTION one:
   a. Restore QUSRBRM objects from the PRODUCTION save file.
   b. Change the BRMS name to PRODUCTION:
      i. Using the BRMS Q1AOLD API with this command:

         ```
         QSYS/CALL QBRM/Q1AOLD PARM('BRMSYSNAME' '*SET ' 'PRODUCTION').
         ```

         This command creates the QUSRBRM/Q1ABRMSYS data-area.
      ii. The BACKUP system acts now as the PRODUCTION BRMS system.
   c. Start the PRODUCTION BRMS IP interface.
   d. Start the Q1ABRMNET subsystem.

6. On the BACKUP system, submit the save-to-tape for independent ASP libraries and directories. Figure 15-8 shows an example of a control group.

```
Create Backup Control Group Entries              DEMOHA

Group . . . . . . . . . . . : IASP1
Default activity  . . . . . *BKUPCY
Text  . . . . . . . . . . . *NONE

Type information, press Enter.

                                Weekly   Retain Save     SWA
     Backup     List ASP        Activity Object While    Message    Sync
Seq  Items      Type Device     SMTWTFS  Detail Active   Queue      ID

  10 *ALLUSR         IASP1      *DFTACT  *ERR   *NO
  20 *LINK           IASP1      *DFTACT  *NO    *NO

                                                                    Bottom
F3=Exit    F5=Refresh    F10=Change item    F11=Display exits
F12=Cancel    F14=Display client omit status    F24=More keys
```

Figure 15-8 BRMS control group for saving IASP1 objects only

Note: SYSBASE objects can also be included in this backup if they are included in the administrative domain and, therefore, are identical on both PRODUCTION and BACKUP nodes.

7. On the BACKUP system, we prepare the comeback for the BRMS environment:
   a. End the Q1ABRMNET subsystem.
   b. End the PRODUCTION BRMS IP interface.
   c. Delete the QUSRBRM/Q1ABRMSYS data area.
8. On the BACKUP system, we need to send the BRMS database to the PRODUCTION system:
   a. Save the QUSRBRM library to a save file.
   b. Send it to the PRODUCTION system.
9. On the PRODUCTION system, we need to restore the BRMS database updated with the save-o-tape operation run on the BACKUP system. If any update has been done to the PRODUCTION BRMS database before this point, it is lost. Take these steps:
   a. Clear the QUSRBRM library.
   b. Restore objects from the save done in step 8.
10. On the PRODUCTION system, start the BRMS IP interface and the Q1ABRMNET subsystem. Any BRMS save and tape information regarding IASP is now on the PRODUCTION system, and from now on this one is used within the BRMS network by other members.

11. On the BACKUP system, we need to set back the usual BRMS BACKUP database so that we can run SYSBASE BACKUP-specific saves to tape:
    a. Clear the QUSRBRM library.
    b. Restore objects from the save taken in step 4 on page 419.
    c. Start the BACKUP BRMS IP interface.
    d. Start the Q1ABRMNET subsystem.
12. If the IASP was made available by changing an ASP session with a detach option, it can be reattached at this time.

Note: BRMS can have only one system name at any time. This means that you cannot run several saves to tapes at the same time for IASP copies from several nodes. You can still use a single node for saves to tape of IASP copies from several nodes, but they must be done in a serial way. You have to run the steps described in this section for each production system, one after the other.

15.10.3 Run production IASP copy saves to tapes after a roles switch

The same concern exists after a role switch (that is, when the preferred backup node becomes the production node and the preferred production node becomes that backup node, but they keep their system name). You will probably want to run the save-to-tape operations on the actual backup node.

After roles switch, when there is no save in progress, the BRMS network is as shown in Figure 15-9.

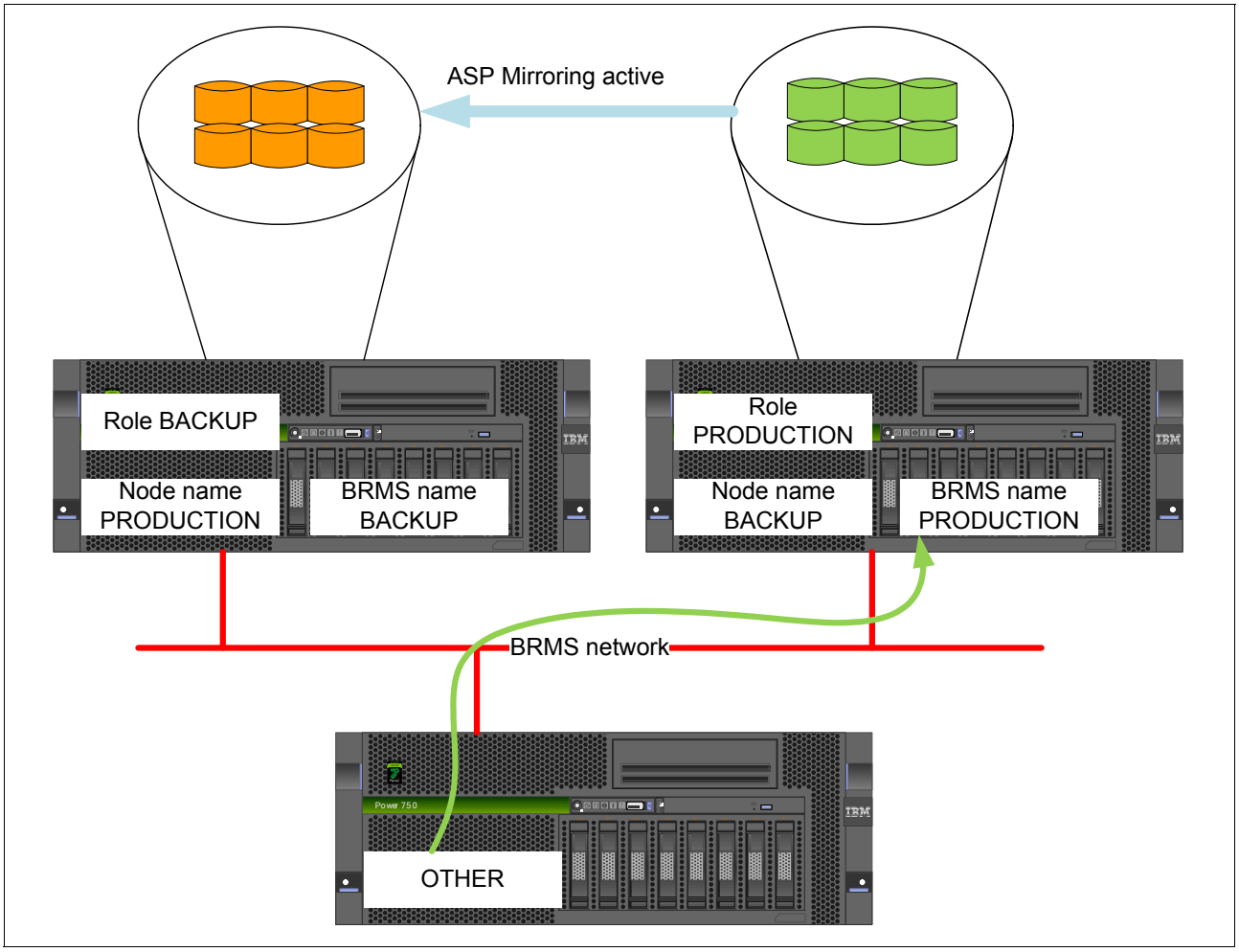


Figure 15-9 BRMS network when no save operation is running and roles are switched ones

When running a save-to-tape from preferred production node while roles are switched ones, the network is as shown in Figure 15-10.

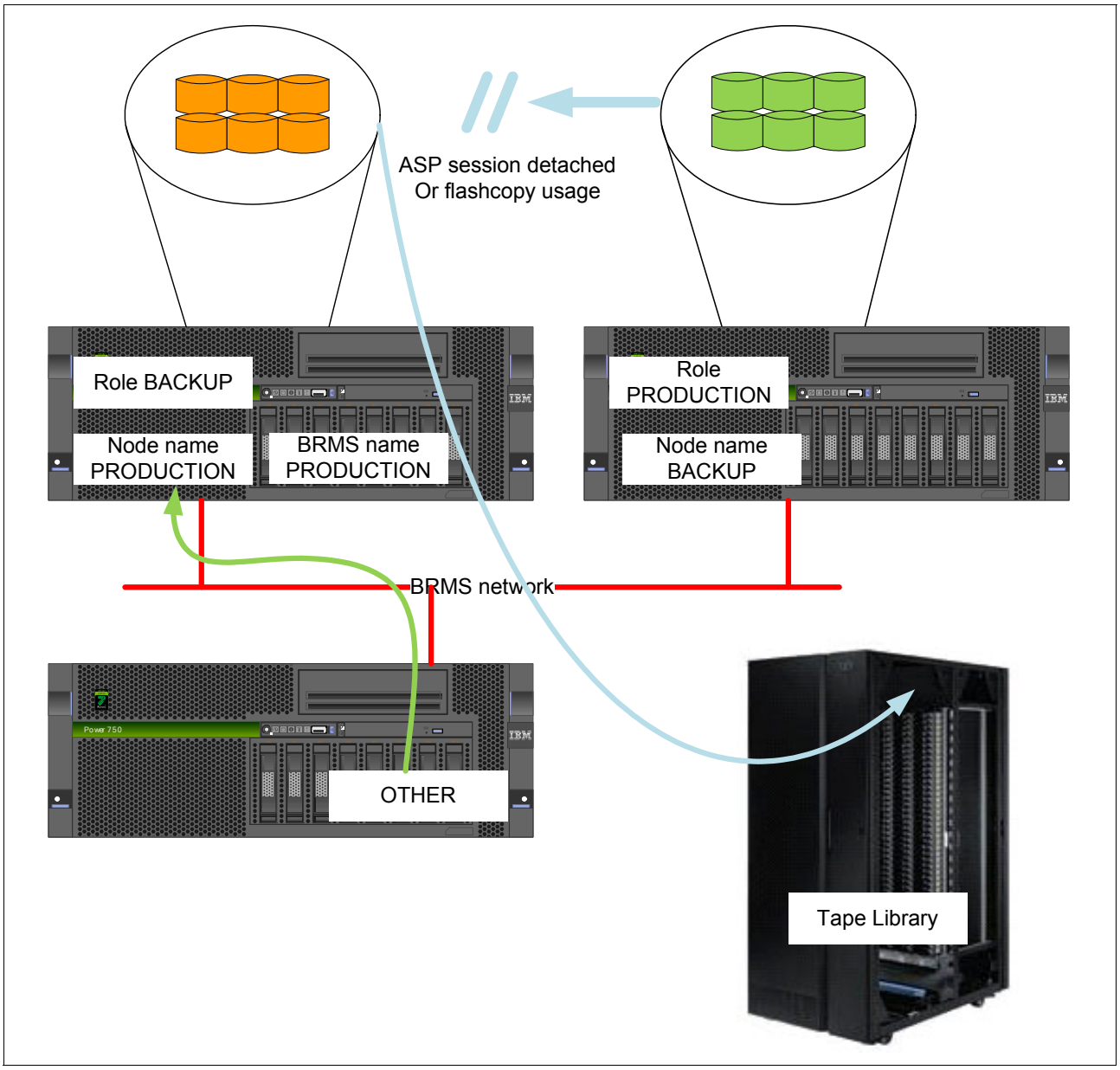


Figure 15-10 BRMS network when running a save-to-tape operation on the backup node and the roles are switched ones

After switching roles, when there is no BRMS activity, BRMS names must be swapped, and related IP interfaces as well:

- On a production role node:
  a. Start the BRMS production IP interface.
  b. Change the BRMS system name to PRODUCTION with Q1AOLD API, as shown above.
- On backup role node:
  a. Start the BRMS backup IP interface.
  b. Change the BRMS system name to BACKUP with the Q1AOLD API, as shown above.

Take these steps to update the PRODUCTION BRMS database when running this ASP save on the current backup node (preferred production node):

1. Make the independent ASP available on the backup by using the ASP session with detach option or FlashCopy procedures.
2. On the production role node, we need to send BRMS information to backup role node:
   a. End the Q1ABRMNET subsystem.
   b. End the production BRMS IP interface.
   a. Save the QUSRBRM library to a save file and send it to the backup role node.
3. Make sure to understand that, from this point, any update to the PRODUCTION BRMS database will be lost. This means that you should not run any save or media operation on the production role node from this time. A restore operation can be run, but any related BRMS history will be lost.
4. On the backup role node, we need to save the current BRMS information:
   a. End the Q1ABRMNET subsystem.
   b. End the backup BRMS IP interface.
   c. Save the QUSRBRM library to a save file.
   d. Clear the QUSRBRM library.
5. On the backup role node, we want the BRMS database to be the PRODUCTION one:
   a. Restore QUSRBRM objects from the production save file.
   b. The backup role node acts now as production BRMS system because of QUSRBRM/Q1ABRMSYS data area content.
   c. Start the production BRMS IP interface.
   d. Start the Q1ABRMNET subsystem.
6. On the backup role node, submit the save-to-tape for independent ASP libraries and directories. Figure 15-8 on page 420 shows an example of a control group.
7. On the backup role node, we prepare the comeback for the BRMS environment:
   a. End the Q1ABRMNET subsystem.
   b. End the production BRMS IP interface.
8. On the backup role node, we need to send the BRMS database to the production role node:
   a. Save the QUSRBRM library to a save file.
   b. Send it to the production role node.
9. On the production role node, we need to restore the BRMS database updated with the save-to-tape operation run on the backup role node. If any update has been done to the PRODUCTION BRMS database before this point, it is lost. Take these steps:
   a. Clear the QUSRBRM library.
   b. Restore objects from the save taken in step 8.
10. On the production role node, start production the BRMS IP interface and the Q1ABRMNET subsystem. Any BRMS save and tape information regarding IASP is now on the production role node, and from now on this one is used again within the BRMS network by other members.

11. On the backup role node, we now need to set back the usual BRMS database to be able to run SYSBASE-specific saves to tape:
    a. Clear the QUSRBRM library.
    b. Restore objects from the save taken in step 4 on page 424.
    c. Start the backup BRMS IP interface.
    d. Start the Q1ABRMNET subsystem.
12. If the IASP was made available with changing an ASP session with the detach option, it can be reattached at this time.

15.10.4 Specific system synchronization

Another way to run saves to tape on a backup system and make the production system own the backup history just like it did it is to use specific system synchronization. More information can be found on the BRMS website:

http://www-03.ibm.com/systems/i/support/brms/new.html#foo7

We decide on the backup system (that is, the one that runs the saves to tapes), which the production system will look like it did it itself. Configuration steps are done either through the BRMS GUI or with the QBRM/Q1AOLD API.

To add a specific system synchronization, change the system name to make it look like the backup was done by this system and synchronize the reference date/time using this command:

```
CALL QBRM/Q1AOLD PARM('HSTUPDSYNC' '*ADD' ‘SYSTEM’ ‘NETWORK ID' 'IASPNAME'
'*CHGSYSNAM')
```

To add a specific system synchronization, keep the name of who did the backup and synchronize the reference date/time:

```
CALL QBRM/Q1AOLD PARM('HSTUPDSYNC' '*ADD' ‘SYSTEM' ‘NETWORK ID' 'IASPNAME '
'*NORMAL')
```

15.10.5 User-defined IASP timestamps

When running saves to tapes on backup systems, after a FlashCopy or after detaching an ASP session, it might be interesting to supply a user-defined timestamp for the backup. A FlashCopy or detach operation can occur at one time, and specific save-to-tape operations later, but you want to keep the FlashCopy or detach operation time for the save reference, and the reference point for future incremental backups. Still using the QBRM/Q1AOLD API, it is possible to define a timestamp for IASP save to tape.

To add a timestamp, use this command:

```
CALL QBRM/Q1AOLD PARM('FLASHTIME' '*ENABLE' 'IASPNAME' 'FILESYSTEM TYPE'
'timestamp')
```

Refer to the same website as given in 15.10.4, “Specific system synchronization” on page 425, for more details.

15.10.6 SYSBASE save-to-tape considerations

We can consider SYSBASE objects included in the administrative domain in the same way as IASP copies. This means that saving them to tape can be run at the same time that the IASP is saved to tape and that they belong to the production role node wherever they are saves to tape.

All other SYSBASE objects not included in the administrative domain belong to the node on which they are installed, whatever its role. This means that they have to be saved *while the BRMS system name is not overridden*, when it is the same as the network attribute system name. This means, for example, that when you are running the preferred configuration, you have to run SYSBASE-specific backup node saves-to-tape outside of a regular production-related control group.

15.11 Hardware replacement in a PowerHA environment

In this section we describe the considerations for replacing either an IBM i server or storage system in an IBM PowerHA SystemMirror for i environment.

15.11.1 Server replacement

Depending on the PowerHA System Mirror solution that you implemented, replacing one or two of the servers in your high availability environment involves a number of steps, which we discuss in this section.

Geographic mirroring

When replacing the backup system, including its disks, in an environment using geographic mirroring with internal disks you cannot simply do a save and restore operation. Perform these steps:

1. To preserve your old backup system in case migration to the new backup system fails, perform a detach of the ASP session using **`CHGASPSSN OPTION(*DETACH)`**.
2. Power down the old backup system.
3. Deconfigure geographic mirroring from the production system using either the GUI interfaces or **`CFGGEOMIR`**. This results in an error message stating that the backup system cannot be found. You can tell the system to ignore this status and to proceed with the deconfiguration. Be aware that your production IASP needs to be varied off to perform this step.
4. Start the new backup system. The system should have unconfigured drives available to become the new backup IASP. Make sure that clustering works between the production system and the new backup system. It might be necessary to remove and add the backup cluster node to and from the cluster and recovery domain.
5. Configure geographic mirroring from the production system to the new backup system using either the GUI interfaces or **`CFGGEOMIR`**. Be aware that your production IASP needs to be varied off to perform this step.
6. When the configuration of geographic mirroring is finished, vary on the production IASP and make sure that geographic mirroring is active. A full resynchronization is required.

Exchanging the production system without first deconfiguring geographic mirroring and reconfiguring it afterwards is also not possible. Consider doing a **`CHGCRGPRI`** to switch over to the backup system, and then follow the steps described above.

To replace a backup server only in a geographic mirroring environment using external storage, make sure to suspend geographic mirroring from the production site first. Then power down the old backup server, attach the new server to the existing external storage, and restart the new backup server. Finally, resume geographic mirroring to run a partial synchronziation.

When replacing the production server only in a geographic mirroring environment using external storage, either switch to the backup system first or make sure to properly end your production system (vary off the IASP, ENDCRG, ENDCLUNOD, before PWRDWNSYS) before exchanging the server hardware.

Remote Copy Services environments

When using IBM PowerHA System Mirror for i with remote Copy Services for either DS6000, DS8000, or SVC/V7000, consider this:

- Consider exporting your old server LPAR configuration to a system plan to have current documentation available for either manual or automatic redeployment by importing the (modified) system plan. Make sure to set up your partitions in the same way that they were set up on the old server. Use specific care for virtual adapter numbers when using VIOS.
- Consider exchanging the server results in new WWPNs for server Fibre Channel adapters. On DS6000 and DS8000, host connections have to be deleted and recreated. On SVC/V7000, first add the new host ports and then delete the old ones, preserving existing vdisk mappings to the host.

Make sure to change your SAN switch zoning for the new server WWPNs.

15.11.2 Storage system replacement

If a remote copy secondary storage system needs to be replaced, plan for these actions:

- Make sure that the IBM i cluster node accessing the secondary storage system is the current backup node.
- If this IBM i backup node has its SYSBAS configured on the secondary storage system, perform an entire system save for this IBM i node before powering it down. Otherwise, just end the remote copy ASP sessions.
- Stop remote mirroring on the storage system.
- Perform the secondary storage system replacement procedure.
- Change any SAN switch zoning information for the new WWPNs from the new secondary storage system.
- Re-create the required logical storage configuration for host ports/connections and volumes on the new secondary storage system.
- If the IBM i backup node had its SYSBAS configured on the secondary storage system, restore SYSBAS for this IBM i node.
- If this IBM i backup node was powered down before, restart the IBM i backup node connected to the new secondary storage system with restarting clustering.
- Re-establish and start the remote copy paths and IASP volume relationships between the primary and new secondary storage system. This requires a full-resynchronization between primary and secondary volumes.
- Change the ASP copy descriptions for the secondary storage system. Even if the logical configuration did not change, at least the serial number information needs to be updated for the new storage system.
- Start the remote copy ASP sessions again.

If a DS8000 or SVC/V7000 remote copy secondary storage system is not replaced but needs to be powered-cycled for a disruptive maintenance action, the primary storage system keeps track of the changes such that only a partial resynchronization of the out-of-sync tracks will be required after the secondary storage system becomes operational again. Usually this

resynchronization would have to be started manually for both DS8000 and SVC/V7000. An exception is DS8000 Global Mirror, which automatically resumes suspended Global Copy relationships and restarts the Global Mirror session.

15.12 Problem data collection

Prior to taking the steps to gather data, ensure that the HA environment is always up to date with PTFs, which can be found on the Recommended Fixes website. There are PTFs added frequently that address various HA-related issues. Also, when missing the suggested PTFs, data capture can also sometimes have difficulties. See 8.1.3, “Power Systems requirements” on page 135, for the PTF requirements.

Note: If you are using the manual method of collecting data, you must collect it on *all* nodes of your cluster.

15.12.1 IBM i clustering

In this section we discuss IBM i clustering.

Node not starting

A node in the cluster cannot be started. Using **STRCLUNOD** results in an error message or hangs indefinitely. Take these steps:

1. Check the joblog of the process in which **STRCLUNOD** was run.

   If there is CPFBB98, it gives reason codes for the problems with a node start.
2. Check the QCSTCTL, QCSTCRGM, or the actual CRG joblog.
3. Check the QHASVR job (PowerHA for i 7.1 and later). Check the joblog if there are any problems.
4. Make sure that user QCLUSTER has no jobs in the QBATCH jobqueue waiting to be started. Usually, the QBATCH job queue has only one maximum job. Check this on all nodes.

Cluster resources group or administrative domain

If there is problem with the cluster resource group or the administrative domain, check that the following information is true:

- The jobs are named the same as the cluster resource groups. Its jobs should not be running in any of the subsystems.
- The job are named the same as the administrative domain.
- Check the QCSTCTL and QCSTCRGM jobs. These are the system jobs not running in any of the subsystems.

Problem data collection to be sent to IBM

In this section we discuss problem data collection to be sent to IBM.

Tip: IBM Support might instruct you to collect additional data not included in these instructions, or they might require less data then listed here. Contact IBM Support with your problem and you will get the instructions for your specific case.

The problem data collection that needs to be sent to IBM can be gathered by using **QMGGTOOLS** (see the 15.12.3, “The Must Gather Data Collector” on page 438) or manually. Using **QMGTOOLS** is the preferred way to collect the diagnostic data for the HA solution problems. In the following sections we describe both of these ways to collect the data.

Collecting the data with QMGTOOLS

To collect the data with **QMGTOOLS**, you need to collect general data on one of your nodes, following the instructions in “Collecting the general cluster data for all nodes” on page 439. Then collect on each of the nodes QSYSOPR and QHST messages for the time of failure.

Collecting the data manually

To prepare data to be sent to IBM for problem analysis, take the following steps on all of the nodes in your cluster:

1. Run this command:

   ```
   DSPCLUINF CLUSTER(cluster_name) DETAIL(*FULL) OUTPUT(*PRINT)
   ```

2. Run this command:

   ```
   DSPCRGINF CLUSTER(cluster_name) CRG(*LIST) OUTPUT(*PRINT)
   ```

3. Run this command:

   ```
   DSPCRGINF CLUSTER(cluster_name) CRG(cluster_resource_group_name) OUTPUT(*PRINT)
   ```

4. Joblogs of QCSTCTL, QCSTCRGM, and the joblogs for the Cluster Resource Group (CRG) and administrative domain: These are named same as CRG and Administrative Domain will be required.

5. Run this command:

   ```
   DMPCLUTRC CLUSTER(cluster_name) CRG(*ALL) NODE(*ALL) LEVEL(*ERROR)
   ```

6. Run this command for each copy description:

   ```
   DSPASPCPYD ASPCPY(asp_copy_description) OUTPUT(*PRINT)
   ```

7. Run this command for each session:

   ```
   DSPASPSSN SSN(session_name) OUTPUT(*PRINT)
   ```

8. Collect the QSYSOPR and QHST messages for the time of failure.

9. Start SST (**strsst)**, log in, and collect the following data:
   a. LIC Logs (strsst → 1. Start a service tool → 5. Licensed Internal Code log → 2. Dump entries to printer from the Licensed Internal Code log).

      On page shown in Figure 15-11, set these options:

      - Dump option 1 and option 3. As a result you get two spool files.
      - Set the starting and ending date and time to at least two hours before the problem occurred.

```
Dump Entries to Printer from Licensed Internal Code Log

Type choices, press Enter.

  Dump option . . . . . .   1          1=Header
                                       2=Header and note entry
                                       3=Header and entire entry
  Entry ID:
    Starting  . . . . . .   00000000   00000000-FFFFFFFF
    Ending  . . . . . . .   FFFFFFFF   00000000-FFFFFFFF

  Entry type:
    Major code  . . . . .   0000       0000-FFFF
    Minor code  . . . . .   0000       0000-FFFF

  Starting:
    Date  . . . . . . . .   09/20/11   MM/DD/YY
    Time  . . . . . . . .   00:00:00   HH:MM:SS
  Ending:
    Date  . . . . . . . .   09/20/11   MM/DD/YY
    Time  . . . . . . . .   00:00:00   HH:MM:SS

F3=Exit   F12=Cancel
```

Figure 15-11 Dump LIC Log

b. Get the product activity log (PAL) entries for a couple of days before the problem occurred (for example, a 7-day period) (strsst → login → 1. Start a service tool → 1. Product activity Log → 1. Analyze log).

When you see the page shown in Figure 15-12, specify the following settings:

- Log type 1 (All logs)
- The From and To dates to include last days (for example, 7) including the time of the problem occurrence

Press Enter.

```
Select Subsystem Data

Type choices, press Enter.

  Log . . . . . .  1          1=All logs
                              2=Processor
                              3=Magnetic media
                              4=Local work station
                              5=Communications
                              6=Power
                              7=Cryptography
                              8=Licensed program
                              9=Licensed Internal Code

  From:
    Date . . . . . .   09/22/11   MM/DD/YY
    Time . . . . . .   14:55:28   HH:MM:SS

  To:
    Date . . . . . .   09/29/11   MM/DD/YY
    Time . . . . . .   14:55:28   HH:MM:SS

F3=Exit            F5=Refresh          F12=Cancel
```

Figure 15-12 PAL options

On the next page (Figure 15-13) chose these options:

- Report type 3 (Print options)
- Optional entries to include, Statistic Y (for Yes)

Press Enter.

```
Select Analysis Report Options

Type choices, press Enter.

  Report type . . . . . .   3   1=Display analysis, 2=Display summary,
                                3=Print options
  Optional entries to include:
    Informational . . . . .   Y   Y=Yes, N=No
    Statistic . . . . . .     Y   Y=Yes, N=No

  Reference code selection:
    Option  . . . . . .       1   1=Include, 2=Omit
    Reference codes
    *ALL                                          *ALL...

  Device selection:
    Option  . . . . . .       1   1=Types, 2=Resource names
    Device types or Resource names
    *ALL                                          *ALL...



F3=Exit       F5=Refresh       F9=Sort by ...        F12=Cancel
```

Figure 15-13 PAL print options

On next page (Figure 15-14), choose these options:

- Report type 4 (full report)
- Include hexadecimal data Y (Yes)

Press Enter.

```
Select Options for Printed Report

Type choices, press Enter.

  Report type  . . . . . .   4     1=Print analysis report
                                   2=Print summary report
                                   3=Print partial report
                                   4=Print full report


    For report type 4:
      Include hexadecimal
        data . . . . . .     Y   Y=Yes
                                 N=No

F3=Exit              F5=Refresh           F12=Cancel
```

Figure 15-14 PAL print report options

c. The general instructions for how to run the tool can be found in “Instructions for running the advanced analysis macro” on page 434.

Data to collect for cluster resource group problem

When a problem is related with the CRG, in addition to general, data you might be required to collect the advanced analysis macros (see “Instructions for running the advanced analysis macro” on page 434) for these:

- IOSW with no options
- HRIFR with -ipl0 option
- IOHRIDEBUG with option -walklhrichildren -oneliner
- IOHRIDEBUB with option -walklhrichildren -detail
- IOHRITOOLHDWRPT with option -hidden
- IOHRITOOLHDWRPT with option -partofcrg
- IOHRITOOLHDWRPT with option -assocrscofcrg

Data to collect when problem is related to administrative domain

In addition to the general data collection, collect the following data:

- If you have a problem with a specific monitored resource run **PRTMRE DETAIL(*FULL)** for this resource.

  The following steps should be done if you are *not* using QMGTOOLS for general data collection.

- Run the following commands and collect the spool files produced by them:

```
DMPSYSOBJ OBJ(QFPATRES) CONTEXT(QSYS)
DMPSYSOBJ OBJ(QFPATATR) CONTEXT(QSYS)
DMPSYSOBJ OBJ(QFPATMAP) CONTEXT(QSYS)
DMPSYSOBJ OBJ(QFPATCAE) CONTEXT(QSYS)
```

- The joblog of the QPRFSYNCH job and any other spoolfiles that were generated by this job.

Data to collect if you are using geographic mirroring

In addition to the general data collection, run the following advanced analysis macros (see “Instructions for running the advanced analysis macro” on page 434):

- GEOSTAT with option -ALL
- DSMINFO without any options
- ASMINFO with option -ASP *nnn* -r 50000 (Where *nnn* is the 3-digit IASP number. For example, for IASP no. 33 you need to specify -ASP 033.)

Data to collect when using DS8000 TotalStorage

When using DS8000 for your IASP and PowerHA solutions, in addition to general data, collect DSCLI logs from `/QIBM/UserData/HASM/hads/dscli/log`.

In case you are *not* using QMGTOOLS, you need to collect XSM logs that can be found in `/QIBM/UserData/HASM/hads`.

Instructions for running the advanced analysis macro

Depending on the problem area within the high-availability environment, IBM Support might ask you to collect advanced analysis macros. The following steps can be used to display or print advanced analysis macros:

1. Start service tools with **strsst**.
2. Log in using the DST user and password (passwords are case sensitive).
3. Select option 1. Start a service tool.
4. Select option 4. Display/Alter/Dump.
5. Select option 2. Dump to printer (or 1. Display/Alter storage if you just want to see the results on the page without creating spoolfile).
6. Select option 2. Licensed Internal Code (LIC) data.
7. Select option 14. Advanced analysis.

You will get a page similar to Figure 15-15.

```
Select Advanced Analysis Command

Output device  . . . . . :   Printer

Type options, press Enter.
  1=Select

Option    Command

          FLIGHTLOG
          ADDRESSINFO
          ALTSTACK
          BATTERYINFO
          CLUSTERINFO
          CONDITIONINFO
          COUNTERINFO
          DISABLEFLASHSYNC
          DSTINFO
          EXCEPTCHAIN
          FINDFRAMES
          FINDPTF
                                                                  More...
F3=Exit   F12=Cancel
```

Figure 15-15 AA panel

Now you can search a macro that you want to run:

1. Specify 1 in the Option column and press Enter.
2. On the next page specify the options that you need (if any), and press Enter.
3. Chose the printing options (if any), and press Enter again.

If you do not find the macro name that you want to run on the list, you can run it by specifying 1 in the Options column in the first row (the empty one) and specifying the macro name in the Command column.

Figure 15-16, Figure 15-17 on page 437, and Figure 15-18 on page 437 show an example of running the AAExample macro with the -ALL option.

```
Select Advanced Analysis Command

Output device  . . . . . :   Printer

Type options, press Enter.
  1=Select

Option    Command
   1      AAExample
          FLIGHTLOG
          ADDRESSINFO
          ALTSTACK
          BATTERYINFO
          CLUSTERINFO
          CONDITIONINFO
          COUNTERINFO
          DISABLEFLASHSYNC
          DSTINFO
          EXCEPTCHAIN
          FINDFRAMES
          FINDPTF
                                                                    More...
F3=Exit   F12=Cancel
```

Figure 15-16 AAExample macro example

```
Specify Advanced Analysis Options

Output device  . . . . . :   Printer

Type options, press Enter.

  Command . . . . :   AAExample

  Options . . . . .   -all









F3=Exit   F4=Prompt   F12=Cancel
```

Figure 15-17 AAExample options

```
Specify Dump Title

Output device  . . . . . :   Printer

Type choices, press Enter.

  Dump title  . . . . . .

  Perform seizes  . . . . . .   1           1=Yes, 2=No

  Partial print page numbers:
    From page . . . . . .           1   1-2147483647
    Through page  . . . . . .    9999   1-2147483647









F3=Exit   F12=Cancel
```

Figure 15-18 AA_Example printing options

15.12.2 PowerHA GUI

In this section we discuss the PowerHA GUI.

Problems with accessing the GUI for PowerHA

The PowerHA for System i provides a graphical user interface within IBM Systems Director Navigator for i that is a part of the HTTP Server ADMIN instance. We recommend that you have the current level of group PTFs for HTTP Server and Java.

If you have problems accessing the IBM Systems Director Navigator for i, make sure that you have started the HTTP Server ADMIN domain. Use **`STRTCPSVR SERVER(*HTTP) HTTPSVR(*ADMIN)`**. Check the ADMIN and ADMIN1-4 jobs in QHTTPSVR subsystem.

Additional information about the problem can be found in the following directories:

- `/QIBM/UserData/HASM/logs`
- `/QIBM/UserData/OS/OSGi/LWISysInst/admin2/lwi/logs`

Collect the Java dump of the ADMIN2 job:

```
Wrkjob admin2
Option 45
Option 32
```

Get the most recent file named `javacore.xxx` from the `/QIBM/UserData/OS/OSGi/LWISysInst/admin2/lwi/runtime/core` directory.

Problems with operations in PowerHA GUI

When problems occur in PowerHA, you can see detailed messages about the problem in the web GUI.

Additional information about PowerHA GUI operations can be found in the `/QIBM/UserData/HASM/logs/` directory.

Problem data collection for PowerHA GUI

To collect the diagnostic data to be sent to IBM Support, see “Collecting the data for the PowerHA GUI with QMGTOOLS” on page 445.

15.12.3 The Must Gather Data Collector

The Must Gather Data Collector is a tool provided by IBM Support for automation of the diagnostic data collection. The tool functionality is continuously developed and new functions are added, so we recommend that you have the latest version of the tool[1].

Installation of the Must Gather Data Collector

A *SAVF containing the QMGTOOLS library can be found at this website:

`https://www-912.ibm.com/i_dir/idoctor.nsf/downloadsMGT.html`

The PTFs listed on that website are related to the iDoctor functions. If you are not planing to use QMPGTOOLS for iDoctor functionality, you do not need to install them. Check the PTFs for high availability mentioned in 8.1.3, “Power Systems requirements” on page 135.

[1] We based our example in this book on the tool version that is current at the time of writing. The tool might have been updated since then.

You need to download QMGTOOLS for your IBM i version, transfer SAVF to the system, and restore the QMGTOOLS library[2]. To restore the QMGTOOLS library, use this command:

```
RSTLIB SAVLIB(QMGTOOLS) DEV(*SAVF) SAVF(SAVF_LIB/QMGTOOLvrm)
```

Where *SAVF_LIB* is the library where you put the downloaded SAVF, and *QMGTOOLvrm* is a SAVF name for your IBM i release.

To collect the data with QMGTOOLS it is enough to install the tool on one of the nodes only.

When you restore the QMGTOOLS library you can then use ADDLIBLE QMGTOOLS and issue **GO MG** to access the tools main menu and then select option 1 to access the HA data collection, or you can go straight to this menu with **GO HASMNU**.

Collecting the general cluster data for all nodes

To collect the diagnostic data for all of the cluster nodes, take the following steps on the node on which you installed the QMGTOOLS:

1. Go to the HASMNU: GO QMGTOOLS/HASMNU.
2. Select option **1. Collect and retrieve cluster data from multiple nodes**.
3. Specify the name of your library, and it will be created (Figure 15-19). Press Enter.

```
(DMPCLU)

Type choices, press Enter.

Library to store data  . . . . .   LIB4HADTA     Library Name
Dump cluster trace only  . . . .   N             Y, N

                                                                      Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 15-19 Options for data collection

[2] There is one menu option used by IBM Support (GO MG, opt.1, opt.4) that requires the 5799PTL to be installed.

4. Specify the user IDs and passwords for the nodes (Figure 15-20). If you are using the same user profile and password on all nodes, you can choose option F6 and specify the same user and password for all nodes (Figure 15-21 on page 441). Press F1 to continue.

Note: To successfully access all nodes, collect the data and get it on single node. User profiles should be able to FTP successfully to and from every node. In case you cannot use FTP among your nodes, you need to collect the data on each node separately, as described in “Collecting the general data on a single node” on page 443.

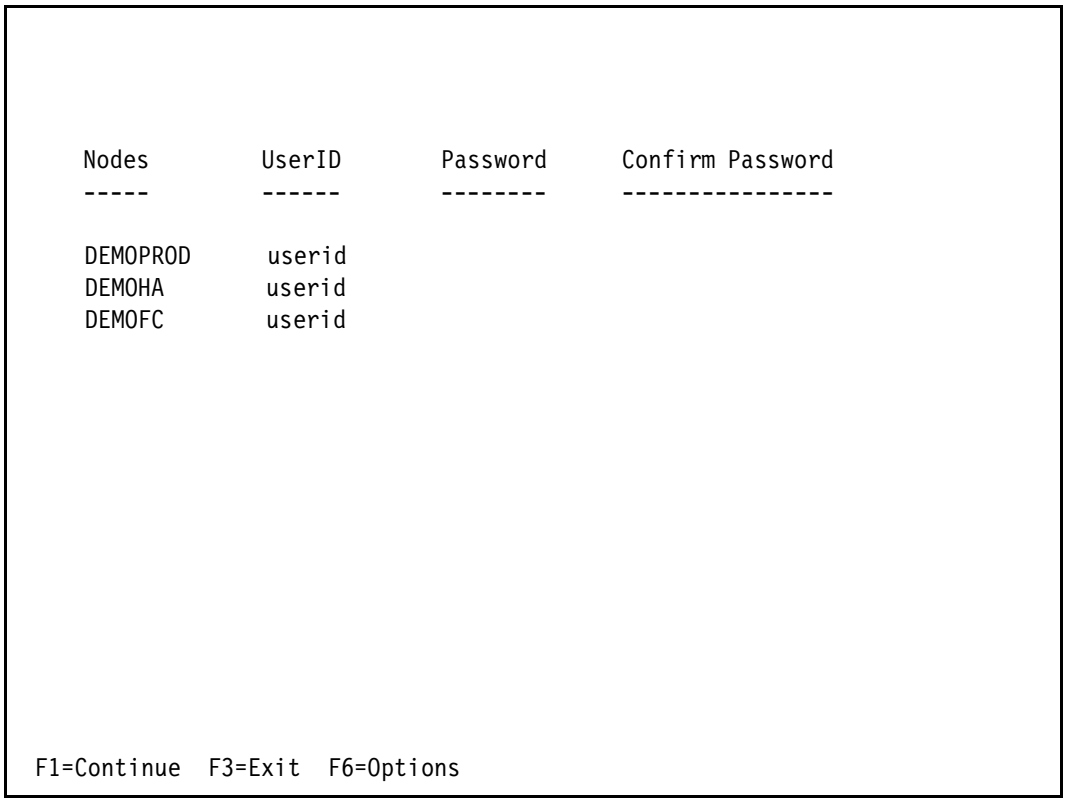


Figure 15-20 Specify UID and PWD for the nodes

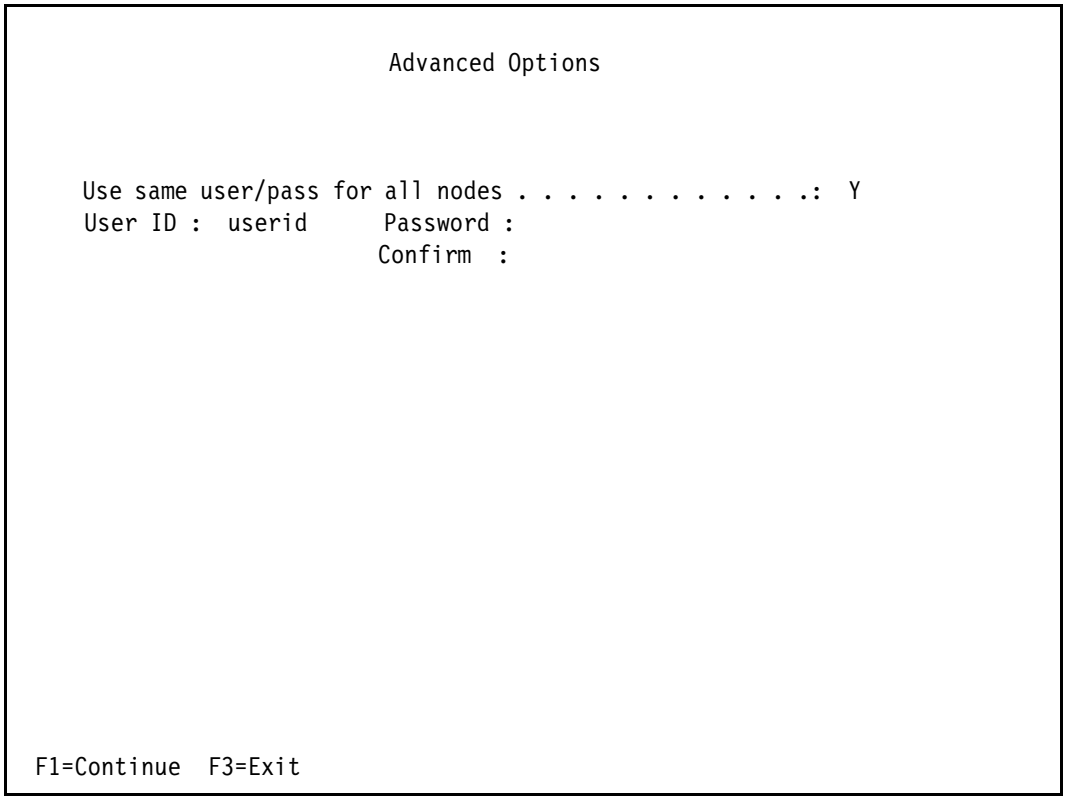


Figure 15-21 Specify same user and password for all nodes (optional)

5. Wait for the collection to complete (Figure 15-22), the press F1.

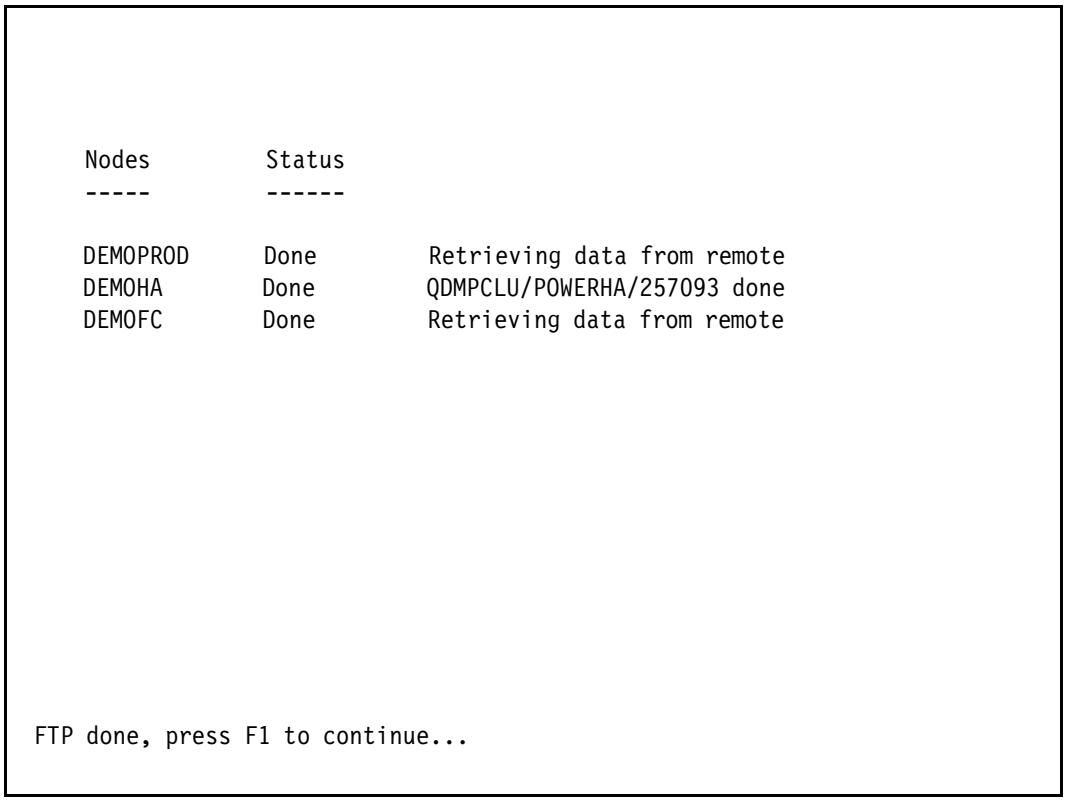


Figure 15-22 Data collection completed

6. Send the data to IBM. SAVF is placed in the library given in step 4, and the name of SAVF is in the message at the bottom of the page (Figure 15-23).

```
HASMNU                          QHASTOOLS menu

Select one of the following:

     1. Collect and retrieve cluster data from multiple nodes
     2. Dump cluster data on local node only
     3.
     4. Cluster Debug Tool (Internal IBM only)
     5. Alternative Debug Tool
     6.
     7.
     8. Collect SBG (Solution Based GUI) data
     9.
    10.
    11.
    12.


Selection or command
===>

F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel
F13=Information Assistant  F16=System main menu
Data saved into save file CLUDOCS001 in library LIB4HAMNU
```

Figure 15-23 Information about the SAVF with collected data

Collecting the general data on a single node

In case you cannot collect data using the QMGTOOLS on a single node, you need to install QMGTOOLS on each node and run the collection on each node. To do this use the following steps on each node.

1. Add the QMGTOOLS library to your library list:

   ```
   ADDLIBLE LIB(QMGTOOLS)
   ```

2. Go to HASMNU: GO MG and select option 1 (or **GO HASMNU**).

3. Select option 2 (Dump cluster data on local node only) and input the correct parameters for data collection (a library will be created) (Figure 15-24).

```
(DMPCLUINF)

Type choices, press Enter.

Cluster Name . . . . . . . . . . > PWRHA_CLU     Character value
Local node name  . . . . . . . . > DEMOHA        Character value
Save into save file? . . . . . .   Y             Y, N
Library to store data  . . . . .   LIB4HALOC     Character value

                                                                      Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 15-24 Parameters for local data collection

4. Wait for the collection to complete. You will get the page shows in Figure 15-25. Press Enter and send the SAVF that was created to IBM.

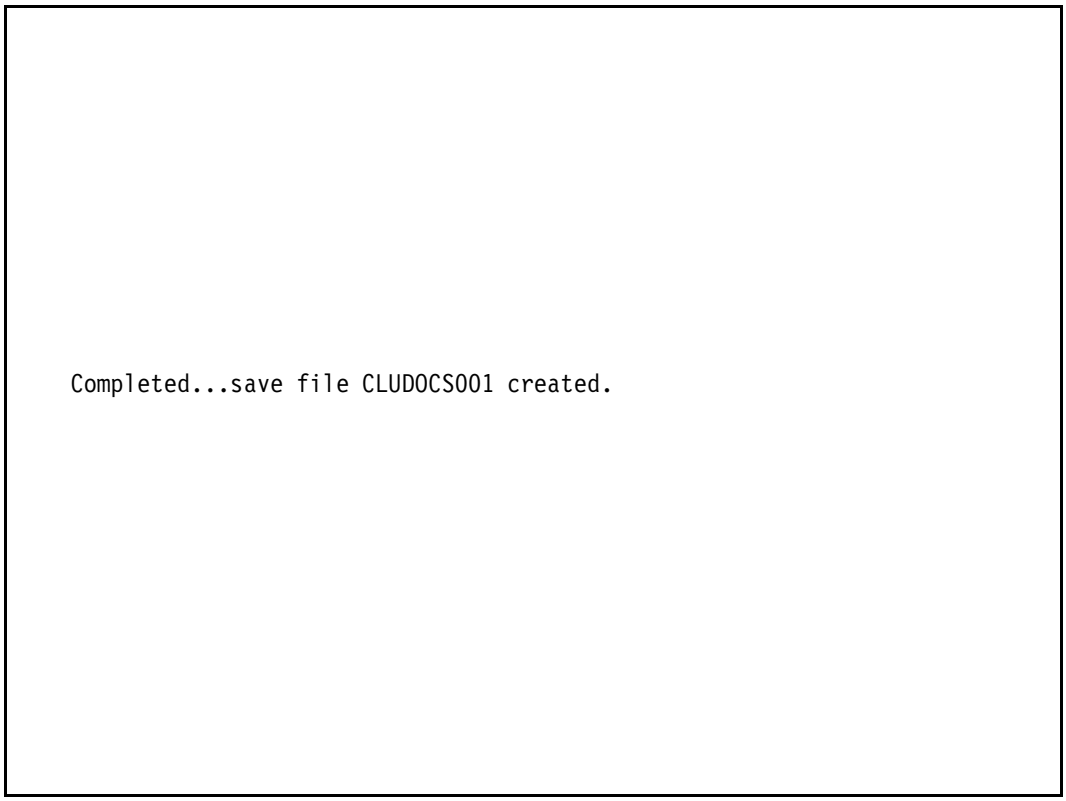


Figure 15-25 Panel with SAVF name to be sent to IBM

Collecting the data for the PowerHA GUI with QMGTOOLS

To collect the diagnostic data for the problems that occurred when you were using the PowerHA GUI, you need to have QMGTOOLS installed on the system on which you are using the IBM Systems Director Navigator for System i.

To collect the data, take these steps:

1. Go to HASMNU: GO QMGTOOLS/HASMNU.
2. Select option 8. Collect SBG (Solution Based GUI) data.
3. Wait for the collection to complete (Figure 15-26) and send the data in the message to IBM.

```
HASMNU                          QHASTOOLS menu

Select one of the following:

     1. Collect and retrieve cluster data from local/remote nodes
     2. Dump cluster data on local node only
     3.
     4. Cluster Debug Tool
     5. Alternative Debug Tool
     6.
     7.
     8. Collect SBG (Solution Based GUI) data
     9.
    10.
    11.
    12.


Selection or command
===>

F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel
F13=Information Assistant  F16=System main menu
Completed, file is in IFS directory /tmp/hasmlogs1003111627.zip
```

Figure 15-26 Send /tmp/hasmlogs1003111627.zip to support

Reviewing the data collected by QMGTOOLS

As a result of your data collection, there is a library created and physical files put into it. There are two files for each node. You can review the content of the files with the available system tool, such as PDM, or with **DSPPFM** or **EDTF**.

The file content might vary due to the development of QMPGTOOLS and your configuration.

Using the Alternative Debug Tool option

In the HASMNU, there is an option that you can use to diagnose certain kinds of problems in your cluster. The types of problems that it can diagnose vary depending on the QMGTOOLS release, so we recommend that you have the latest version of the tool. To use it, enter **GO HASMN**U and select option 5. Alternative Debug Tool. You will be presented a page similar to the one shown in Figure 15-27. Input the name of the library containing the data that you collected previously.

```
Cluster Debug (CLUDBGALT)

Type choices, press Enter.

Library containing dumps . . . .   LIB4HAMNU     Library Name

                                                                    Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 15-27 Alternative debug tool options panel

As a result of this analysis, there will be a spoolfile created in your interactive job. In the file you can find the problems that the tool has found and additional information about what to collect to debug the problem further, if necessary.

Development of the Must Gather Data Collector

As it was stated previously, QMGTOOLS is being continuously developed and new functions are being added. Therefore, we advise that you have the most current version of the tool to use the new features of this tool.

If you have any suggestions for improvements that might help us improve this tool, you can contact Benjamin Rabe (brabe@us.ibm.com) or the authors of this publication to have your suggestions forwarded to the appropriate development team.

15.12.4 PowerVM Virtual I/O Server

If you have a problem with storage hosted by Virtual I/O Server, check the following items:

1. The VIOS LPAR operability and configuration. Go to HMC and check the VIOS LPAR properties for the status and the allocated resources.
2. You can log in to the VIOS LPAR and check the mapping of the devices with **`lsmap -all`**.
3. You can use **`diagmenu`** to diagnose problems in your VIOS partition.

   More information about Virtual I/O Server problem solving and management can be found in the Power Systems Information Center at:

   `http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp`
4. Collect the snap data. When you sign on to the VIOS LPAR use **`snap`** with no parameters. When it ends there will be file.

15.12.5 DS8000 Copy Services

When using DS8000 for your IASP storage and PowerHA for System i, you can log onto the TotalStorage device and check the status of the replication using the DSCLI commands:

1. The Metro Mirror status of a volume can be checked with **`lspprc`** (Figure 15-28).

```
dscli> lspprc -fmt default 6000-6002 6100-6102
Date/Time: October 3, 2011 4:44:47 PM CEST IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY031
ID        State       Reason Type         SourceLSS Timeout (secs) Critical Mode First Pass Status
=============================================================================================
6000:6000 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6001:6001 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6002:6002 Full Duplex -      Metro Mirror 60        60             Disabled      Invalid
6100:6100 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6101:6101 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
6102:6102 Full Duplex -      Metro Mirror 61        60             Disabled      Invalid
```

Figure 15-28 lspprc for volumes used in Metro Mirror

2. The Global Mirror status for the volumes can be checked with **lsgmir** (Figure 15-29).

```
dscli> lsgmir 0
Date/Time: October 6, 2011 4:55:24 PM CEST IBM DSCLI Version: 7.6.10.530 DS:
IBM.1750-13ABGAA
SessionID MasterID         ID State   %Success CGtime
==========================================================================
0x02      IBM.1750-13ABGAA 02 Running 100      10/06/2011 15:32:12 CEST
dscli> showgmir 0
Date/Time: October 6, 2011 4:55:26 PM CEST IBM DSCLI Version: 7.6.10.530 DS:
IBM.1750-13ABGAA
ID                           IBM.1750-13ABGAA/02
Master Count                 1
Master Session ID            0x02
Copy State                   Running
Fatal Reason                 Not Fatal
CG Interval Time (seconds)   0
Coord. Time (milliseconds)   50
Max CG Drain Time (seconds)  30
Current Time                 10/06/2011 15:32:15 CEST
CG Time                      10/06/2011 15:32:14 CEST
Successful CG Percentage     100
FlashCopy Sequence Number    0x4E8DADDE
Master ID                    IBM.1750-13ABGAA
Subordinate Count            0
Master/Subordinate Assoc     -
```

Figure 15-29 lsgmir for Global Mirror

3. The FlashCopy status for the copied volumes can be checked with **lsflash** (Figure 15-30).

```
dscli> lsflash -fmt default a010-a013
Date/Time: October 3, 2011 4:12:40 PM CEST IBM DSCLI Version: 7.6.10.530 DS: IBM.2107-75AY032
ID        SrcLSS SequenceNum Timeout ActiveCopy Recording Persistent Revertible SourceWriteEnabled TargetWriteEnabled BackgroundCopy
=============================================================================================================================
A010:A020 A0     0           60      Disabled   Enabled   Enabled    Disabled   Enabled            Enabled            Enabled
A011:A021 A0     0           60      Disabled   Enabled   Enabled    Disabled   Enabled            Enabled            Enabled
A012:A022 A0     0           60      Disabled   Enabled   Enabled    Disabled   Enabled            Enabled            Enabled
A013:A023 A0     0           60      Disabled   Enabled   Enabled    Disabled   Enabled            Enabled            Enabled
```

Figure 15-30 lsflash for volumes in FlashCopy

If you need additional help with the command you can use the **-help** option with any of them. The list of available commands is shown in DSCLI when you issue **help**.

For more information about DS8000 see this website:

http://publib.boulder.ibm.com/infocenter/dsichelp/ds8000ic/index.jsp

15.12.6 SVC/V7000 Copy Services

To collect problem data when using the SVC, collect the information for Virtual I/O Server (15.12.4, “PowerVM Virtual I/O Server” on page 448). Also collect this information:

1. `DSPSVCSSN SSN(session_name) OUTPUT(*PRINT)`.
2. Collect the XSM logs from `/QIBM/UserData/HASM/hads`.

A

IBM i data resilience options

Whether you need continuous availability for your business applications or are looking to reduce the amount of time that it takes to perform daily backups, IBM i high-availability technologies provide the infrastructure and tools to help achieve your goals.

Modern IBM i high-availability solutions are built on IBM i cluster resource services, or more simply clusters. A cluster is a collection or group of multiple systems that work together as a single system from an application perspective. Clusters provide the underlying infrastructure that allows resilient resources, such as data, devices, and applications, to be automatically or manually switched between systems or nodes (partitions on a single frame or on multiple frames). It provides failure detection and response, so that in the event of an outage, cluster resource service responds accordingly, keeping your data safe and your business operational. PowerHA SystemMirror for i offers complete end-to-end integrated solutions for high availability and disaster recovery, with special focus on the application availability through planned or unplanned outage events.

This appendix discusses two other options for providing data resiliency:

- Full-system storage-based Copy Services solutions
- Logical replication solutions

We also provide a comparison of the resiliency solutions available for IBM i:

- PowerHA SystemMirror for i
- Full-system storage-based Copy Services
- Logical replication

IBM i full-system storage-based Copy Services solutions

The introduction of IBM i boot from SAN further expanded IBM i availability options by exploiting solutions such as FlashCopy, provided through IBM System Storage Copy Services functions. With boot from SAN introduced with i5/OS V5R3M0, you no longer needed to use remote load source mirroring to mirror your internal load source to a SAN-attached load source. Instead, the load source can be placed directly inside a SAN-attached storage subsystem and with IBM i 6.1 provide multi-path attachment to the external load source for redundancy.

Boot from SAN makes it easier to bring up a system environment that has been copied using Copy Services functions such as FlashCopy or remote Copy Services. During the restart of a cloned environment, you no longer have to perform the *Recover Remote Load Source Disk Unit* through Dedicated Service Tools (DSTs), thus reducing the time and overall steps required to bring up a point-in-time system image after FlashCopy or remote Copy Services functions have been completed.

Cloning IBM i

Cloning has been a concept for the IBM i platform since the introduction of boot from SAN with i5/OS V5R3M5. Previously, to create a new system image, you had to perform a full installation of the SLIC and IBM i. When cloning IBM i, you create an exact copy of the existing IBM i system or partition. The copy can be attached to another System i/Power Systems models, a separate LPAR, or, if the production system is powered off, the existing partition or system. After the copy is created, you can use it for offline backup, system testing, or migration.

Boot from SAN enables you to take advantage of some of the advanced features that are available with IBM system storage. One of these functions is FlashCopy. It allows you to perform a point-in-time instantaneous copy of the data held on a LUN or group of LUNs. Therefore, when you have a system that only has SAN LUNs with no internal drives, you can create a clone of your system.

Important: When we refer to a clone, we are referring to a copy of a system that only uses SAN LUNs. Therefore, boot from SAN is a prerequisite for this.

Full system FlashCopy

FlashCopy not only allows you to take a system image for cloning but is also an ideal solution for increasing the availability of your IBM i production system by reducing the time for your system backups (Figure A-1).

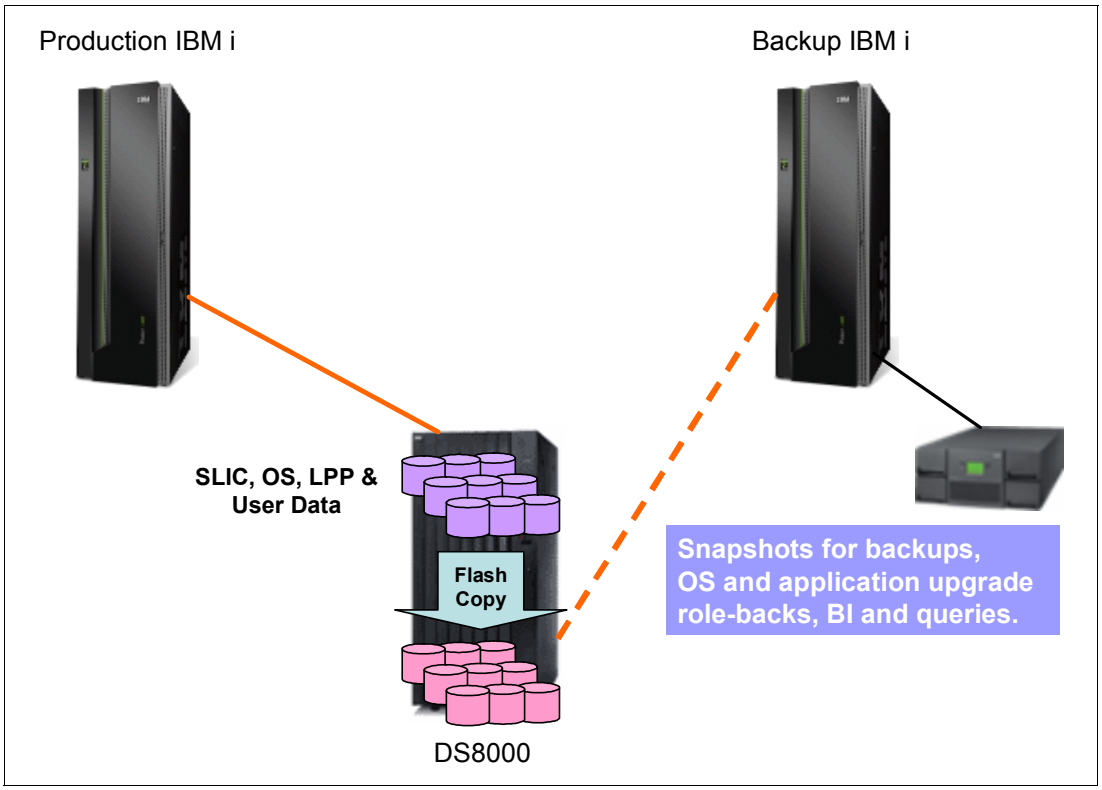


Figure A-1 Full system FlashCopy

To obtain a full system backup of IBM i with FlashCopy, either a system shutdown or, as of IBM i 6.1, a quiesce function is supplied that flushes modified data from memory to disk. FlashCopy only copies the data on the disk. The IBM i quiesce for Copy Services function (CHGASPACT) introduced with 6.1 allows you to suspend all database I/O activity for *SYSBAS and IASP devices before taking a FlashCopy system image, eliminating the requirement to power down your system.

Full system replication by Metro Mirror or Global Mirror

Metro Mirror offers synchronous replication between two DS models or between a DS and ESS model 800. In Figure 3-9, two IBM i servers are separated by distance to achieve a disaster recovery solution at the second site. This is a fairly simple arrangement to implement and manage. Synchronous replication is desirable because it ensures the integrity of the I/O traffic between the two storage complexes and provides you with a recovery point objective (RPO) of zero (that is, no transaction gets lost). The data on the second DS system is not available to the second IBM i system while Metro Mirror replication is active (that is, it must be powered off).

The main consideration with this solution is distance. The solution is limited by the distance between the two sites. Synchronous replication needs sufficient bandwidth to prevent latency in the I/O between the two sites. I/O latency can cause application performance problems.

Testing is necessary to ensure that this solution is viable depending, on a particular application's design and business throughput.

With Global Mirror all the data on the production system is asynchronously transmitted to the remote DS models. Asynchronous replication via Global Copy alone does not guarantee the order of the writes, and the remote production copy will lose consistency quickly. To guarantee data consistency, Global Mirror creates consistency groups at regular intervals, by default, as fast as the environment and the available bandwidth allows. FlashCopy is used at the remote site to save these consistency groups to ensure that a consistent set of data is available at the remote site, which is only a few seconds behind the production site. That is, when using Global Mirror a RPO of only a few seconds can be achieved normally without any performance impact to the production site.

Figure A-2 shows an overview of a full system remote copy solution by using IBM system storage.

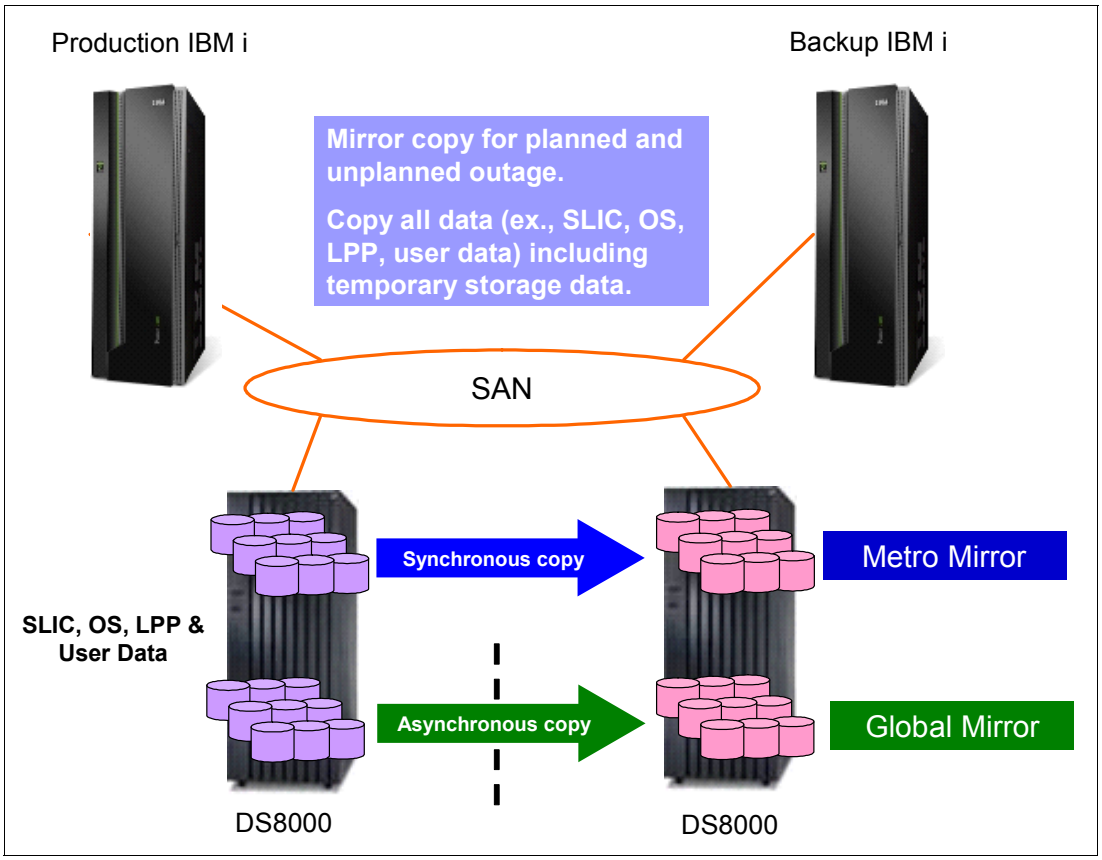


Figure A-2 Full system remote Copy Services

This is an attractive solution because of the extreme distances that can be achieved with Global Mirror. However, it requires a proper sizing of the replication link bandwidth to ensure that the RPO targets can be achieved. Testing should be performed to ensure that the resulting image is usable.

When you recover in the event of a failure, the IPL of your recovery system will always be an abnormal IPL of IBM i on the remote site.

Note: Using IBM i journaling along with Metro Mirror or Global Mirror replication solutions is highly recommended to ensure transaction consistency and faster recovery.

Logical replication solutions

Logical replication is currently the most widely deployed multisystem data resiliency topology for high availability (HA) in the IBM i world. It is deployed by using IBM iCluster or a High Availability Business Partner (HABP) solution package. Replication is executed (via software methods) on objects. Changes to the objects (for example, file, member, data area, or program) are replicated to a backup copy. The replication process is near real time. Typically, if the object, such as a file, is journaled, replication is handled at a record level. For such objects as user spaces that are not journaled, replication is handled at the object level. In this case, the entire object is replicated after each set of changes to the object is complete.

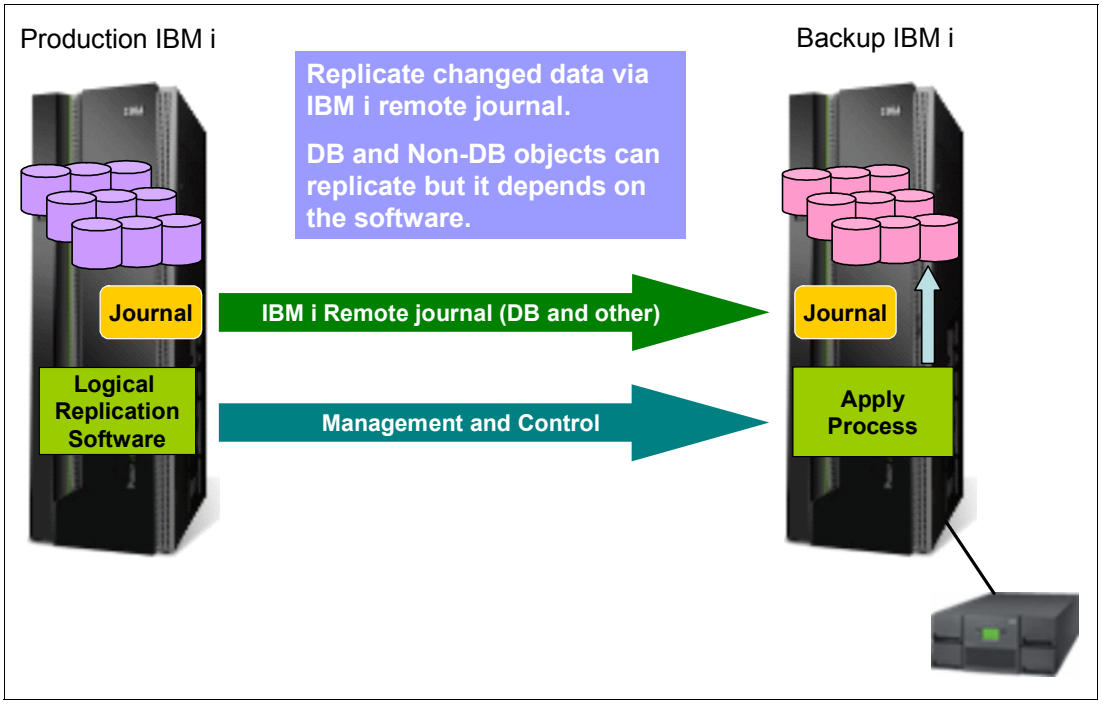


Figure A-3 Logical replication

Most logical replication solutions allow for additional features beyond object replication. For example, you can achieve additional auditing capabilities, observe the replication status in real time, automatically add newly created objects to those being replicated, and replicate a subset of objects in a given library or directory.

To build an efficient and reliable multi-system HA solution using logical replication, synchronous remote journaling as a transport mechanism is preferable. With remote journaling, IBM i continuously moves the newly arriving database data in the journal receiver to the backup server journal receiver. At this point, a software solution is employed to “replay” these journal updates, placing them into the object or replacing the object on the backup server. After this environment is established, there are two separate yet identical objects, one on the primary server and one on the backup server.

With this solution in place, you can activate your production environment on the backup server via a role-swap operation.

One characteristic of this solution category is that the backup database file is “live”. That is, it can be accessed in real time for backup operations or for other read-only application types, such as building reports.

Another benefit of this type of solution is that different releases of IBM i can be utilized on the primary and backup nodes. This means that the solution can be used to assist in migrations to new operating system levels. As an example, you can be replicating from IBM i 6.1 on the production system to IBM i 7.1 on the backup system, and when you want to migrate the production system to IBM i 7.1 you can perform a roleswap to the backup, perform your OS upgrade on production, and when the migration has been validated and tested, roleswap your users back to production.

The challenge with this solution category is the complexity that can be involved with setting up and maintaining the environment. One of the fundamental challenges lies in not strictly policing undisciplined modification of the live copies of objects residing on the backup server. Failure to properly enforce such a discipline can lead to instances in which users and programmers make changes against the live copy so that it no longer matches the production copy. Should this happen, the primary and backup versions of your files are no longer identical. There are new tools provided by the software replication providers that perform periodic data validation to help detect this situation.

Another challenge associated with this approach is that objects that are not journaled must go through a checkpoint, be saved, and then be sent separately to the backup server. Therefore, the granularity of the real-time nature of the process might be limited to the granularity of the largest object being replicated for a given operation.

For example, a program updates a record residing within a journaled file. As part of the same operation, it also updates an object, such as a user space, that is not journaled. The backup copy becomes completely consistent when the user space is entirely replicated to the backup system. Practically speaking, if the primary system fails, and the user space object is not yet fully replicated, a manual recovery process is required to reconcile the state of the non-journaled user space to match the last valid operation whose data was completely replicated. This is one of the reasons why the application state and the state of the replicated data at the target box are inherently asynchronous.

Another possible challenge associated with this approach lies in the latency of the replication process. This refers to the amount of lag time between the time at which changes are made on the source system and the time at which those changes become available on the backup system. Synchronous remote journal can mitigate this for the database. Regardless of the transmission mechanism used, you must adequately project your transmission volume and size your communication lines and speeds properly to help ensure that your environment can manage replication volumes when they reach their peak. In a high-volume environment, replay backlog and latency might be an issue on the target side even if your transmission facilities are properly sized. Another issue can arise when the apply process on the target system cannot keep up with the incoming data. The target system cannot be used until this lag data is fully applied.

Comparison characteristics

In this section, we selected major characteristics to consider. However, you might have other characteristics that are equally or more important to your environment. To compare the various availability techniques that use some form of data resiliency, we use the characteristics in the technology comparison shown in Table A-1 and Table A-2.

Table A-1 Comparison with data resiliency options (technology)

	PowerHA SystemMIrror for i	**Full-system storage-based Copy Services**	**Logical replication**
Base technology	IBM i Storage management	Storage	Replication software
Cluster resource management	Yes	No	No
Multi-site cluster resource management	Yes	No	No
Multi-site data replication service	Yes	Yes	Yes
Synchronous to application state	Yes	Yes	No
Environment resiliency	Cluster Admin Domain	No	Depends on software
Integrated heartbeat	Yes	No	No
Integrated flash copy	Yes	Yes	N/A
Automated/user control fail over	Yes	No	Depends on software and applications
Internal disk support	Yes	No	Yes
Switched disk-based resiliency	Yes	No	No
Host-based replication	Geographic Mirroring	No	Remote journal
Storage-based replication	Metro Mirror Global Mirror	Metro Mirror Global Mirror	No

Table A-2 Comparison with data resiliency options (IT operating environment)

	PowerHA SystemMIrror for i	**Full-system storage-based Copy Services**	**Logical replication**
IT operation skill base	IBM i/Storage	Storage	IBM i/Replication software
Technical skill support base	IBM	IBM	IBM/HABP
Data synchronization	Disk mirroring	Disk mirroring	Journal replay
Required bandwidth	Data change in IASP	Data change in full system	Journal changes

	PowerHA SystemMIrror for i	Full-system storage-based Copy Services	Logical replication
Save window	FlashCopy	FlashCopy	Save while active
Primary/secondary storage synchronously mirrored	Yes	Yes	No
Use of journal	Yes, best practice	Yes, best practice	Source of replicated data
Application environment change management for replication services	Not required	Not required	Required, after changing or adding applications, setting up replication
Object type support	IASP data	All	Most
IPL is required when switchover/failover	No	Yes	No
Recovery point objective	Last data written to IASP	Last data written to Storage	Depends on transaction boundary
Recovery time objective	IASP vary on time	Abnormal IPL time	Time of applying lag plus switchover overhead plus sync check

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *Striving for Optimal Journal Performance on DB2 Universal Database for iSeries*, SG24-6286
- *Journaling - User ASPs Versus the System ASP*, TIPS0602
- *Journaling at object creation on DB2 for iSeries*, TIPS0604
- *Journaling: Why Is My Logical File Journaled?*, TIPS0677
- *Journaling - How Can It Contribute to Disk Usage Skew?*, TIPS0603
- *Journaling · Journal Receiver Diet Tip 1: Eliminating Open and Close Journal Entries*, TIPS0607
- *Journaling - *RMVINTENT: The preferred fork in the road for heavy journal traffic*, TIPS0605
- *Journaling · Journal Receiver Diet tip 2: Consider using skinny headers*, TIPS0654

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

These websites are also relevant as further information sources:

- IBM i Information Center

  http://publib.boulder.ibm.com/infocenter/systems/scope/i5os/index.jsp
- IBM PowerHA

  http://www-03.ibm.com/systems/power/software/availability/i5os.html
- IBM iCluster for i

  http://www-03.ibm.com/systems/power/software/availability/i5os.html#icluster

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

R

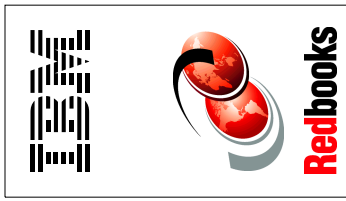
S

T

U

V

W



PowerHA SystemMirror for IBM i Cookbook

PowerHA SystemMirror for IBM i Cookbook

Take advantage of PowerHA to configure and manage high availability

Find guidance on planning and implementing PowerHA

Benefit from the latest PowerHA solution enhancements

IBM PowerHA SystemMirror for i is the IBM high-availability disk-based clustering solution for the IBM i 7.1 operating system. When combined with IBM i clustering technology, PowerHA for i delivers a complete high-availability and disaster-recovery solution for your business applications running in the IBM System i environment. PowerHA for i enables you to support high-availability capabilities with either native disk storage or IBM DS8000 or DS6000 storage servers or IBM Storwize V7000 and SAN Volume Controllers.

The latest release of IBM PowerHA SystemMirror for i delivers a brand-new web-based PowerHA graphical user interface that effectively combines the solution-based and task-based activities for your HA environment, all in a single user interface.

This IBM Redbooks® publication provides a broad understanding of PowerHA for i. This book is intended for all IBM i professionals who are planning on implementing a PowerHA solution on IBM i.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-7994-00 ISBN 0738436364