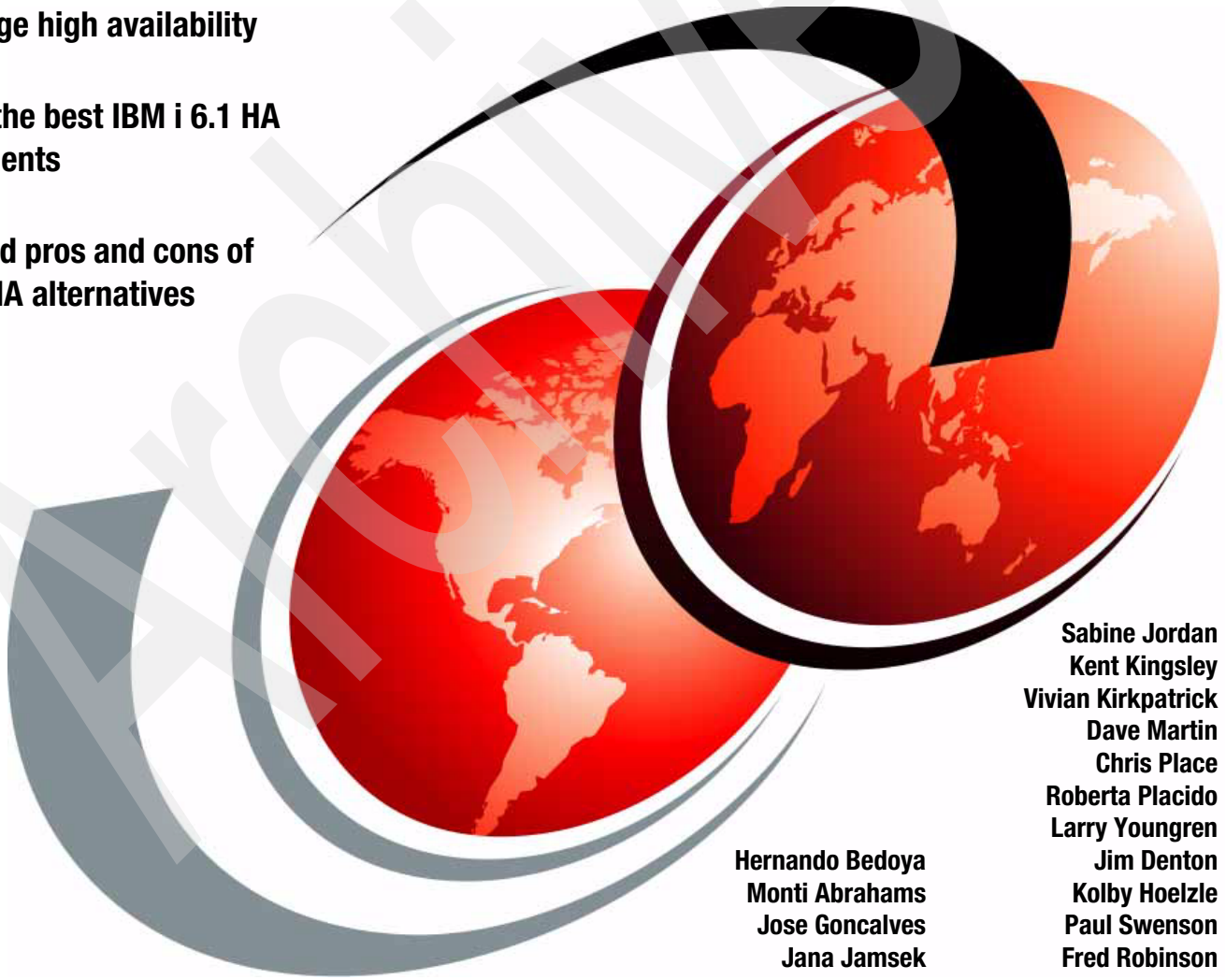
IBM

Implementing PowerHA for IBM i

- Embrace PowerHA for i to configure and manage high availability
- Leverage the best IBM i 6.1 HA enhancements
- Understand pros and cons of different HA alternatives



Sabine Jordan
Kent Kingsley
Vivian Kirkpatrick
Dave Martin
Chris Place
Roberta Placido
Larry Youngren
Jim Denton
Kolby Hoelzle
Paul Swenson
Fred Robinson
Hernando Bedoya
Monti Abrahams
Jose Goncalves
Jana Jamsek

Redbooks

ibm.com/redbooks

International Technical Support Organization

Implementing PowerHA for IBM i

November 2008

SG24-7405-00

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

First Edition (November 2008)

This edition applies to IBM i 6.1.

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 about 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 Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites 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:

AIX®
Alerts®
AS/400®
Cross-Site®
DB2 Universal Database™
DB2®
DS6000™
DS8000™
eServer™
FlashCopy®
i5/OS®
IBM®
iCluster®
Integrated Language Environment®
iSeries®
Language Environment®
MVS™
Operating System/400®
OS/400®
POWER™
Power Systems™
POWER6™
PowerHA™
Redbooks®
Redbooks (logo) ®
System i®
System Storage™
System/36™
System/38™
TotalStorage®
Virtualization Engine™
WebSphere®

The following terms are trademarks of other companies:

Oracle, JD Edwards, PeopleSoft, Siebel, and TopLink are registered trademarks of Oracle Corporation and/or its affiliates.

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

Java, JDK, and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

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

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.

Preface

IBM® PowerHA™ for i (formerly known as HASM) is the IBM high availability disk-based clustering solution for the IBM i 6.1 operating system. PowerHA for i when combined with IBM i clustering technology 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.

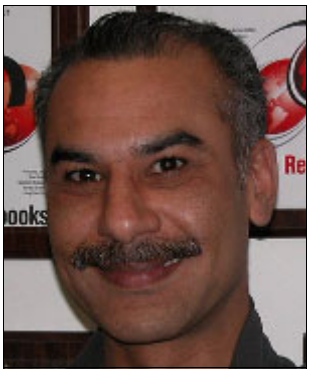
This IBM Redbooks® publication gives a broad understanding of PowerHA for i. This book is divided in four major parts:

- Part 1, “Introduction and background” on page 1, provides a general introduction to clustering technology and some background.
- Part 2, “PowerHA for i setup and user interfaces” on page 69, describes and explains the different interfaces that PowerHA for i has. It also describes the migration process to this product and some sizing guidelines.
- Part 3, “Implementation examples using PowerHA for i” on page 319, explains how to use PowerHA for i with three major ERP solutions, such as SAP®, Lawson M3, and Oracle® JD Edwards®.
- Part 4, “Other IBM i 6.1 high availability enhancements” on page 349, explains additional IBM i 6.1 announced enhancements in high availability.

The team that wrote this book

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

Hernando Bedoya is an IT Specialist at the IBM ITSO, in Rochester, Minnesota. He writes extensively and teaches IBM classes worldwide in all areas of DB2® for i5/OS®. Before joining the ITSO more than seven years ago, he worked for IBM Colombia as an IBM AS/400® IT Specialist doing parsleys support for the Andean countries. He has 24 years of experience in the computing field and has taught database classes at Colombian universities. He holds a master’s degree in computer science from EAFIT, Colombia. His areas of expertise are database technology, application development, and high availability.



Monti Abrahams is a Senior IT Specialist for IBM South Africa, based in Cape Town. He is an IBM Certified Solutions Expert: DB2 Content Manager On Demand and a Certified SAP Technical Consultant with more than 15 years of experience in Operating System/400® (OS/400®) and DB2 Universal Database™ for iSeries®. Monti provides technical support and consulting to IBM System i clients throughout South Africa and Namibia. He also conducts training courses for IBM IT Education Services. Monti has co-authored several IBM Redbooks publications on various iSeries topics.

Jose Goncalves is a Senior Certified IT Specialist for System i of the European STG-Lab Services team in La Gaude, France. For the past seven years he has helped EMEA clients modernize legacy RPG applications and System i High Availability design and implementation. He has 25 years of experience. He has his master's degree in IT computer science from University PARIS VI, and in 1982 he joined IBM as Customer Engineer (MVS™). Since 1992, he has worked on the AS/400 platform, mainly in application development (ICMS – IBM Telecommunication Billing and Customer Care software) and OS/400 System management and support.

Jana Jamsek is an IT Specialist in IBM Slovenia. She works in Storage Advanced Technical Support for Europe as a Specialist for IBM Storage systems and i5/OS systems. Jana has eight years of experience in the System i and AS/400 area and six years of experience in storage. She holds a master's degree in computer science and a degree in mathematics from the University of Ljubljana, Slovenia. She was an author of the IBM Redpapers publication, *The LTO Ultrium Primer for IBM eServer iSeries Customers*, REDP-3580, and the IBM Redbooks publications *iSeries in Storage Area Networks*, *iSeries and IBM TotalStorageIBM i and IBM System Storage: A Guide to Implementing External Disks on IBM i*, SG24-7120; *IBM Virtualization EngineThe IBM Virtualization Engine TS7510: Getting Started with i5/OS and Backup Recovery and Media Services*, SG24-7510; *Implementing IBM Tape in i5/OS*, SG24-7440; *IBM System Storage DS8000: Copy Services in Open Environments*, SG24-6788; and *IBM System Storage DS6000 Series: Architecture and Implementation*, SG24-6781.

Sabine Jordan is a Consulting IT Specialist working in IBM Germany. She has worked on several XSM implementations for both SAP and non-SAP environments. Among these implementations, she has created concepts for the XSM design and implemented the entire project (cluster setup, application changes), as well as performed customer education and testing. In addition, Sabine presents and delivers workshops (internal and external) on XSM, as well as high availability and disaster recovery.

Kent Kingsley is a Senior Software Engineer for Vision Solutions. He works on design and development of the MIMIX Cluster1 product and also works on design and implementation of customer cluster solutions. He also works with IBM in the cluster lab to test both IBM and MIMIX solutions and do proof-of-concept testing for potential customers.

Vivian Kirkpatrick is a Staff Software Engineer in the Rochester Support Center located in the United States. He has a degree in Computer Information Science from Minnesota State University, Mankato. He worked for three years in database support before moving to the SLIC/VMC/Internals/PHYP area, where he worked for six years. In 2007 a new high-availability team was created to focus on cluster-related issues. Vivian has developed and taught multiple courses for BIM Rochester employees and worldwide about support on Cluster, independent auxiliary storage pool (iASP), and geographic mirroring.

Dave Martin is a Certified IT Specialist in Advanced Technical Support (ATS). He started his IBM career in 1969 in St. Louis, Missouri, as a Systems Engineer. From 1979 to 1984, he taught System/34 and System/38™ implementation and programming classes to customers and IBM representatives. In 1984 he moved to Dallas, Texas, where he has provided technical support for S/36, S/38, AS/400, iSeries, and System i in the areas of languages, operating systems, systems management, DB2 for i5/OS, and Business Intelligence. For the last four years his focus has been primarily on System i high availability considerations and solutions.

Chris Place is an IT Specialist in the Americas System i Technical Sales Support Organization in Rochester, Minnesota, focusing on HA and DASD performance. He has worked at IBM for 14 years. Chris has over five years of experience analyzing various aspects of System i performance. Prior to that, Chris worked at SAP Germany on the porting team in software logitics. Chris' areas of expertise include LIC, LPAR performance, and DASD performance. He has written extensively on LPAR performance. He works with customers directly in the areas of I/O hardware performance. He presents at various user groups and the IBM System i Technical Conference.

Roberta Placido joined IBM in December 1990 and started working as a software CE in the iSeries support center in Milan. Her responsibilities involve onsite services and advanced technical training on LPAR technologies, HA solutions, and clustering/iASP. She worked on the back end for two years for the internal team placed in Rochester supporting countries world wide. She deliver HA solution services, and last year she became a member of the HA team of the Rochester Support Center.

Larry Youngren after more than 30 years of experience leading the design efforts for System i database and journal support at IBM, Larry Youngren recently retired from IBM and now lectures, writes, and consults on high-availability issues. He also coordinates the efforts of other recently retired System i professionals who help out at colleges and at annual COMMON conferences as mentors. For 30 years he served as a Microcode Designer for the lower layers of the i5/ OS operating system and frequently consulted with customers regarding High Availability and Journal performance issues. During his IBM career, he worked exclusively with microcode, first for the S/38 and then for the System i. Larry helped the teams responsible for Data Base, Commit, SMAPP, and Journal. His interests involve future performance and recovery improvements affecting journaling and IPL duration. He and the team he led have authored a popular IBM Redbooks publication entitled *Striving for Optimal Journal Performance on DB2 Universal Database for iSeries*, SG24-6286, numerous magazine articles regarding both Ragged Save While Active and Remote Journaling, in addition to over a dozen journal-related IBM Technotes that address popular journal questions. These technotes can be accessed from the IBM Redbooks publication Web site.

Jim Denton is a Senior Software Engineer at IBM Rochester where he has been employed since 1981. He has held a variety of positions in operating system development and performance analysis on the S/38, AS400, System i, and Power Systems™. His recent experience includes five years as a DB2 Specialist and a three-year assignment to the IBM Benchmark and Briefing Center in Montpellier France. Since 2005, Jim has been a member of the ERP development team with the mission of working closely with EnterpriseOne and World developers, producing performance benchmarks, and providing performance and tuning guidance for EnterpriseOne on i.

Kolby Hoelzle is an Advisory Software Engineer working at the IBM i development lab in Rochester, Minnesota. He is currently a member of the SAP on i Development Team working closely with SAP development. He joined IBM in 1999 and has over seven years of experience with SAP on the IBM i platform, including two years on assignment in Germany working at the SAP Development Lab in Walldorf, Germany. Kolby has extensive experience both implementing and developing iASP-based high-availability solutions for SAP applications. Kolby graduated from Utah State University in 1999.

Paul Swenson is a member of the IBM i ERP Development Team, which is part of the IBM i development lab in Rochester, Minnesota, and focuses solely on Lawson. He joined IBM in 2000 and has over four years of experience with Lawson applications on the IBM i platform. Prior to working in ERP Paul worked in IBM i Performance focusing on performance for Java™ and WebSphere® applications for four years. Paul graduated from the University of North Dakota in 2000.

Fred Robinson is a Consulting Educator with the IBM i Technology Center in Rochester, Minnesota. Fred joined IBM in 1978 as a Systems Engineer in Joplin, Missouri. In 1986, he moved to Rochester to work on the convergence of the System/36™ and System/38 into the AS/400. Fred has technical and practical skills in most areas of IBM i including Business Continuity, Storage, Virtualization, and Database. Fred has worked with many IBM Fortune 100 customers and small business customers alike to enhance their knowledge and understanding of both the IBM Systems family and IBM i. In addition to his work at the IBM i Technology Center, Fred is sought out by companies around the globe as a speaker and consultant.

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

Thomas Gray
Joanna Pohl-Miszczyk
James Hansen
International Technical Support Organization, Rochester Center

Jenny Dervin
Peg Levering
Vicki Morey
Curt Schemmel
Ron Peterson
Rick Dunsirn
James Lembke
Scott Helt
Lilo Bucknell
Amanda Fogarty
Raymond Bills
William Seurer

Selwyn Dickey
Tim Klubertanz
Daniel Degroff
Eric Hess
Fred Robinson
IBM Rochester

Special thanks to the IBM System i Benchmark Center for providing the required HW and SW for this project:
Dan Daley
Ken Wise
Jerry Evans
IBM System i Benchmark Center:

http://www-03.ibm.com/systems/services/benchmarkcenter/

Become a published author

Join us for a two- to six-week residency program! Help write a book dealing with specific products or solutions, while getting hands-on experience with leading-edge technologies. You will have the opportunity to team with IBM technical professionals, Business Partners, and Clients.

Your efforts will help increase product acceptance and customer satisfaction. As a bonus, you will develop a network of contacts in IBM development labs, and increase your productivity and marketability.

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 publications form found at:

  ibm.com/redbooks
- Send your comments in an e-mail 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

Part 1

Introduction and background

In this part we discuss high-availability foundations, clustering concepts, and the different high-availability technologies available on IBM i.

This part includes the following chapters:

1

Introduction: PowerHA for i

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 would eventually enable us to deliver a complete integrated IBM i solution. The announcements that we made with IBM i 6.1 represent a major part of achieving that goal. 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.

In this chapter we introduce the different solutions that IBM has to address business continuity on IBM i.

1.1 IBM i Business Continuity Solutions

It is estimated that there are over 800 customers big and small around the world that have deployed pre-6.1 independent auxiliary storage pool (iASP)-based clustering solutions. Some of our largest customers have moved to this capability using the Copy Services Tool Kit, which is a lab services offering that combines IBM i iASP and clustering technologies with IBM storage server technology. Customers utilizing a SAN storage strategy appreciate this approach because it addresses all of their availability needs with minimal day-to-day administrative activity while allowing them to fully capitalize on their SAN server investment. With 6.1 we announced the High Availability Solutions Manager (HASM) license program (LP), now called PowerHA for i, which enables IBM i-based management of an integrated storage cluster or a SAN-server-based cluster. We extended the Cross Site Mirroring (XSM) capability of IBM i to include both geographic mirroring (IBM i mirroring over IP) as well as IBM DS8000TM (and IBM DS6000TM) storage replication technologies, metro mirror and global mirror. Both geographic mirroring and metro mirror are synchronous replication solutions, which is the optimal approach for high-availability deployments. If you are looking for geographic dispersion for a disaster recovery solution, the IBM DS8000 global mirror replication solution may be used. Another disaster recovery approach is iCluster®, the IBM software replication solution for HA and disaster recovery (DR). iCluster is one of the well-known logical replication solutions that is well suited for geographic dispersion implementations as well as for data replication/recovery operations. With PowerHA for i, iCluster, and the DS8000, IBM can now offer IBM i customers a full menu of HA/DR solution choices, as shown in Figure 1-1.

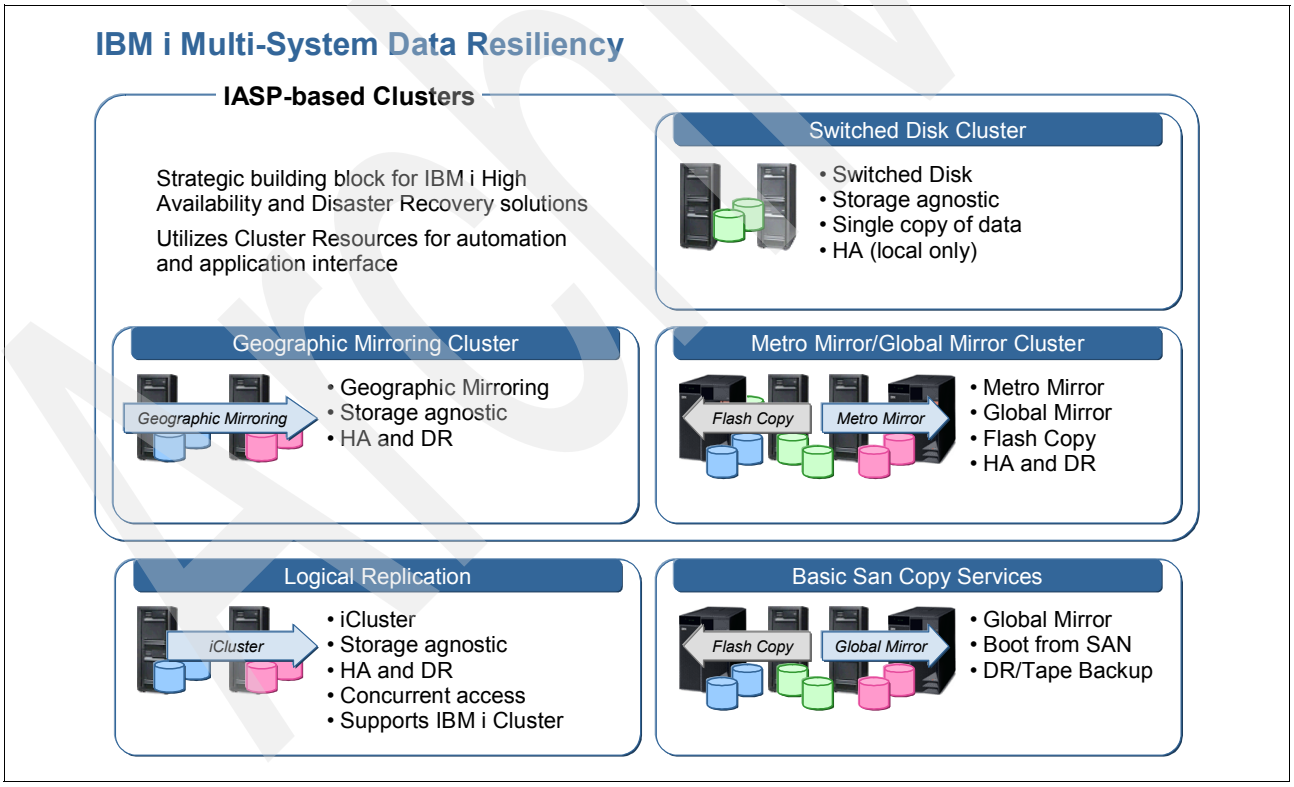


Figure 1-1 IBM i multi-system data resiliency solutions

V6R1 availability capabilities include:

- Simplified cluster management using a new Web-based browser.
- Implementation of your high availability with a solution-based graphical interface that guides the user through the verification, setup, and management of the chosen solution.
- Synchronization of objects that are not in the iASP through the use of the PowerHA function called administrative domain (no third-party replication product required).
- Metro mirror and global mirror are integrated as an extension of IBM PowerHA XSM for DS8000-based and DS6000-based solutions.
- Significant Fibre Channel performance and capacity improvements that put DS8000 performance on par with IBM i native disk storage deployments.
- iCluster software replication optimized for geographic dispersion and data replication/recovery.

1.2 Choosing a solution

Now that IBM offers a full menu of solution approaches that address high availability and disaster recovery requirements, our clients need to do some homework before choosing which solution to deploy. As IBM has discussed in previous articles and papers, the basis for choosing the correct solution must be based on your requirements. Some of the more typical requirements are recovery point objectives (RPOs), recovery time objectives (RTOs), geographic dispersion objectives, staffing, skills, and day-to-day administrative requirements. The underlying data resiliency technology that your solution utilizes is the most fundamental bifurcation.

Multiple system data resiliency methods can be fit into either the logical replication category or the hardware category. The hardware category is further divided into operating-system-based replication called geographic mirroring and the storage-server-based replication of metro or global mirroring. PowerHA for i manages the iASP-deployed solutions in the hardware category. Some of the physical aspects are illustrated in Figure 1-2. Note that in Figure 1-2 the straight arrow implies synchronous replication and the jagged arrow implies asynchronous replication.

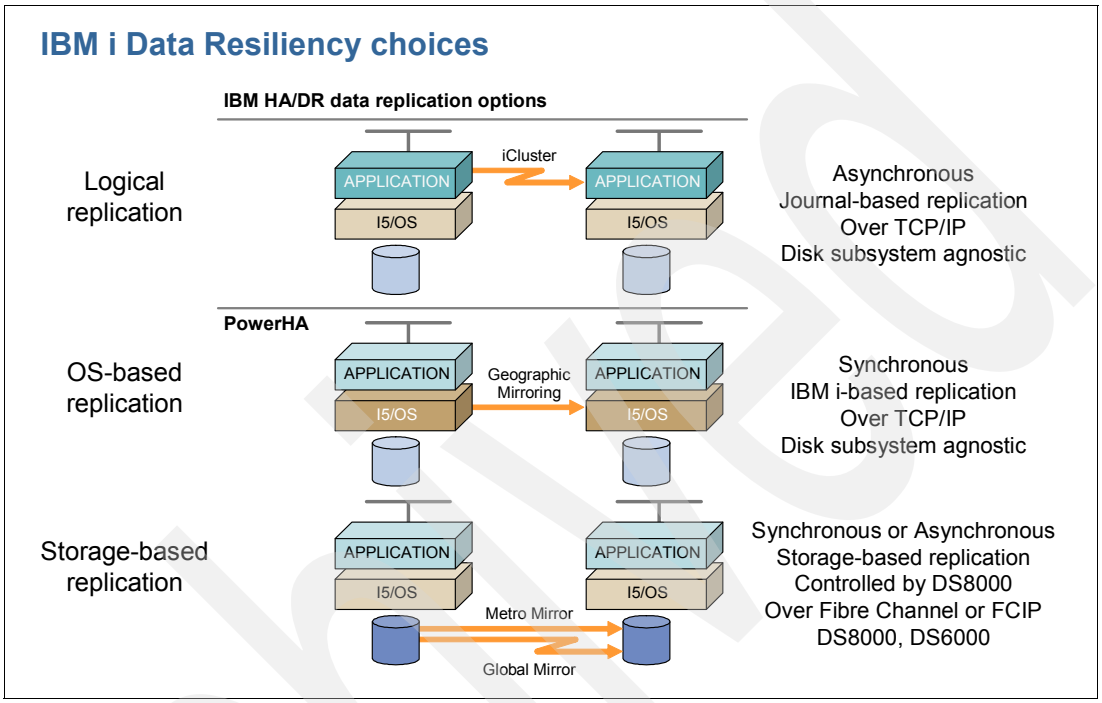


Figure 1-2 IBM i HA/DR data replication options

Depending on the version of your IBM i there is an IBM Availability solution that will fit your needs. In Figure 1-3 we illustrate the different IBM Availability solutions.

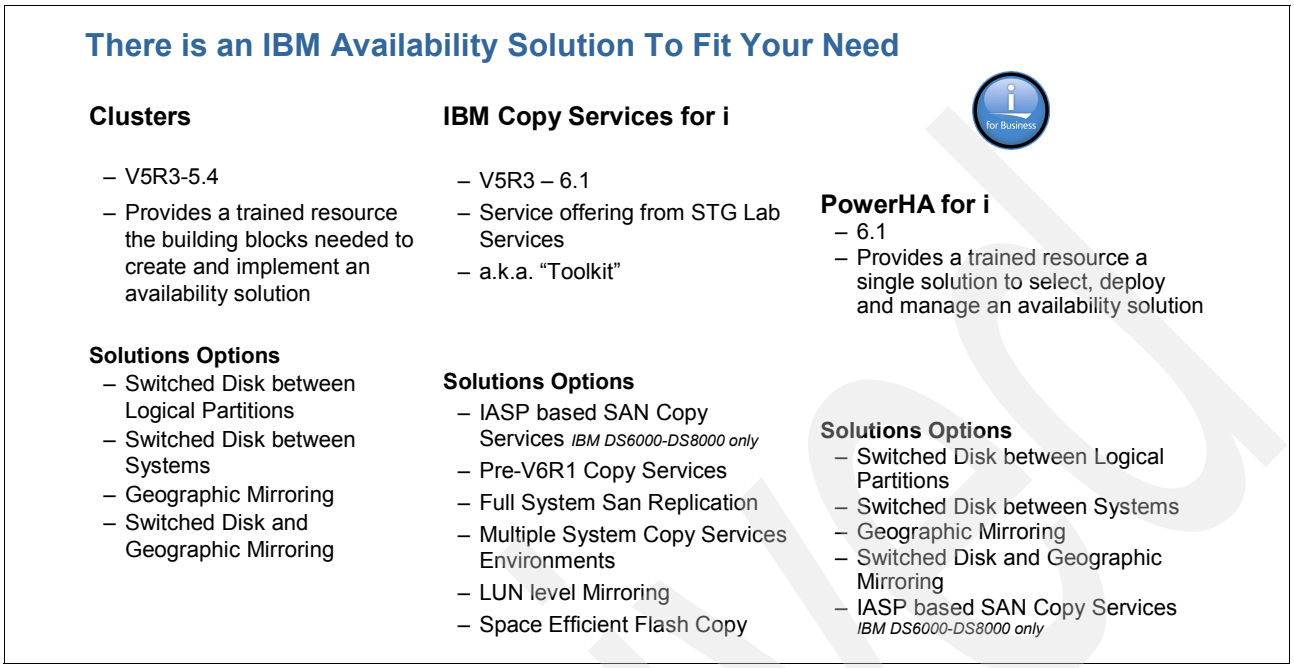


Figure 1-3 IBM Availability solutions for iBM i

1.3 Further considerations

Going beyond the physical characteristics illustrated in Figure 1-2 on page 6, we need to talk a little bit about the underlying replication characteristics from a synchronicity perspective. A storage-based synchronous replication method is one in which the application state is directly tied to the act of replication just as it is when performing a write operation to local disk. The reason that we call these functions *mirroring* is because they are mirroring over IP. 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 as well as 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. The huge benefit in comparison to a logical replication approach is that the two copies are identical, 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 insures 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 metro mirror 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. If your shop does not conduct regular and sustained role swaps, your shop does not have a high availability solution deployment.

The traditional approach for asynchronous data replication in the IBM i environment is called logical replication. This solution approach 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 (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 insure 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, since 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. iCluster is the IBM logical replication-based solution, and there are also solutions provided by other IBM HA ISV partners.

Global mirror is the asynchronous IBM SAN storage server replication technology. Both global mirror and metro mirror can be used independently of iASPs for full system data replication, which addresses RPO but not RTO. There are nearly 100 customers with over 500 partitions deployed in customer environments using the Copy Services Tool Kit. From a practical point of view this means that you do no have to wait until 6.1 to take advantage of your DS8000 to deploy iASP-based clustering to address your HA/DR operations.

1.4 Clustering

Clustering provides the underlying infrastructure that allows the resilient resources (data, devices, and applications) to be automatically or manually switched between systems. It also provides failure detection and response. In the event of an outage, cluster resource services provide mechanisms that enable critical resources to be automatically available on backup systems. The complexity associated with multiple systems and multiple partitions involved in a high-availability topology is managed and simplified with clustering.

1.5 Summary

IBM i 6.1 represents the culmination of over a decade of strategy and development investment, giving you the opportunity to deploy IBM state-of-the-art high availability and disaster recovery solutions. Figure 1-4 illustrates a complete hardware solution for high availability from IBM. PowerHA provides the cluster management, the DS8000 provides the data resiliency with metro mirror and FlashCopy® for nondisruptive offline backups. If it was a disaster recovery environment farther away than approximately 50 miles we would use global mirror for geographic dispersion (conditions vary by client application environment and communications throughput).

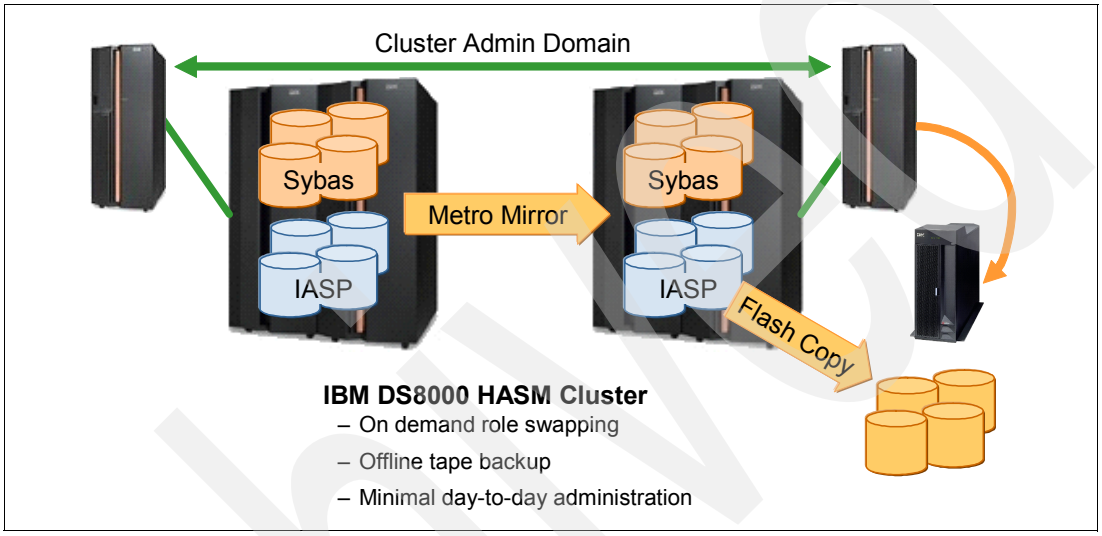


Figure 1-4 Complete hardware solution for HA from IBM

You asked us to deliver this capability. We listened. We delivered. The next move is up to you. Table 1-1 is a summary of IBM i solution choices for your consideration.

Table 1-1 IBM i 6.1 HA/DR offering summary

	Logical Replication	iASP Clustering Integrated Storage	iASP Clustering SAN	Full System SAN Replication
IBM SW offering	iCluster. iCluster SMB.	PowerHA/XSM - geographic mirroring	PowerHA/XSM ► Geographic mirroring ► Metro mirror ► Global mirror (for DR)	► Global mirror ► Metro mirror ► FlashCopy ► Copy Services Toolkit for i
IBM HW 0ffering	IBM i CBU integrated storage.	IBM i CBU Integrated Storage	► IBM i CBU ► IBM DS8000, DS6000	► IBM i CBU ► IBM DS8000, DS6000
Best fit	DR operations can be adapted for HA.	HA operations PowerHA with geographic mirroring	► HA operations PowerHA with metro mirror ► HA operations PowerHA with global mirror	Disaster recovery

2

High-availability building blocks

In this chapter we provide you with an overview of the various high-availability cluster technologies, as well as with components that make up a cluster in an IBM i environment.

2.1 Building blocks: Clustering for enhanced high availability

IBM i *clustering technology* offers you state-of-the-art, yet relatively easy-to-deploy, mechanisms to put your business on the path to near continuous availability. Clustering technology can provide a single point of control. It is through the cluster resource services support that you establish a heartbeat between the systems, monitor and react to outage events, and establish the relationships that assure that your production and target system stay in sync. It is through the cluster resource services that you establish automatic switchover.

In the sections that follow, we discuss some of the key components of IBM i cluster technology. In doing so think about three areas of protection: your data, your application, and your surrounding environment, as illustrated in Figure 2-1.

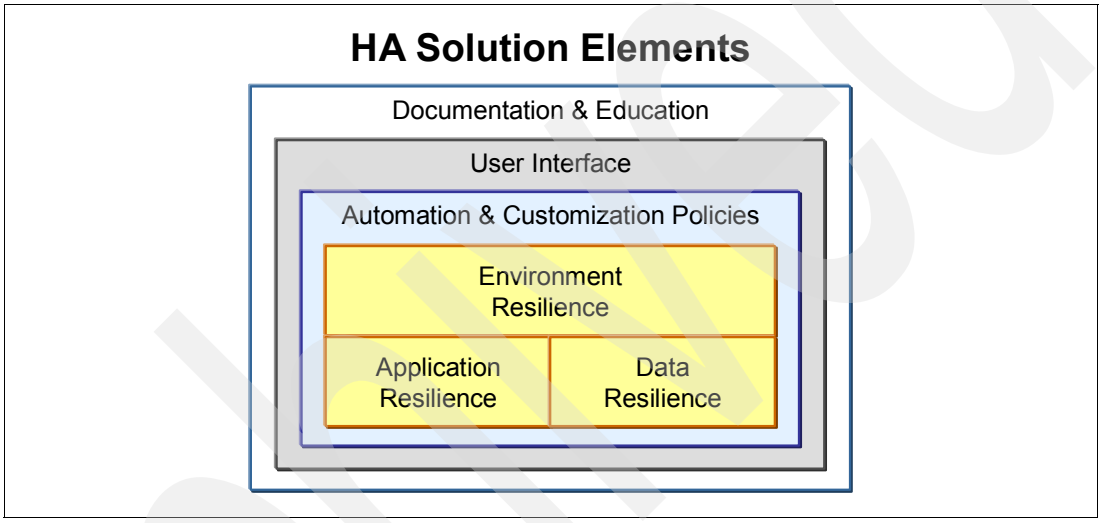


Figure 2-1 The three areas of resilience to be considered

2.1.1 Definition of a cluster

In its most simplistic form, a cluster can be described as collection of complete computing systems (nodes) that are interconnected and utilized as a single, unified computing resource, as shown in Figure 2-2. Of course, in IBM i terms, this need not be a standalone separate system (although you may want one to achieve the highest degree of isolation). Instead, any IBM i logical partition (LPAR) that has the i OS installed may serve as a cluster node.

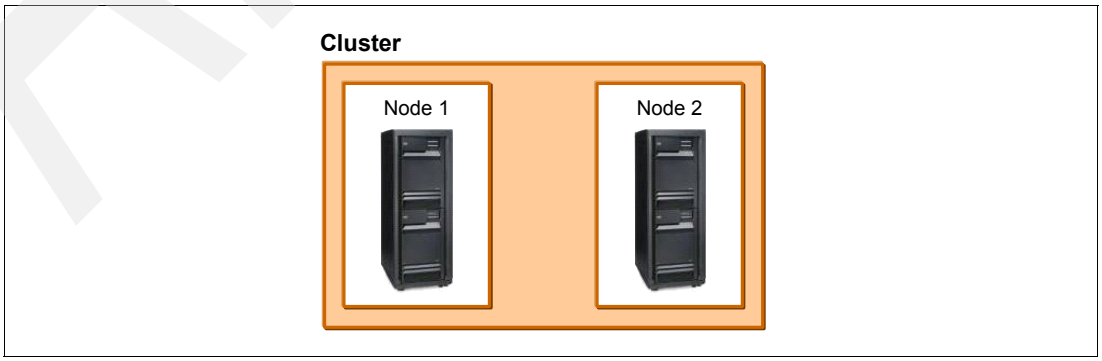


Figure 2-2 A cluster consisting of two nodes

One of the main benefits of a clustered environment is that it delivers coordinated, distributed processing across all the participating nodes. This results in very high levels of system availability, as well as simplified resource administration.

All the high-availability solutions that we discuss in this publication can be built upon *cluster resource services*. Clusters provide the underlying infrastructure that can make your resources (not just your data) resilient to outages. These resources may include data, hardware devices, and applications. You can also configure cluster resource services for the purpose of switching resources between systems either automatically or manually in the event of an outage. If you elect to do so, you should practice such role swaps regularly.

2.1.2 What clustering gives you

With IBM i clustering technology, each cluster node is a complete, fully configured IBM i server or LPAR. While many computer vendors provide a clustering solution to address only a particular business need, the fundamental design of IBM i clustering technology addresses all of the following business requirements:

- Near-continuous system availability
- Simplified system management
- Flexible and scalable resource utilization
- High-speed interconnect communication
- Shared resources

In short, IBM i clustering gives you transparent server backup and failover capacities. It also provides you with redundancy of systems, peripherals, and data, all of which are critical for your complete high availability solution requirements.

IBM i cluster technology is known as cluster resource services (CRS). CRS enables you to define the cluster, as well as the individual resources that you want to protect against outages. It also detects and manages outage conditions and allows for the automatic, coordinated transfer of critical resources to a backup system.

In order to achieve near-continuous availability, it is not sufficient to simply have robust and redundant systems. You also need to ensure that your data, as well as your applications, are resilient to outages. IBM i cluster technology focuses on all these aspects in order to provide you with a complete solution.

2.1.3 Cluster components

As previously mentioned, an IBM i cluster consists of one or more IBM i servers or logical partitions (LPARs) that work together to form a single, unified computing resource. As such, there are a number of elements that need to interact in order to deliver and support one coherent cluster entity. We discuss some of these cluster elements in the sections that follow.

Cluster node

A cluster node is any IBM i system or a logical partition that you have designated to be a member of a cluster.

When you create a cluster, you specify which systems or LPARs you want to include in the cluster. You typically specify a name for each cluster node to identify that system in the cluster. We recommend that you use the host name of the system as the cluster node name.

Each cluster node is associated with one or more TCP/IP addresses. All communications between the cluster services of each node (the so-called heartbeat) occur through the TCP/IP communication interfaces.

All the cluster nodes that you have configured as part of the cluster are often referred to as the cluster membership list.

Cluster resource groups

A cluster resource group (CRG) is an IBM i system object that defines a collection of cluster resources that are used to manage events that occur in a high-availability environment. A cluster resource is a computing resource that is required to be highly available in your business environment. Examples of these computing resources could include application programs, a library containing specific data, or a physical disk unit. You can move or replicate cluster resources between one or more nodes within a cluster.

A CRG allows you to easily monitor and manage a collection of individual cluster resources. It also defines the relationship between the different nodes that are associated with a particular CRG. For example, a CRG allows you to specify which nodes may use a particular cluster resource, which node currently has the resources allocated, and which node should get the resources in the event of a failure.

CRG types

Currently, IBM i cluster defines and supports the following CRG types:

- Device CRG

  The device CRG controls the switching between the mirrored copies of an independent disk pool in cross-site mirroring (XSM) environments. For example, if you experience an outage at your production site, the device CRG switches production to the mirrored copy of your independent auxiliary disk pool (iASP). A device CRG can also contain a collection of hardware resources that can be switched as a single entity via a switchable IOP or tower.

- Application CRG

  An application CRG provides the mechanisms required to restart your application on a backup system if you experience an outage in your production environment.

- Data CRG

  A data CRG is an IBM i system object that assists in the logical replication of data between the primary node and backup nodes in the recovery domain.

- Peer CRG

  A peer CRG is a non-switchable cluster resource group that provides peer resiliency for groups of objects or services in the cluster. It is used when defining an administrative domain to synchronize objects residing in the system ASP.

2.2 Building blocks: Independent auxiliary storage pools

One of the key technology components of IBM i high availability is independent disk pools. Independent disk pools allow you to store data and applications on disk storage units that view themselves as independent from the surrounding system and name space to which they are currently connected. As a consequence, they are self-contained and as a result can be moved from one server to another. This represents an ideal packaging approach for managing data that may need to be accessed from different nodes in the cluster at different times. In a clustered environment, the data and applications stored within independent disk pools can be switched to other systems or logical partitions when required.

A number of IBM i cluster technologies are based on independent disk pools. Some of these include switched disks, geographic mirroring, metro mirror, and global mirror.

Auxiliary storage pools (ASPs)

The concept of auxiliary storage pools is a fundamental design feature of the IBM i architecture. An ASP is a collection of disk units that are grouped together to form a single pool of auxiliary storage. This allows for greater control over the placement of objects on disk and, in turn, it protects data in one ASP from being affected by disk failures in another ASP.

Basic ASPs

By default, every IBM i server has at least one ASP, known as the system ASP, which is designated as ASP1. If you do not configure any additional ASPs, all objects that you create on your system will be stored in the system ASP (often called *SYSBAS). That is probably not what you want for your HA solution, By instead configuring your system with multiple ASPs, you can simplify your system management tasks, reduce recovery time, and (provided that you configure a sufficient quantity of disk units within each ASP) increase your system performance by eliminating disk arm contention. IBM i allows you to create up to 32 basic user ASPs.

Such basic user ASPs share the same name space with the system ASP. Hence, your applications can easily access objects residing in any of the 32 ASPs without any application changes. This follows from the fact that, to your application, the separate physical isolation of disk unit pools is seamless when you employ basic (that is, user) ASPs. That is, they all are part of one unified set of names known as your namespace. In a sense there are no boundaries from your application's point of view and it does not really matter which basic ASP houses the objects that it wants to reference. Such is not the case for the variety of ASP we describe next.

Independent ASPs

There is another variety of auxiliary storage pools known as independent ASPs (iASPs). Such independent ASPs are user auxiliary storage pools that can either be used on a single system or switched between multiple systems or logical partitions. iASPs are often referred to as independent disk pools. IBM i allows you to create up to 223 iASPs.

Because they are so independent (not tightly coupled to the system ASP) and because they are self-contained (allowing their name-space to be temporarily blended with the name space of a surrounding system ASP), they are the ideal building block for many of the high-availability solutions described herein.

However, your applications will notice a difference. Unlike basic ASPs, these independent ASPs require application changes to assure that your application can see the name-space provided by the iASP and thereby access the objects residing within.

If you use third-party packages you will want to verify with your software vendor that they have enabled their product to work in an iASP environment.

Switchable versus non-switchable iASPs

An iASP that is used with a single system is sometimes referred to as a non-switchable or private iASP. When you configure an iASP to be used on a single system, the contents of the iASP can dynamically be made available or unavailable for use on that system only.

The act of making the contents of that iASP visible is known as *varying on* the iASP. This vary-on step is similar to the IPL processing steps that your system experiences, except limited to the objects residing within the iASP. This isolation gives you some flexible options. For example, you might elect to IPL the rest of your system but keep the iASP offline so as to

reduce IPL time. Alternatively, you might elect to perform certain time-consuming housekeeping steps such as reclaim storage and limit its scope to only the objects residing within the iASP.

You can also configure an iASP to be used across multiple systems. These are commonly referred to as switchable iASPs. When you switch an iASP to another system, the entire contents of the iASP can be accessed from that system without having to restart (IPL) the entire system. However, this does not mean that there are no priming steps required when a switched iASP is transferred. Quite the contrary. Under the covers a mini-IPL (also known as a *vary on*) of the objects residing within the switched iASP does ensue, and certainly you need to factor in the duration of this vary-on step when contemplating your RTO. It is not instantaneous.

Note: Keep in mind that switchable iASPs require IBM i clustering support, which is available through separately priced IBM i License Product 5761-SS1 Option 41.

The other consideration regarding high-availability approaches built upon independent disk pools is that not all objects can reside within an iASP. There are certain classes of objects that may not reside within an iASP (user profiles, for example). These objects must reside in the system ASP. If those objects need to be synchronized across the cluster nodes, then the iASP-based replication should be supplemented with the administrative domain.

Note: The IBM i Information Center contains a complete list of objects that are not supported in iASPs. For more information about iASPs, refer to the topic Disk Pools in the IBM i Information Center.

2.3 Building blocks: Journaling and commitment control

Journaling is an inherent feature of IBM i. The system employs journals to ensure both data integrity and recoverability. Journaling also serves as the cornerstone upon which many logical replication solutions are built.

2.3.1 Journaling protection in a clustered environment

The support afforded by journaling can provide you with mechanisms to minimize the possibility of the loss of critical enterprise data, thereby helping you achieve your RPO. You can achieve high levels of data resiliency by journaling your database tables, as well as other supported objects that contain critical data. Some of these objects include data areas, data queues, and the contents of the Integrated File System (IFS).

With IBM i 6.1, you now also have the ability to journal library (*LIB) objects. When you elect to do so, any journal-eligible object that you create in such a library hereafter will automatically inherit the journaling attributes that you have specified for that library. In addition, you now also have more fine-grained control over which object types in a library you automatically want to start journaling, as well as which operations and data images you want to include or omit from the journal. You can start journaling library objects by using the new Start Journal Libraries (STRJRNLIB) CL command.

2.3.2 Commitment control in a clustered environment

Journaling alone only assures RPO to a point in time. Alone, it does not assure that the recovery point represents a natural boundary for an application transaction. Consider, for example, a travel agency application that books an airline seat, rental car, and hotel room for an upcoming trip. If you enable journaling protection for your database files but fail to incorporate commitment control protection into your applications, the operating system has no way to discern that the trio (airline, rental car, hotel room) are viewed as related aspects of a singular reservation. Use of commitment control solves that problem. It ensures that your recovery point will honor the transaction boundaries built into your application. In our travel agency example, use of commitment control ensures that anyone who has an airline seat also has a matching rental car and hotel room reservation. That is, it ensures that you recover to a point where the entire trip reservation is complete, thereby ensuring transaction integrity, which is what your users probably expect.

With commitment control, you can define and process a group of changes to resources, such as database files or tables, as a single, coherent transaction. Without consistent use of commitment control your travel agency customer could end up with no place to sleep for the night when he lands at his destination even though your HA approach recovered to a point in time. As a consequence, robust transaction recovery should be included as part of any serious high-availability plan.

By making use of commitment control, you can ensure that an entire group of individual changes occurs on all systems that participate in the transaction, or that no changes occur if any of the individual changes were unsuccessful. That is, you achieve transaction integrity not only on your production system but on your target system as well.

Commitment control employs IBM i journaling to ensure that at role swap time any in-flight transactions are rolled back and that any completed transactions are replayed.

You can use commitment control to design your application in such a way that the system can restart the application if a job or an activation group within a job or the system ends abnormally. You can therefore ensure that when the application is restarted, no partial updates are in the database due to incomplete transactions from a prior failure. In short, any serious RPO ought to include consideration of the resulting recovered state. Rigorous use of commitment control in your applications can help ensure that the resulting state will satisfy your users who expect to see complete transactions rather than partially completed travel reservations.

2.4 Data resilience

In the sections that follow we look at some of the data replication technologies that can be used in a clustered environment.

2.4.1 Logical replication

Software solutions built upon a logical replication approach typically make use of journaling to capture information regarding transactional data changes made on the source system and then transport and replay those same changes logically (as though an identical application were executing at a distance) to one or more target systems. The approach is called logical replication because, rather than operating at the physical level (disk sector images replicated or main memory page frames replicated), the replication is driven at the object level or

meaningful recognized entities within the objects (such as rows within a database file or messages within a data queue).

2.4.2 Switched disk

Independent auxiliary storage pools are one of the key elements of an IBM i clustered environment.

Often referred to as a switched disks or switchable disk pools, you can switch access to an iASP between the participating nodes within a cluster as and when required. In order to facilitate the switching mechanism, the iASP must reside on a switchable device. A switchable device could be either an external expansion unit (tower) if you want to switch disks between two physical systems or an IOP on a shared bus or assigned to an IO pool if you want to switch the attached disks between two logical partitions in the same physical system.

Hardware that does not have a physical IOP uses a virtual logical representation of the IOP. Figure 2-3 shows an example of an i5/OS cluster with switched disks.

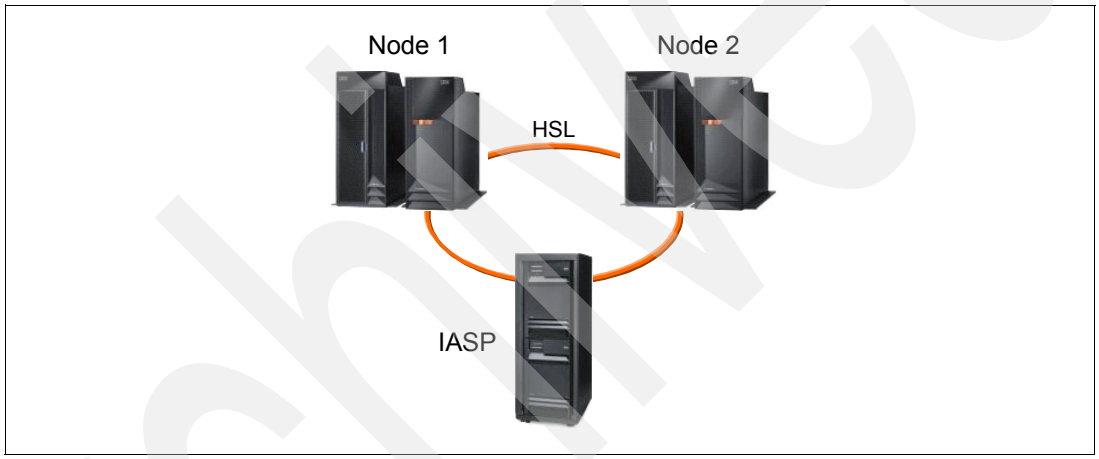


Figure 2-3 Switched disks in IBM i cluster

iASPs are controlled by a device cluster resource group. The CRG allows the iASP to be switched automatically in the case of an unplanned outage, or you can initiate the switchover manually. We discuss CRGs in more detail in 2.1.3, “Cluster components” on page 13.

By combining iASPs with IBM i cluster technology, you can create a simple and cost-effective high availability solution that will take care of planned and some unplanned outages.

For more information about iASPs refer to the topic Disk Pools in the System i Information Center.

2.4.3 Cross-site mirroring

Cross-site mirroring is a collective term that we use to describe several IBM i supported mirroring technologies that you can use to achieve disaster recovery and high availability. XSM not only maintains a mirrored copy of your data, it also manages the replication process and controls the point of access to the data. In the event of an outage of the source system, you can make the mirrored copy of the data available to your users by performing either an automatic or a manual switchover.

XSM extends the advantages of switchable iASPs by giving you increased data and application resilience through geographic dispersion. XSM gives you the ability to mirror the contents of an iASP to a secondary iASP attached to another, often remote, system, over a communications fabric. So, as you write data to your production iASP, a mirrored image of that data is automatically written to a backup iASP on another system. Depending on the physical implementation that you choose this process is either under the control of the IBM i storage management or under the control of the IBM System Storage™ unit.

All changes that you write to the production iASP on the source system are guaranteed to be written to the mirror copy on the target system and in the same order. You therefore have a real-time hot backup copy of your data. In the event of the production iASP failing or if you need to shut it down for any reason, the backup iASP can become the production copy of your data.

XSM has built-in functionality that enables the switching or automatic failover to a mirrored copy of the iASP, as well as local switching between systems. This addresses the single-point-of-failure issue of the basic switchable device structure.

In addition, XSM also provides real-time replication support for System i hosted environments such as Microsoft® Windows®, Linux®, and AIX®. This is not generally possible with an i5/OS journal-based logical replication solution.

The sub-categories of XSM include geographic mirroring, metro mirror, and global mirror. We discuss these in more detail in the sections that follow.

Geographic mirroring

Geographic mirroring, when used with IBM i cluster technology, provides a high availability solution where your production iASP data is mirrored to a backup iASP that is attached to another, remote system. Geographic mirroring maintains a consistent backup copy of an iASP by using either internal or external storage. Figure 2-4 shows a high-level diagram of a basic two-node geographic mirroring environment.

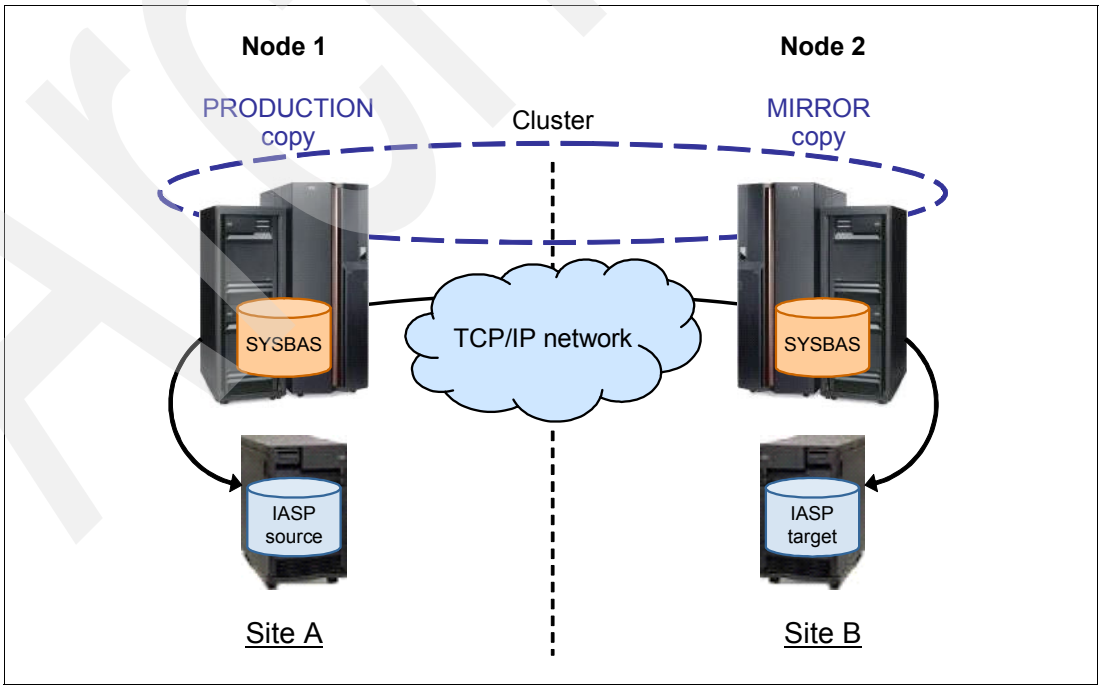


Figure 2-4 Geographic mirroring environment

The benefits of geographic mirroring are essentially the same as that of basic switchable device solutions. The added advantage is that it provides disaster recovery to a second iASP at a geographically dispersed location. This gives you disaster recovery protection in the event of a site-wide outage. The biggest benefit still remains in the operational simplicity. All the data that you store in the production iASP, including objects such as journal receivers, are constantly mirrored to a second iASP that is attached to a remote system. Your data is mirrored before the write operations on the production system complete. This is a great advantage if you have applications that cannot afford any data loss in the event of an outage.

Geographic mirroring provides logical page-level mirroring between iASPs through the use of data port services running over TCP/IP. Data port services are used to spread one logical connection (the mirroring process) over multiple TCP/IP addresses. This provides redundancy and greater bandwidth in geographic mirroring environments.

The switching operations are largely the same as that of switchable device solutions.

Considerations to take into account in terms of geographic mirroring are:

- You cannot make an iASP available on a system that is at a lower IBM i release level than the primary system.
- At any point in time you can only access data on the current production iASP. Concurrent access to the backup iASP is not possible while geographic mirroring is running.
- Some objects (like user profiles) cannot reside in an iASP. To ensure data integrity and consistency, any objects that reside in the system ASP of the production system must be replicated to the backup system via some other mechanism. This can be either the use of an administrative domain (see 2.6, “Environment resilience: Administrative domain” on page 23, for details) or some other automated or manual process.
- Greater distances between the iASPs may require additional communication bandwidth to improve response times. It is important that you understand your distance needs and their associated implications in order to establish your communication infrastructure requirements.

Metro mirror

Metro mirror is part of the optional copy services feature of IBM System Storage external storage units.

Metro mirror, previously known as Synchronous Peer-to-Peer Remote Copy (PPRC), is designed to continuously replicate any changes that you make to the data on your primary data volume, to a secondary data volume. The primary and secondary volumes may be on the same storage server or on separate, geographically dispersed storage servers, up to 300 km apart. Figure 2-5 shows an example of a basic metro mirror landscape.

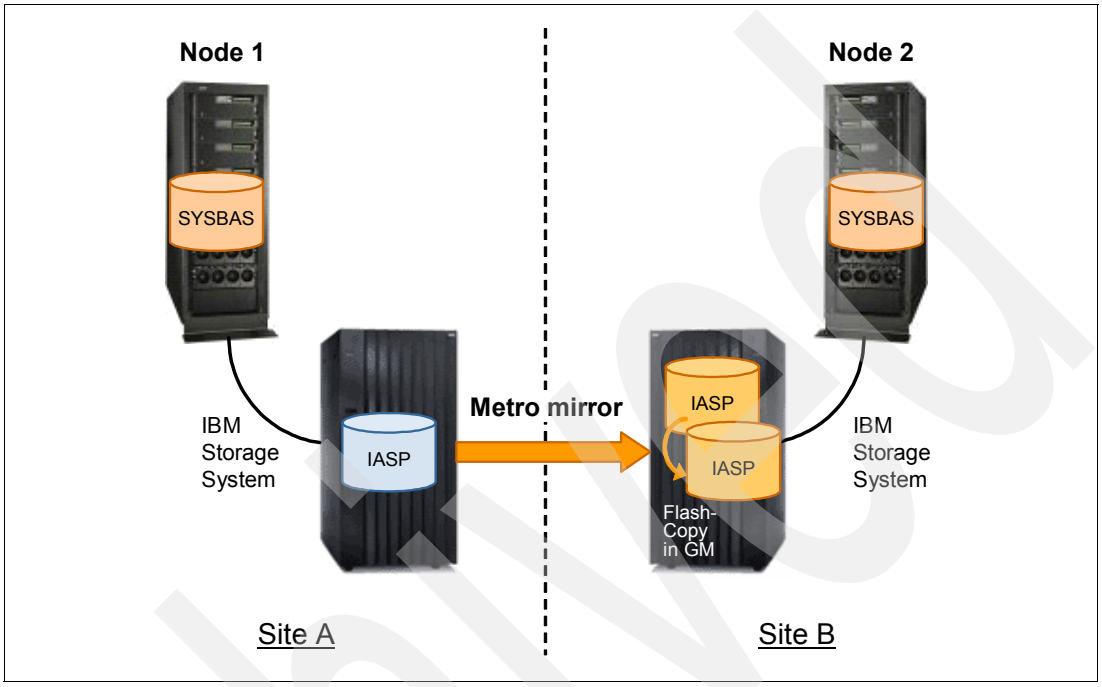


Figure 2-5 Metro mirror landscape

Take into account the following considerations about metro mirror:

- The data mirroring function occurs at the disk subsystem level. Your applications, therefore, are not aware of this process.
- Your data is replicated at the disk volume level. You might therefore find that additional time is required for the automatic database recovery process to complete when you bring your mirrored copy online.
- Greater distances between the iASPs may require additional communication bandwidth to improve response times. It is important that you understand your distance needs and the associated implications in order to establish your communication infrastructure requirements.

Global mirror

Global mirror, previously known as Asynchronous PPRC and PPRC Extended Distance (PPRC-XD), is part of the optional copy services feature of IBM System Storage external storage units.

Global mirror provides disk I/O subsystem level mirroring between two IBM System Storage external storage units. It is designed to provide a long-distance remote copy solution between two sites through asynchronous replication technology.

By making use of high-speed Fibre Channel communication links, global mirror maintains a complete and consistent remote mirrored copy of your data. This process occurs asynchronously at virtually unlimited distances without any noticeable impact on application response times. By separating your data centers over larger distances, you can achieve high-availability protection in the event of a regional outage.

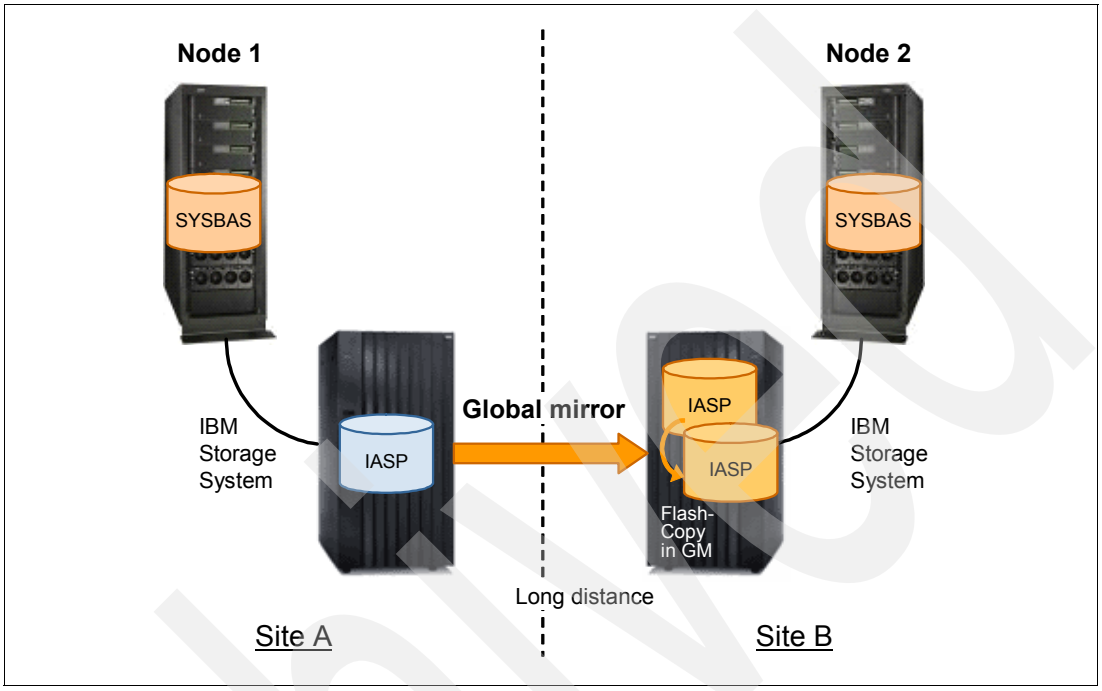


Figure 2-6 Global mirror environment

This asynchronous replication technique also provides better performance at larger distances than metro mirror does. However, the secondary site will typically lag a few seconds behind the primary site.

With global mirror, you can achieve a high-performance and cost-effective global distance data replication and disaster recovery solution. However, take into account the following considerations about global mirror:

- The remote copy of your data can potentially lag quite a number of transactions behind the production copy, depending on the amount of I/O activity and bandwidth availability. You should take this into consideration in terms of your RPO.
- Your data is replicated at the disk volume level. You might therefore find that additional time is required for the automatic database recovery process to complete when you bring your mirrored copy online.

2.5 Application resilience

Now that you have taken steps to ensure that you have access to your data in the event of a disaster, you also need to ensure that the applications that use the data are accessible to the users in the case of a system outage. Application resilience is one of the key elements in a high-availability environment.

Automatic or manual

Again, depending on your requirements, there are a number of different levels of application resiliency options that you can choose from. Your RTO will dictate how much automation is required for an application to restart in the event of a system outage. Maybe your business can afford to wait for human intervention to restart the application after a system outage. If, on the other hand, your business cannot afford any application interruption, you might want to consider a solution that incorporates automatic failover and application restart. The clustering support helps you automate your response and monitors for node failures.

With sufficient automation, a resilient application can be restarted on a different cluster node without any need for you to reconfigure the clients. In addition, the data that is associated with the application (at least as much of it as your RPO approach assured had reached disk) will also be available after the planned switchover or unexpected failover. Your users can therefore experience minimal interruption, or even near-seamless switchover, while the application and its data switch from the primary node to the backup node.

Do your planning

While application resiliency is a cherished objective, it takes some careful planning and some serious analysis of your application environment along with a willingness to roll up your sleeves and provide some customized restart software in order to achieve best-of-breed application resiliency.

Application resiliency requires that your applications meet certain availability specifications. There are a number of characteristics that must be present in the application in order for it to be switchable and therefore highly available within the cluster. For more information about how to make your applications resilient to system outages, refer to the topic Planning application resiliency in the IBM i 6.1 Information Center.

The IBM i clustering framework gives you the ability to automate the application recovery for different types of failures. The amount of automation that you can achieve depends on how much of the manual switchover procedures you elect to automate and also the types of applications that you are using. By making use of customized exit programs (which you would need to write), steps required to switch over the application can be automated. You should also ensure that your applications run in a client-server environment. A client-server application design ensures that application availability is separated from the application data availability. Remember, not every application will give you all the availability benefits of clustering.

Redundancy matters

If possible, for even more versatility and resilience, configure multiple nodes in the cluster so that any of these nodes can automatically assume the role of the application server in the event of the primary application server failing. Also, by making use of TCP/IP address take-over capabilities, you can put measures in place to automatically restart the application on the backup node.

Redundancy is the key to a well-designed cluster. It is important to ensure that you eliminate every single point of failure (SPOF) that could potentially cause a system outage or interruption. This includes networking components, power supplies, and so on.

2.6 Environment resilience: Administrative domain

In an IBM i High Availability environment, a cluster administrative domain provides a mechanism to maintain a consistent operational environment across the defined cluster

nodes. It ensures that highly available applications and data function as expected when a failure occurs or when you manually switch over to a backup node.

Applications and application data often have associated configuration data or parameters and we often refer to these as the operational environment for the application. Some of these operational environments may include user profiles that are used to access the application or its data, or system environment variables that control the behavior of the application.

In a high-availability environment, you must ensure that the operational environment is identical on every system where the application can run, or where the application data resides. If you change configuration parameters or data on one system, that same change has to be replicated to all the other systems.

A cluster administrative domain provides you with the mechanisms to identify resources that need to be maintained consistently across the systems in your IBM i high availability environment. In addition, the cluster administrative domain also monitors for changes to these specified resources and then keeps these changes synchronized across the active domain.

The type of objects that can be managed in a cluster administrative domain, also known as monitored resources, have been enhanced in IBM i 6.1.

Table 2-1 Monitored resource entry type supported

Object or attribute description	Type
Device description for an auxiliary storage pool device	*ASPDEV
Class description (*)[a]	*CLS
System environment variable (*)	*ENVVAR
Line description for an Ethernet line	*ETHLIN
Job description (*)	*JOBD
Network attribute (*)	*NETA
Network server configuration	*NWSCFG
Network server description	*NWSD
Device description for a network server host adapter	*NWSHDEV
Network server storage space	*NWSSTG
Device description for a optical device	*OPTDEV
Subsystem description	*SBSD
System value (*)	*SYSVAL
Device description for a tape device	*TAPDEV
TCP/IP attribute (*)	*TCPA
Line description for an token-ring network line	*TRNLIN
User profile (*)	*USRPRF

a. (*) Available from i5/OS V5R4

Some of these objects (or attributes) that are part of the operational environment used to run the client applications need to be the same on every system node of the cluster where these applications can run or where the application data reside. When a change is made to one or

more of these objects on one system, the same change needs to be replicated on all system nodes, part of the device domain in the cluster. The IBM i administrator will use the cluster administrative domain to identify each resource that needs to be maintained consistently across the systems. Any changes made on resources defined in the cluster administrative domain are monitored and synchronized across all the nodes defined in 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 peer CRG's recovery domain. As said by its name, in a *peer* CRG all the nodes that make up the cluster administrative domain are all peer nodes. Any change made on one node is replicated to other nodes.

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

You can manage a cluster administrative domain using either CL commands or the Cluster Resource Services graphical interface in IBM Systems Director Navigator for i5/OS.

Monitored resources

A monitored resource is a system resource that is managed by a cluster administrative domain. Any changes that you make to a monitored resource are synchronized across the nodes in the cluster administrative domain and are applied to the resource on each of the active nodes. Monitored resources can be system objects such as user profiles or job descriptions. It may also be a system resource that is not represented by a system object type, such as a single system value or a system environment variable (see Table 2-1 on page 24). Monitored resources are represented in the cluster administrative domain as monitored resource entries (MREs).

A cluster administrative domain supports monitored resources with simple attributes, as well as compound attributes. A compound attribute contains zero or more attribute values (for example, subsystem descriptions (*SBSD) and network server descriptions (*NWSD)), while a simple attribute contains only a single value.

In order 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 will be created. Also, if a node is added to the cluster administrative domain at a later stage, the monitored resource will also be created on this node if it did not exist before. You can only add MREs to the cluster administrative domain if all nodes in the domain are active and participating in the group. You cannot add MREs to the cluster administrative domain if the domain has a status of partitioned.

Once you have added an MRE to the cluster administrative domain, any changes that you make to the resource on any active node in the cluster administrative domain will be propagated to all nodes in the active domain. If a node in a cluster administrative domain is inactive, the synchronization option controls the way changes that are propagated throughout the cluster.

If you set the synchronization option to Active Domain, any changes that you make to the resource on the inactive node are discarded when the node rejoins the cluster. If you select the Last Change synchronization option, any changes that you make to the resource on the inactive node are only discarded if there was a more recent change to the resource propagated in the cluster administrative domain. If you delete a cluster administrative domain, all monitored resource entries that are defined in the cluster administrative domain are removed. However, the actual resource is not removed from any node in the active domain.

2.7 Building blocks: Practice, practice, practice

Always exercise your HA plan. It is not enough simply to set up an environment for HA. The plan must be exercised regularly to ensure that both the automated and manual processing involved yield the desired results and that your operations staff are familiar with the plan. Planned role swaps should be a normal and well-rehearsed part of running your business.

3

Introducing PowerHA for i

This chapter provides a description of the new license product IBM System i PowerHA for i, previously known as High Availability Solutions Manager (HASM), available with IBM i 6.1 (license program number 5761-HAS).

3.1 PowerHA for i introduction

IBM System i PowerHA for i is a new end-to-end, high-availability offering for the IBM i 6.1 operating system. PowerHA for i, when combined with independent auxiliary storage pool (iASPs) and HA Switchable Resources (HASR- i5/OS option 41), enables a complete end-to-end, hardware-based clustering solution for high-availability and disaster recovery operations. PowerHA for i along with cross-site mirroring (XSM) enables solutions to be deployed via IBM DS8000 storage server or internal disk. It also includes tools that enable administrators to configure and manage high-availability solutions using one of two GUIs, as well as the option for a command-line interface.

This new license product allows IBM i administrators to configure and manage their high-availability and disaster recovery solutions simply by using two new Web Graphical User Interfaces that have been integrated in IBM Systems Director Navigator for i5/OS:

- Cluster Resource Services GUI (CRS GUI)
- High Availability Solutions Manager GUI (HASM GUI)

See Figure 3-1.

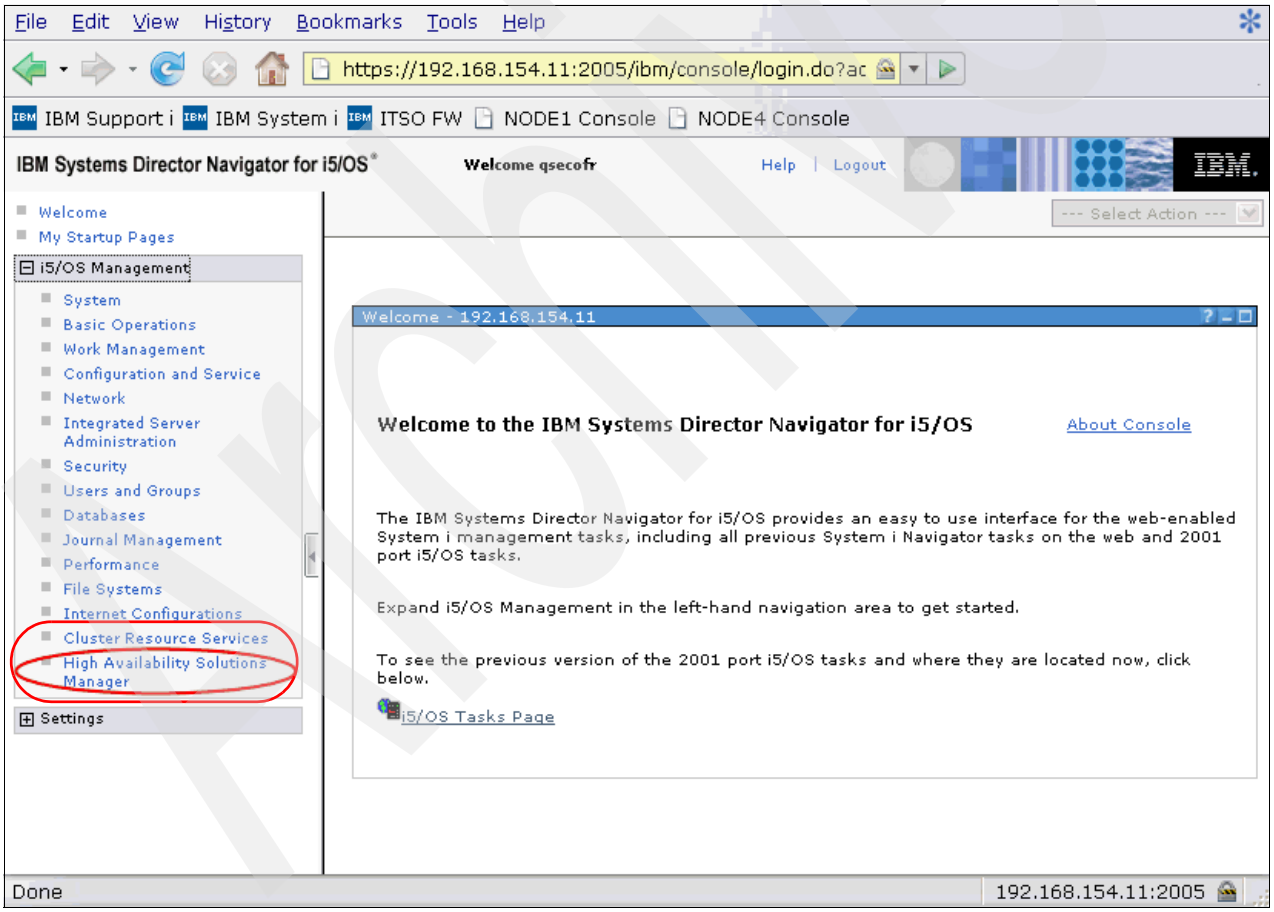


Figure 3-1 IBM Systems Director: HASM GUI interfaces

These user interfaces invoke cross-site mirroring to manage geographic mirroring (native i5/OS replication), metro mirror (DS8000 synchronous replication), or global mirroring (DS8000 asynchronous replication). Using either the command-line interface or the CRS GUI, you can also perform FlashCopy operations on the DS8000 server.

PowerHA for i or (HASM) helps protect critical business applications from outages. Combined with IBM i 6.1, PowerHA for i delivers tools for configuring, monitoring, and managing your high-availability clustering solution.

PowerHA for i, along with IBM i 6.1, offers the following functions and benefits:

- IBM i 6.1 enables integrated support that uses IBM i and PowerHA for i to enable an end-to-end high-availability deployment and management solution. You can cluster partitions on a single system, multiple systems, or a combination of systems. PowerHA for i capabilities include switching of additional switchable devices as well as iASPs. These include printers, tape and optical devices, various communication line descriptions, and network server descriptions.
- Cluster administrative domain covers the majority of system-based objects required by switchable applications. This means that in most cases, the application environment on the cluster nodes are synchronized. Application and data can be activated seamlessly on an alternate node.
- PowerHA for i also includes integrated source and target side tracking for geographic mirroring. This means that when you detach a target system, the resync operation, after reattaching, includes only the changed objects on the source and target system.
- PowerHA for i enables you to perform role-swap operations using metro mirror, a synchronous replication product for the DS8000 server. You can readily perform both planned and unplanned switching operations with minimal impact on operational procedures. You should use metro mirror for best-case recovery point objective (RPO) and recovery time objective (RTO).

For disaster recovery solutions you must have a remote recovery center. In such cases the primary objective is not RTO but rather RPO. You can use PowerHA for i to manage global mirror to perform asynchronous data replication operations.

3.2 Graphical interfaces

PowerHA for i includes three types of interfaces for controlling and managing the high-availability environment:

- Solution-based approach or HASM GUI for small business environments (deploying on a single iASP topology, which can include a mirrored copy via geographic mirroring)
- Task-based approach: Cluster resource services for all environments
- Traditional command-line interface

GUI operations are deployed via IBM Systems Director for i5/OS, a Web-based console.

IBM Systems Director Navigator for i5/OS console is accessed from a Web browser using the following http address (as console in previous i5/OS releases):

```
http://you_server_IPaddress:2001
```

You will be redirected to:

```
https://you_server_IPaddress:2005/ibm/console/logon.jsp
```

This new console is secured. You must enter your i5/OS user ID and password to access it, as shown in Figure 3-2.

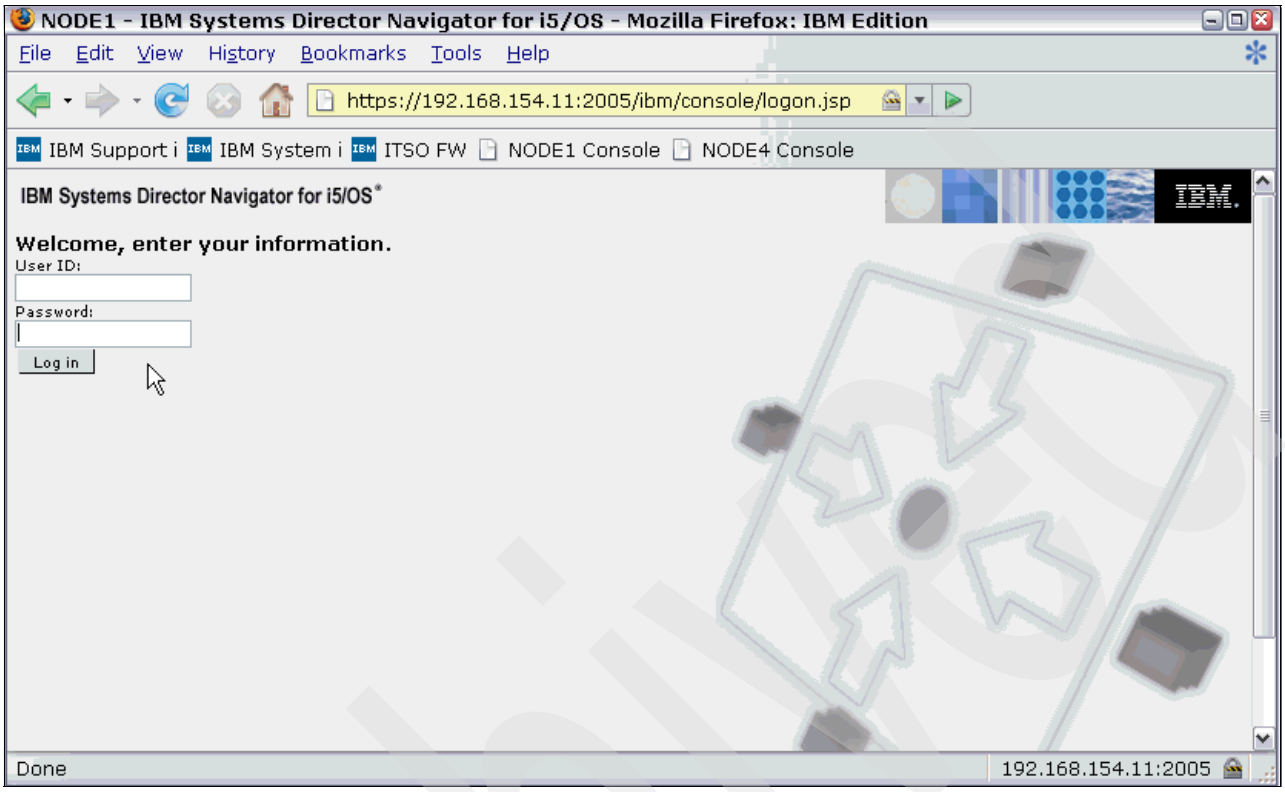


Figure 3-2 IBM Systems Director Navigator for i5/OS login window

3.2.1 High Availability Solutions Manager GUI (HASM GUI)

PowerHA for i provides a solution-based approach for customers willing to implement a high-availability environment by selecting one of four solutions proposed:

- Switched disk between logical partitions
- Switched disk between systems
- Switched disk with geographic mirroring
- Cross-site mirroring with geographic mirroring

The graphical interface option called High Availability Solutions Manager (HASM) GUI available from IBM Systems Director for i5/OS allows administrators to select and configure for the first time one of the predefined high availability or disaster recovery solutions that are based on IBM i high availability technologies, such as independent disk pools and clusters.All the hardware requirements must be in place to allow the creation of the new HA solution environment.

As soon as the solution has been set up, System i Administrators will be able to manage their new HA environment from both GUI interfaces.

The HASM GUI guides users through the process of selecting, configuring, and managing a high-availability solution. The user must complete each step before continuing to subsequent steps. When PowerHA for i is installed you can access the HASM GUI in the IBM Systems Director Navigator for i5/OS. The HASM GUI has the features shown in Figure 3-3.

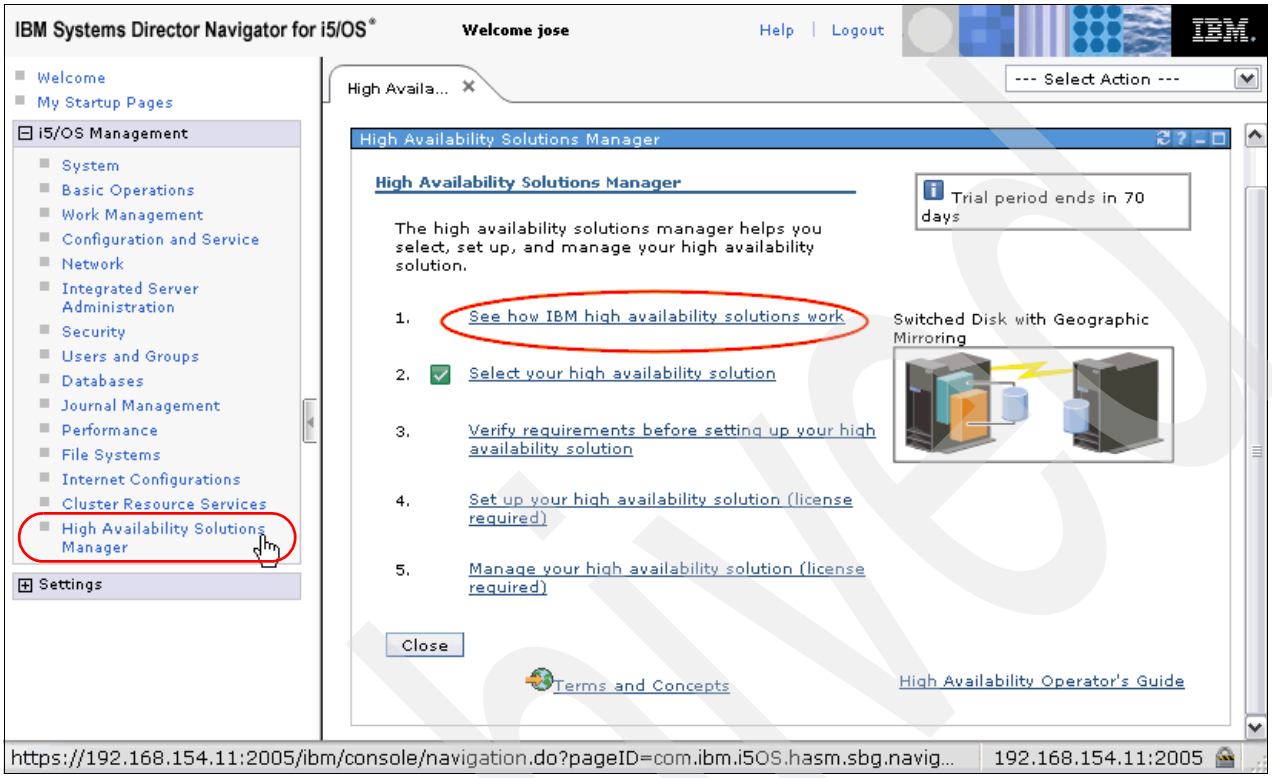


Figure 3-3 HASM GUI

The options are:

- Option 1 allows the user to display a flash video showing IBM System i PowerHA for i. It gives more details on each of the four solutions that can be chosen to be setup via the HASM GUI.
- Option 2 allows the System i Administrator to select one of the possible high-availability solutions that can be implemented in an HA environment.
- Option 3 performs the verification of hardware requirements required to set up the chosen solution.
- Option 4 performs the high availability solution environment setup according to the choice made in step 2 and checking made in step 3. The time needed to perform this step can be long. Many processes must be run in order to fully create the environment. See Chapter 6, “High Availability Solutions Manager GUI” on page 89, for more information.
- Option 5 gives the System i Administrator the possibility to manage its High Availability environment that has been selected and set up in previous steps.

Important: To perform options 2, 3, and 4, the System i Administrator *must* be connected to the QSECOFR profile.

The GUI interface only allows System i Administrators to perform steps 2 to 5 in the correct order, and prevents us from running these steps in the wrong order. One step cannot be

performed if the previous one has not been done (green check box on the left side of the option).

3.2.2 Cluster resource service GUI: Task-based approach

The Cluster Resource Services graphical interface (CRS GUI) lets System i Administrators perform tasks with i5/OS high-availability technologies, such as working with clusters, cluster resource group, device domain, and cluster administrative domain to configure their high availability solution and manage it by performing planned switchovers, as shown in Figure 3-4.

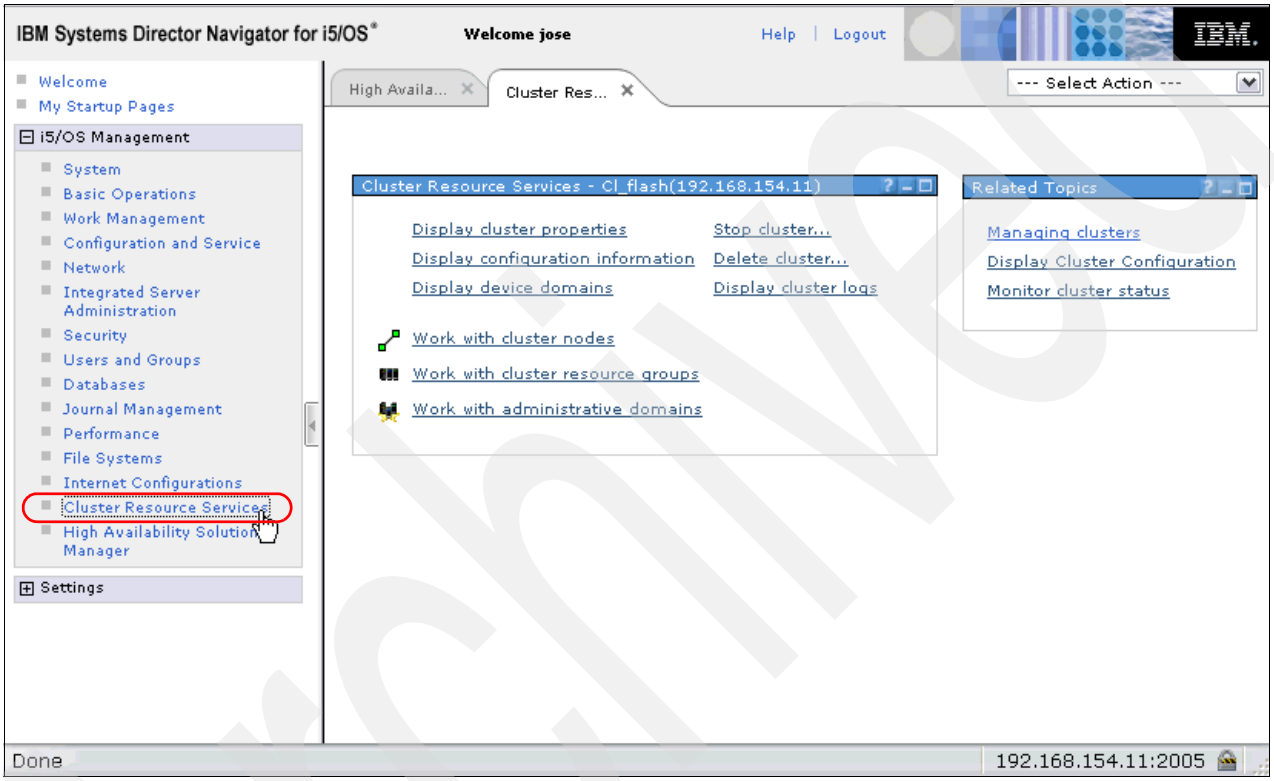


Figure 3-4 Cluster Resource Services GUI

The CRS GUI allows you to configure and manage clustered resources and environments. Unlike the HASM GUI, this CRS GUI is based on task-oriented goals.

This interface allows you to:

- Create and manage a cluster.
- Create and manage cluster nodes.
- Create and manage cluster resource groups (CRGs).
- Create and manage cluster administrative domains.
- 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.

Also, depending on the high availability solution that has been set up either previously to i5/OS V6R1 or from the PowerHA for i license product, you still need to configure and manage

additional technologies, such as geographic mirroring or independent disk pools, which are outside of the Cluster Resource Services graphical interface.

These functionalities can still be accessed by selecting the **Configuration and Service** option on the console and then selecting **Disk Pools**, as shown in Figure 3-5:

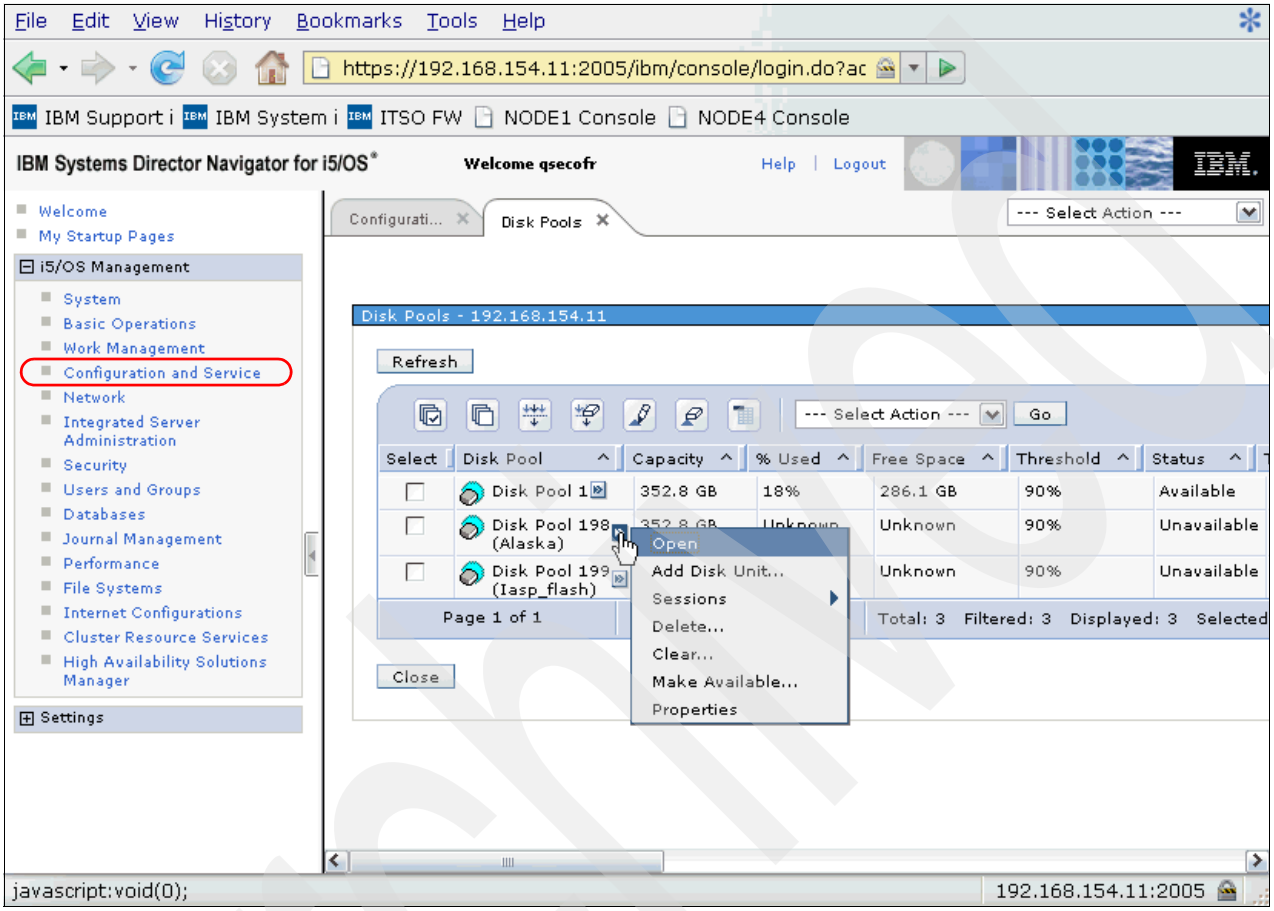


Figure 3-5 Configuration and Services

3.2.3 PowerHA for i and IBM i commands

For System i Administrators who prefer the 5250 green screen interface, the PowerHA for i license product also contains a command-line interface. It allows automation of high-availability processes by simply integrating these CL commands in your own CL programs.

The PowerHA for i license product is based on cluster commands and functions existing in previous i5/OS releases. Most of the CL commands have been moved from system library QSYS to the new library QHASM delivered with the new license product PowerHA for i, which also contains brand new commands and functionalities that are described later in this book.

Commands that remain in IBM i

A few CL commands remain in IBM-supplied system library QSYS. They are available with IBM i 6.1 and do not require the PowerHA for i license product to be installed.

Command description	
Change cluster recovery	CHGCLURCY
Delete cluster resource group	DLTCRG
Dump cluster trace	DMPCLUTRC
End clustered hash table server	ENDCHTSVR
Start clustered hash table server	STRCHTSVR

Commands moved from IBM i (i5/OS) to PowerHA for i

The following commands that were delivered in previous releases of i5/OS are now available through the license product PowerHA for i only and are supplied in the new IBM library called QHASM.

Command description	
Add cluster node entry	ADDCLUNODE
Add CRG device entry	ADDCRGDEVE
Add CRG node entry	ADDCRGNODE
Add device domain entry	ADDDEVDMNE
Change cluster node entry	CHGCLUNODE
Change cluster version	CHGCLUVER
Change cluster resource group	CHGCRG
Change CRG device entry	CHGCRGDEVE
Change CRG primary	CHGCRGPRI
Change device description for ASP	CHGDEVASP
Create cluster	CRTCLU
Create cluster resource group	CRTCRG
Create device description for ASP	CRTDEVASP
Delete cluster	DLTCLU
Delete CRG from cluster	DLTCRGCLU
Display cluster information	DSPCLUINF
Display CRG information	DSPCRGINF
End cluster node entry	ENDCLUNOD
End cluster resource group	ENDCRG
Remove cluster node entry	RMVCLUNODE
Remove CRG device entry	RMVCRGDEVE
Remove CRG node entry	RMVCRGNODE
Remove device domain entry	RMVDEVDMNE
Start cluster node	STRCLUNOD
Start cluster resource group	STRCRG
Work with cluster	WRKCLU

New commands available only with PowerHA for i

These commands are brand new in IBM i 6.1 with the license product PowerHA for i and are supplied in IBM library QHASM. These are part of the new capabilities and enhancements made by IBM to System i high availability, improving and simplifying System i Administrator effort to manage a System i high-availability environment.

Command description	
Add ASP copy description	ADDASPCPYD
Add cluster administrative domain MRE[1]	ADDCADMRE
Add cluster administrative domain node entry	ADDCADNODE

[1] MRE: Monitored Resource Entry

Change ASP copy description	CHGASPCPYD
Change ASP session	CHGASPSSN
Change cluster administrative domain	CHGCAD
Change cluster	CHGCLU
Change device description (ASP)	CHGDEVASP
Create cluster administrative domain	CRTCAD
Delete cluster administrative domain	DLTCAD
Display ASP copy description	DSPASPCPYD
Display ASP session	DSPASPSSN
End cluster administrative domain	ENDCAD
Remove administrative domain MRE	RMVCADMRE
Remove administrative domain node entry	RMVCADNODE
Remove ASP copy description	RMVASPCPYD
Start ASP session	STRASPSSN
Start cluster administrative domain	STRCAD
Work ASP copy description	WRKASPCPYD

Commands renamed

The following commands existing in i5/OS V5R4 are now renamed in PowerHA for i (the old V5R1 commands were deleted):

Command description	**Old V5R1**	**New HASM**
Change cluster configuration	CHGCLUCFG	CHGCLU
Create cluster admin domain	CRTADMDMN	CRTCAD
Delete cluster admin domain	DLTADMDMN	DLTCAD

4

High-availability technologies

This chapter describes the different solutions available on IBM System i server to implement a high-availability or disaster recovery solution. It aims to simply describe each of them and how they work. We discuss:

- Switched disk
- Geographic mirroring
- FlashCopy
- Metro mirror
- Global mirror

4.1 Introduction

The difference between a high availability and a disaster recovery solution depends on the recovery time objective (RTO) and recovery point objective (RPO) that you aim to achieve and also the distance between both systems or sites used as production and backup.

The recovery time objective is measured in minutes, hours, or seconds and is determined by the answer to this question: How long can you afford to be without your systems?

The recovery point objective is measured in minutes, hours, or seconds and is determined by the answer to this question: When it is recovered, how much data can you afford to loose or recreate?

If your needs and requirements intend to have low RTO and RPO objectives for a high-availability solution, the servers must be close, either in the same room or less than a few kilometers distance apart, allowing you to use synchronous replication.

On the other hand, if the servers are in the same room or very close to each other, we cannot consider this solution to be a disaster recovery solution. For instance, in case of a tornado or an earthquake, a distance of less than 100 km between your production and backup site is probably not enough to guarantee that all your protected applications will be able to restart on the backup site.

When choosing your high availability solution make sure that you take all your requirements into account.

4.2 Switched disk solution

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

How switched disk solutions work

There are two different ways that you can set up a switched disk environment. You can either configure it so that you are switching disks between two different systems or you can switch between logical partitions. When looking at these different options as part of a cluster environment, they actually look exactly the same. The difference between them is in how the switched disk is configured and how the hardware is connected.

Switched disk between systems

In this environment, shown in Figure 4-1, you have two different systems running IBM i in the same data center. Both of these systems have high-speed link (HSL) loop connections to a set of common towers.

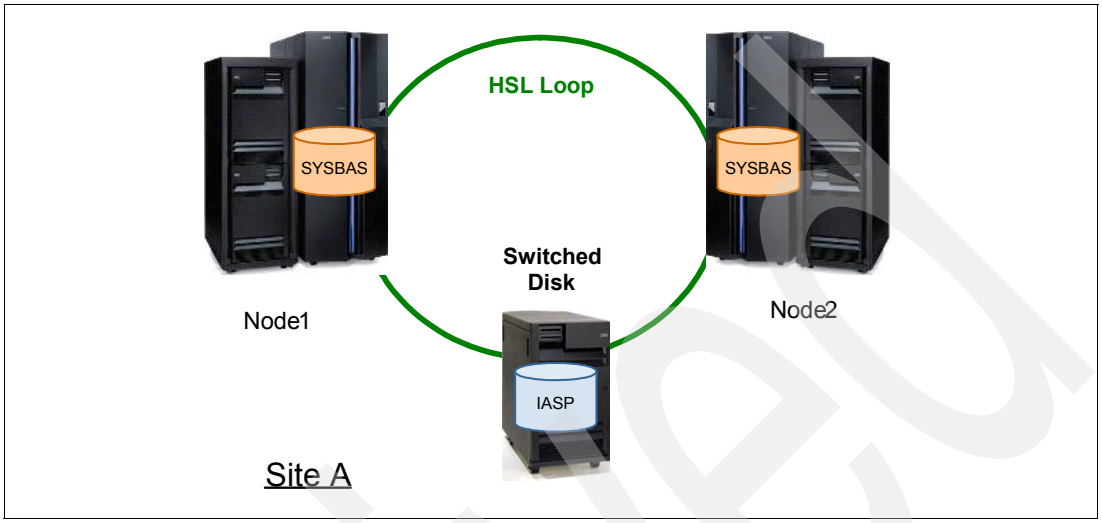


Figure 4-1 Switched disk high availability environment

When you do a switch from the primary node to the backup node in this environment, everything that is contained in the switchable towers is switched from being owned by Node1 to being owned by Node2.

Due to hardware limitations, you are limited to only being able to connect the towers to two systems. This means that in order to have more than just a primary and a single backup node you will have to use logical partitioning on at least one of the systems. If you do this, you will be able to define your cluster relationships such that you will be able to switch the hardware from the primary node to a backup node that is either a logical partition on the same system or the other system.

Note: POWER6™ servers will be the last server support for switched iASPs between servers. For strategic planning, you may want to consider switched iASPs between LPARs.

Switched disk between logical partitions

This environment essentially looks the same as the switched disk between the systems shown in Figure 4-1. The only difference between these two environments is that in this environment we are not switching the hardware between different systems, we are switching it between logical partitions on the same system.

When configuring this type of environment you need to group all of the hardware resources that will be switching 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 will be switched and which ones will not be.

4.3 Geographic mirroring solution

In this section we geographic mirroring, how it works, recommendations when using it, and additional topics related to it.

4.3.1 Overview

Geographic mirroring maintains a consistent backup copy of an independent disk pool (also called iASP, independent auxiliary storage pool) using internal or external storage at two sites to provide high availability or disaster recovery, as shown in Figure 4-2. Geographic mirroring is a subfunction of cross-site mirroring (XSM[1]) that has been implemented in i5/OS from V5R3M0.

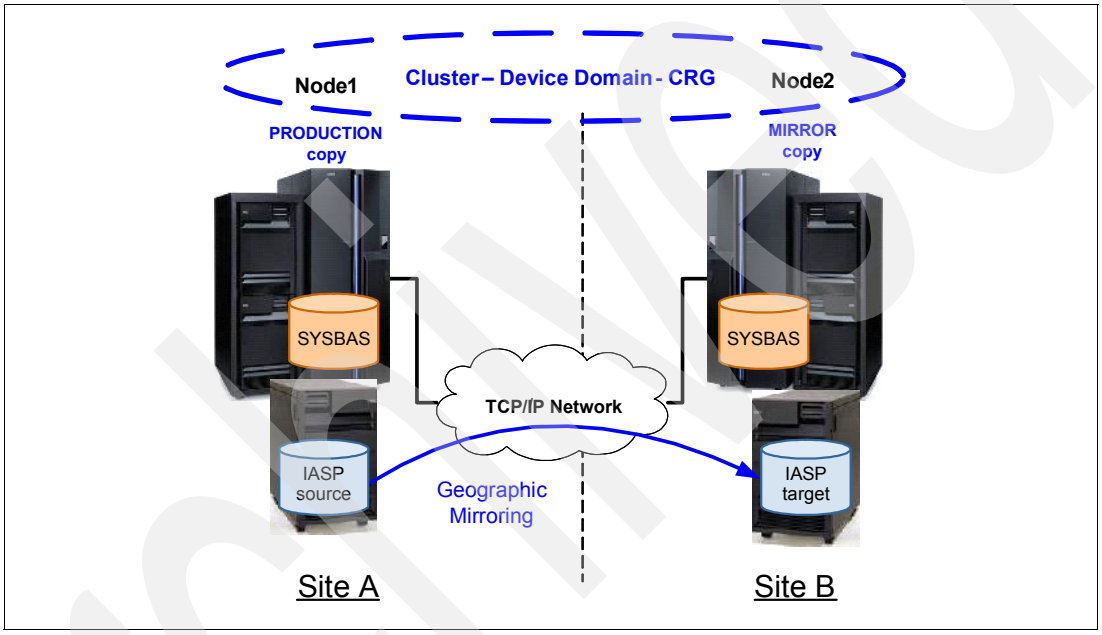


Figure 4-2 Geographic mirroring between two sites

Note: The data replication process between the two sets of disks (production copy and mirror copy), described by the blue arrow in Figure 4-2, 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).

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. Communication

[1] Cross-site mirroring is a collective term used for several different high availability technologies, including geographic mirroring, metro mirror, and global mirror. Each one of these technologies has specific tasks related to configuration.

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.

Geographic mirroring functionality involves the use of the following cluster components:

- Cluster
- Device domain
- Cluster resource group
- Cluster administrative domain
- Independent auxiliary storage pools

Important: Before implementing geographic mirroring for your high availability or disaster recovery solution, you must verify that your application can be migrated into the independent auxiliary storage pool.

4.3.2 How geographic mirroring works

In this section we describe how geographic mirroring works, the different modes that it works with (synchronous and asynchronous), and some operations that can be done with it such as suspending, configuring, and so on.

Configuring

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.

When geographic mirroring is configured, the mirror copy has the same disk pool number and name as the original disk pool, the production copy. 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 loose your high availability. If, on the other hand, 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 this would compromise your high availability environment.

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. In a high availability perspective, we recommend for redundancy purposes to use at least two different IP interfaces, connected to two different networks, if possible, that can be used for the geographic mirroring and for the cluster heartbeat function.

Synchronous mirroring mode

When geographic mirroring is active in synchronous mode, as described in Figure 4-3, the write on disk operation (#1 for journal entry) waits until the operation is complete to the disk (actually 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 in green arrows shown in Figure 4-3.

Important: The operation numbers 1–4 do not reflect the exact timing order. They are just a reference to explain this process.

The mirror copy is always eligible to become the production copy, because the order of writes is preserved on the mirror copy. We recommend trying synchronous mode first. If your performance remains acceptable, continue to use synchronous mode.

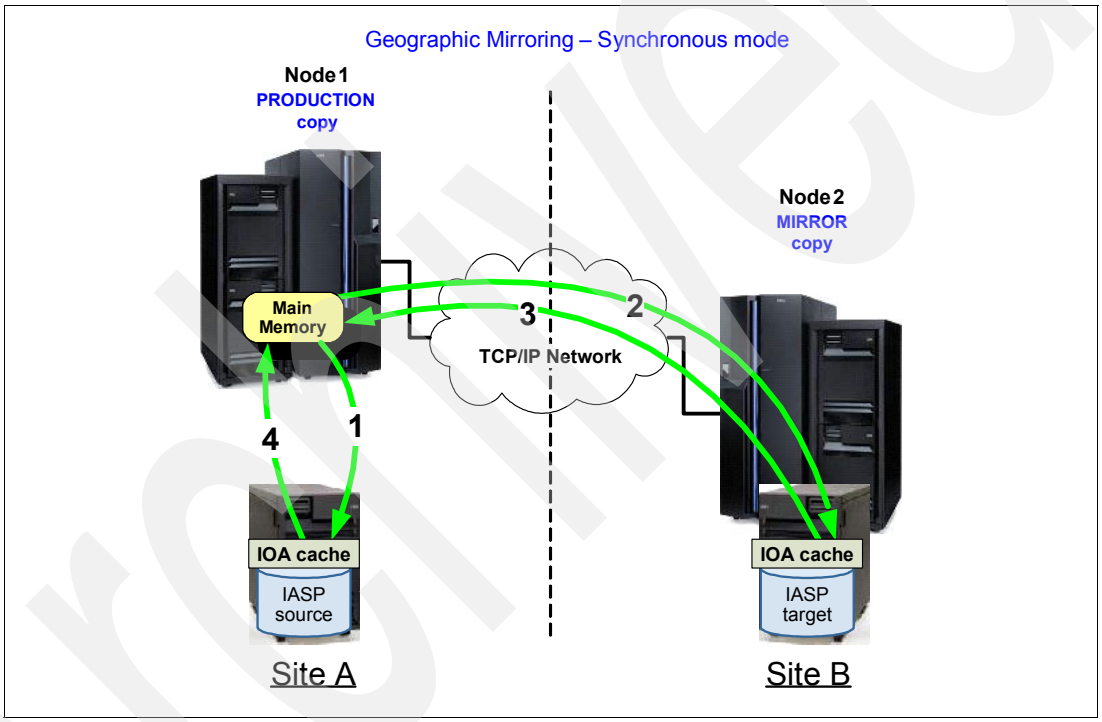


Figure 4-3 Geographic mirroring: Synchronous mode

Semi-asynchronous mode

When geographic mirroring is using asynchronous mode, the write on disk operation (operation 1 for journal entry in production copy) 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 #) only. See the operations numbered 1–4 in orange arrows shown in Figure 4-4. The physical write operation, number 5 in red in the figure, is performed later (asynchronously) to the disk on the mirror copy (target system).

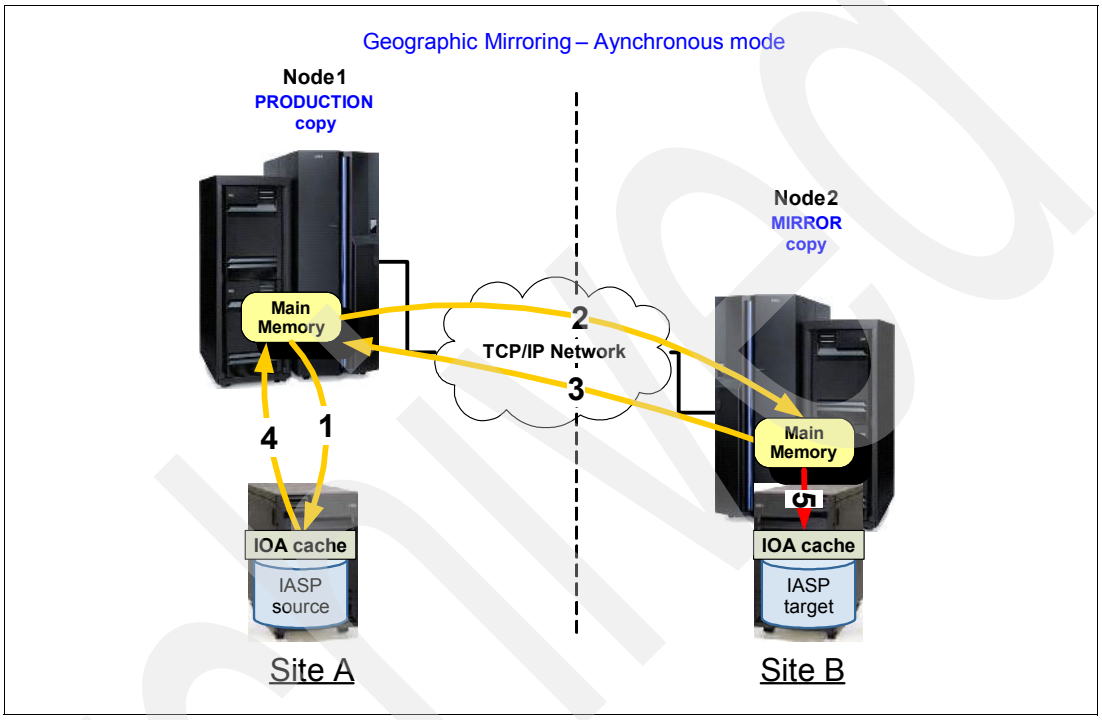


Figure 4-4 Geographic mirroring: Asynchronous mode

In asynchronous 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 may 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.

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

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. When migrating from 5.3, tracking space is automatically added to your existing iASPs. Tracking space is allocated inside of the independent ASPs. The more tracking space 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. 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.

Managing

After geographic mirroring is configured, the production copy and mirror copy run as a unit. When the production copy is made available, the mirror copy status is varied on now that geographic mirror can be performed. A full synchronization occurs when you make the disk pool available for the first time after you configure geographic mirroring. Normal operation on the production system is possible while this first synchronization is run. When geographic mirroring is active, changes to the production copy data are transmitted to the mirror copy across TCP/IP connections (data port). Changes can be transmitted either synchronously or asynchronously, as explained below.

To maintain the data integrity of the mirror copy, the user cannot access the mirror copy while geographic mirroring is being performed. If you want to perform save operations, to create reports, or to perform data mining from the mirror copy, you must detach the mirror copy in order to allow it to be made available on to the backup server. However, the mirror copy must be synchronized with the production copy after it is reattached. The synchronization will be partial or full depending on whether changes have been tracked. See “Detaching mirror copy” on page 45, for more details.

Be aware that although journal entries are not required to replicate data to the backup iASP journaling your data in the iASP is still highly recommended. This is due to the concept of single-level storage of IBM i. When the application running on the system changes an object, this change is made in main memory and acknowledged from this point to the application as considered done. It is then up to storage management to decide when to actually move these changes from main memory to disk. As geographic mirroring only replicates changed disk pages you could loose all the changes that were still in main memory at the point of a system crash. This potentially leads to a situation where data on your backup system might be inconsistent because different files were last written to disk at different points in time.

When journaling is in use, the operating system first does a physical write of the changed data into the journal receiver and then changes the corresponding data in main memory. In case of a system failure the object will be automatically recovered from the journal entries at the next vary on of the iASP or at IPL time.

Suspending geographic mirroring

You suspend geographic mirroring when you need to stop the backup system (mirror copy of the iASP), for instance, to do hardware or software maintenance on the backup site. In this case the mirror copy of the iASP is varied off and will not be used from the backup server. In suspended mode you are still not allowed to access the backup copy of the iASP.

With tracking

Starting from V5R4, if you suspend geographic mirroring with tracking, the system attempts to track changes made to those disk pools. This may reduce the synchronization process by performing partial synchronization when you resume geographic mirroring, since the system only sends what has been changed while geographic mirroring was suspended.

Note: If tracking space is exhausted, then when you resume geographic mirroring, a complete synchronization is required.

Also, if geographic mirroring is interrupted, for example, by a communications failure, the system suspends geographic mirroring automatically after the specified time out. If a tracking

space was defined for that iASP during configuration (which is the default behavior) then tracking automatically occurs.

Without tracking

If you suspend geographic mirroring without tracking changes, then when you resume geographic mirroring, a complete synchronization is required between the production and mirror copies. Complete synchronization can be a very lengthy process, anywhere from minutes to several hours, or even longer. The length of time it takes to synchronize is dependant on the amount of data in your iASP, the number of objects in the iASP, and the communication bandwidth available for geographic mirroring.

Detaching mirror copy

The detach function also suspends geographic mirroring, but it allows the mirror copy to be brought online on the backup system. If you are using geographic mirroring and want to access the mirror copy to perform save operations or data mining, or to create reports, you must detach the mirror copy from the production copy.

The user can either use the new CL command CHGASPSSN to detach the geographic mirroring (see Chapter 8, “Commands” on page 229) or the IBM Systems Director Navigator GUI by selecting **Configuration and Services** → **Disk Pools** (see Chapter 7, “Cluster Resource Services graphical user interface” on page 163 for more details) or **iSeries Navigator**.

With tracking

This function was brought to IBM i 5.4 by PTF MF40053, allowing changes made to the production and the mirror copy to be tracked while in detached status. 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 overlaid 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.

Tracking of changes to the detached mirror copy will allow the user to achieve a reasonable re-synchronization time when the detached mirror copy needs to be used for some reason (for example, backup).

With IBM i 5.4, the only way to invoke detach with tracking was to call the QYASSDMO API in your own program. You also had to vary off the production iASP before doing the detach. With IBM i 6.1, detach with tracking can be performed easier from IBM Systems Director for i5/OS or from the new CHGASPSSN command. You can detach with tracking while the production copy of iASP remains available. The new command CHGASPACT allows you to quiesce the activity of database and IFS operations against the iASP. The information still in memory will be written to the disks. See the description of this command in Chapter 8, “Commands” on page 229.

Restriction: In some cases it is possible that the information cannot be written to the disk (for example, if the page is in use by an application or by the operating system). It might therefore be better to vary off the iASP in order to be sure that all data are fully written to disk (and therefore also available on the backup system) before you detach geographic mirroring.

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. Messages are 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 may shorten the synchronization time considerably because a complete synchronization is unnecessary. In this case only pages changed on the production copy of the iASP during the time where the geographic mirroring was suspended or detached will be pushed to the mirror copy, and all the changes made in the mirror copy will be undone by getting the original information from the production copy.

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. 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.

- Recovery time-out

  In addition to synchronization priority, you can also set the recovery time-out. The recovery time-out 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 recovery time-out before suspending geographic mirroring, which allows your application to continue.

The trade-off is between blocking your application or requiring synchronization after suspending geographic mirroring. When your application is blocked for an extended time, other jobs might also be blocked waiting for resources and locks owned by the applications using the geographic mirrored disk pool. When geographic mirroring is suspended, you no longer have the protection of the mirror copy. When deciding on this parameter you should also consider whether your iASP has a tracking space. As the use of a tracking space reduces the need for full synchronization after a suspend, the recovery time-out can be shorter.

4.3.3 Requirements for geographic mirroring

Important: When using geographic mirroring to protect your application's data, the data must be stored in an independent auxiliary storage pool. You must ensure that your application can fully run in an iASP environment.

To use geographic mirroring on IBM i as a high-availability solution, the minimum hardware and software requirements are:

- Hardware requirements
  - All independent disk pool (independent auxiliary storage pool) hardware requirements must be met.
  - You must have enough disks on both systems to set up iASPs of about the same size.
  - As a general rule of thumb we recommend having one disk arm in the system ASP per three disk arms in the iASP. This is due to the fact that temporary objects are placed into the system ASP.
  - We recommend configuring a separate storage pool for jobs using geographic mirrored independent disk pools. Performing geographic mirroring from the main storage pool can cause the system to hang under extreme load conditions.
  - Geographic mirroring is performed when the disk pool is available. When geographic mirroring is being performed, the system value for the time of day (QTIME) should not be changed.
- Software requirements
  - To use advanced features of geographic mirroring with IBM i 6.1, the license program PowerHA for i (5761-HAS) must be installed.
  - To use new and enhanced functions and features of this technology, we recommend installing the most current release and version of the operating system on each system or logical partition that is participating in a high availability solution based on this technology.

    Note: For systems on the same HSL loop, see the High Availability Web site to ensure that you have compatible versions of i5/OS:

    `http://www-03.ibm.com/systems/power/software/availability/i/index.html`

  - One of the following graphical interfaces is required to perform some of the disk management tasks necessary to implement independent disk pools:
    - IBM Systems Director Navigator console for i5/OS
    - System i Operations Navigator
  - You need to install i5/OS Option 41 HA Switchable Resources. Option 41 gives you the capability to switch independent disk pools between systems. It is also required for working with high availability management interfaces, which are provided as part of the PowerHA for i licensed program.
  - To switch an independent disk pool between systems, the systems must be members of a cluster and the independent switched disk must be associated with a device cluster resource group in that cluster.

Communication requirements

When you are implementing an i5/OS high availability solution that uses geographic mirroring, you should plan communication lines so that geographic mirroring traffic does not adversely affect system performance.

Communications requirements for independent disk pools are particularly critical as they affect throughput. Up to four IP communication lines (data ports) can be configured between both source and target systems for geographic mirroring. We recommend minimum of two lines for better availability.

As said previously, for the cluster heartbeat we recommend using two different IP addresses for redundancy. These two IP addresses can be among the ones used for geographic mirroring communication.

Application performance

When implementing geographic mirroring, different factors can influence the performance of systems involved in this HA solution. In order 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.

CPU considerations

Geographic mirroring creates an additional 5 to 20% workload 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 is no formula to calculate this exactly, because 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 source iASP to target iASP.

Machine pool size considerations

Geographic mirroring also requires extra memory in the machine pool. Calculation can be done using the following formula and then using the WRKSHRPOOL command to set the machine pool size:

```
Extra Machine Pool Size = 300MB + (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.

When the system value QPFRADJ is equal to 2 or 3**,** for automatic performance adjustment, to prevent the performance adjuster reducing the machine pool size, you should set the machine pool minimum size to the calculated amount (the current size plus the extra size for geographic mirroring from the formula) using the Work with Shared Storage Pools (WRKSHRPOOL) command or the Change Shared Storage Pool (CHGSHRPOOL) command.

Otherwise, you have to set the system value QPFRADJ to zero, which prohibits the performance adjuster from changing size of the machine pool.

Storage pool considerations

Performing geographic mirroring from the main storage pool *BASE can cause the system to hang under extreme load conditions. Configure a separate storage pool for the jobs using geographic mirrored independent disk pools, especially if you specify a long recovery time out.

Disk unit 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 (like all job's temporary library QTEMP, temporary indexes built by SQL engine or QUERY/400, and so on) 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. You must also remember that the operating system and basic functions occur in the SYSBAS disk pool.

Network configuration considerations

Network configuration can potentially affect the geographic mirroring performances. The available bandwidth between the remote systems depends on distance, number, and characteristics of data ports used for geographic mirroring.

4.3.4 Recommendations when using geographic mirroring

In this section we give recommendations about using geographic mirroring.

Journaling

When you use the high availability solution based on geographic mirroring, you should start journaling of your data (files, data area, data queues). It is important to understand that on IBM i, information written by an application is first recorded on main memory, principle of the single level storage, and then the system will decide when and where it will be stored on disk. In the case of the system ending abnormally, the information still in memory is not saved and will be lost and impossible to recover.

Journal management prevents this situation. When you start journaling an object, the system will first record changes to journal receivers on disk and then change the data itself in main memory. In the case of an abnormal system end, at the next IPL or vary-on of an iASP the system will first recover the objects, checking journal receivers to apply changes that are not in the objects yet.

System value QTIME

Geographic mirroring is performed when the disk pool is available. When geographic mirroring is being performed, the system value for the time of day (QTIME) should not be changed.

Communication

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, as shown below. 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 adaptor (VIPA) can be used to define the geographic mirroring IP addresses.

Improving use of geographic mirroring

To reduce the time to vary-on independent disk pools, for either a planned switchover or when unplanned outages occur, and ensure that the restart occurs quickly and efficiently, you can use the methods discussed in this section.

Synchronizing user profile name, UID, and GID

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 actually used when looking at object ownership. To reduce the amount of internal processing that occurs during a switchover, where the independent disk pool 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 CRG. 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 switch over or failover.

Synchronization of user profiles including IUD and GID can be accomplished by using the administrative domain support.

Using recommended structure for independent disk pools

The system disk pool and basic user disk pools (SYSBAS) should contain primarily 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. Vary on and switchover times are optimized with this structure.

You should expect longer vary-on and switchover times if you have a large number of database objects residing in the system ASP because additional processing is required to merge database cross-reference information into the disk pool group cross-reference table.

Specifying a recovery time for the independent disk pool

To improve vary-on performance after an abnormal vary off, consider specifying a private customized access path recovery time specifically for that independent disk pool by using the Change Recovery for Access Paths (CHGRCYAP) command rather than by relying upon the overall system-wide access path recovery time. This will limit the amount of time spent rebuilding access paths during the vary on.

4.3.5 Combining geographic mirroring and switched disk

There is also the possibility to combine a switched disk environment with geographic mirroring, as shown in Figure 4-5. Here you would use the switched disk environment locally to achieve high availability, whereas the system using geographic mirroring could sit in a remote location and therefore help to also achieve disaster recovery. All three nodes are part of the recovery domain. Node 1 is the primary node, node 2 is the backup node 1, and node 3 is backup node 2. Should just node 1 fail, then node 2 would become the primary node (by switching over the tower containing the iASP to that node), whereas node 3 would become backup node 1. Geographic mirroring would still be active. If the complete site hosting node 1 and node 2 should fail, then node 3 would become the primary node and work could continue there.

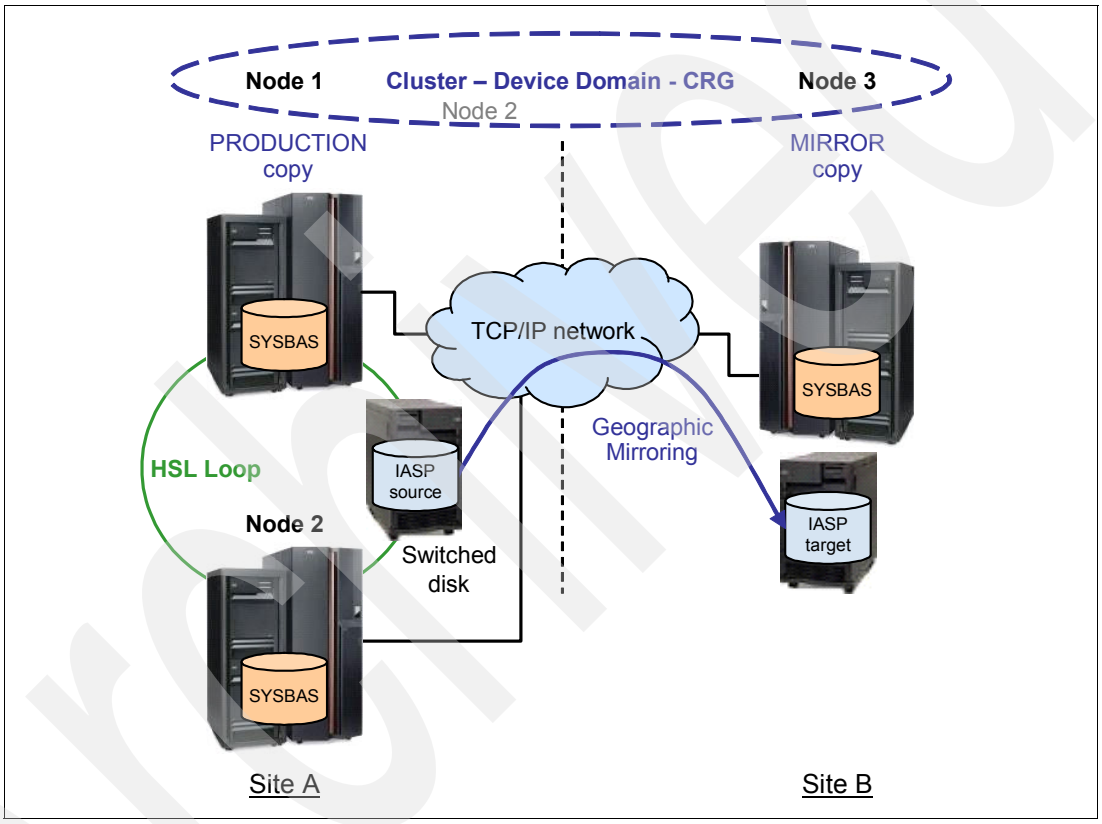


Figure 4-5 Combination of geographic mirroring and switched disk

4.4 FlashCopy

In this section we describe solutions using FlashCopy, a point-in-time copy of logical disk units within the same external Storage system. FlashCopy is managed by the microcode of the Storage system. Data is copied on a logical volume level.

Note: A disk unit residing on a Storage system is referred to as logical volume, volume, or logical unit number (LUN).

4.4.1 FlashCopy overview

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 (whole volume) or, when using the *nocopy* option, only those parts are really copied that are changing 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.

Within PowerHA for i, the classic FlashCopy is supported. It uses normal volumes as target volumes. This target volume has to have the same size (or larger) as the source volume, and that space is allocated in the storage subsystem. There is also a version of FlashCopy called space efficient FlashCopy (FlashCopy SE), which is not currently supported on the PowerHA product.

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.

Next we discuss the basic characteristics of a FlashCopy operation.

Establish 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 4-6.

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.

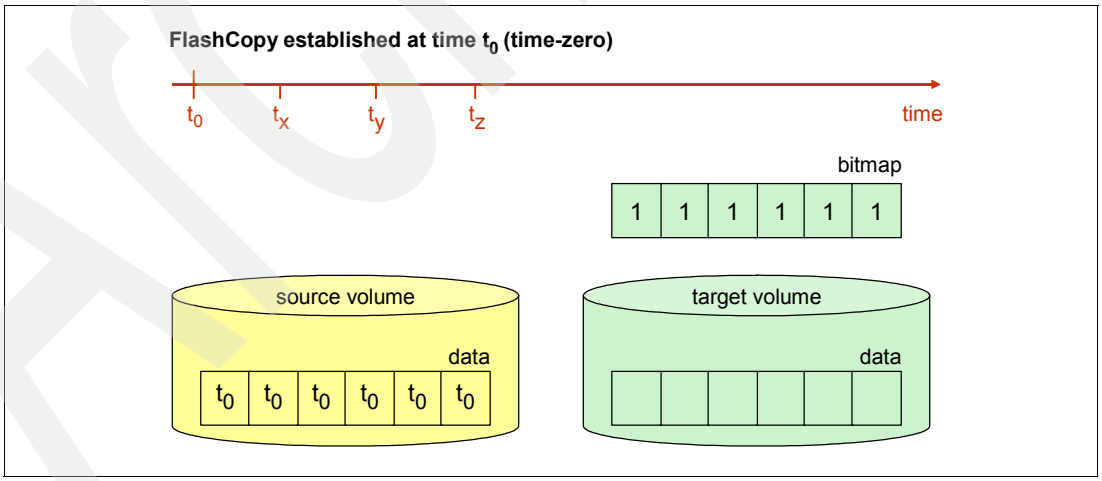


Figure 4-6 FlashCopy at time t0

Once 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 4-7.

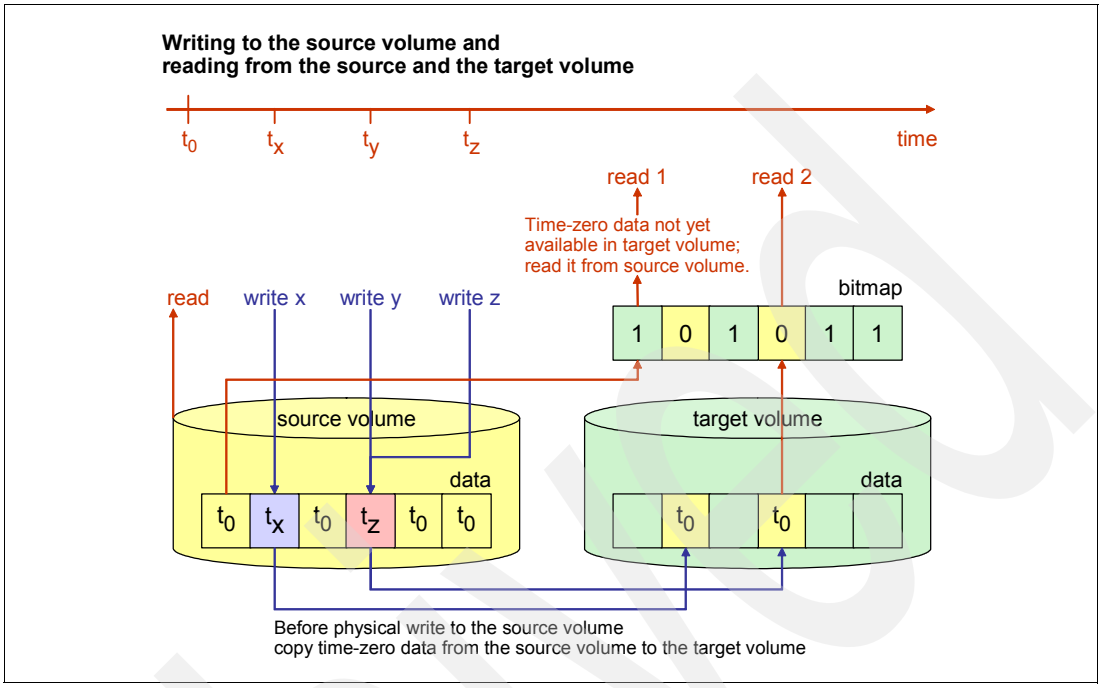


Figure 4-7 Reads from source and target volumes and writes to source volume

Reading from the source

The data is read immediately, as can be seen in Figure 4-7.

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 done.

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, as illustrated in Figure 4-7.

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. The concept of writes to the target volume is illustrated on Figure 4-8.

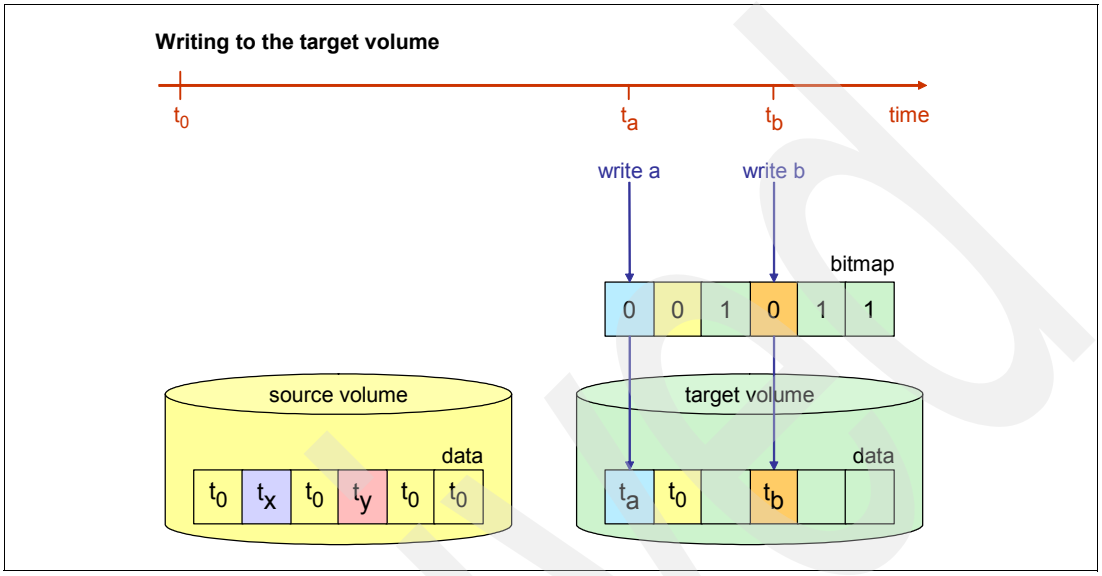


Figure 4-8 Writes to target volume

Terminating the FlashCopy relationship

The FlashCopy relationship is *automatically ended* when all tracks have been copied from the source volume to the target volume. 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 4-6 on page 52, Figure 4-7 on page 53, and Figure 4-8.

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.

The *nocopy* option is default for FlashCopy SE.

Note: For more information about external storage with i5/OS refer to *IBM i and IBM System Storage: A Guide to Implementing External Disks on IBM i*, SG24-7120.

More information about FlashCopy can be found in:

- *IBM System Storage DS8000: Copy Services in Open Environments*, SG24-6788
- *IBM System Storage Copy Services and IBM i: A Guide to Planning and Implementation*, SG24-7103

4.4.2 FlashCopy and PowerHA for i

The PowerHA for i product allows a FlashCopy of an iASP. The critical data resides in an iASP that is made up of logical volumes on external storage. FlashCopy is used to provide a point-in-time copy of only the data in the iASP. The copy of the iASP can be varied on to another partition.

A solution with FlashCopy provides minimal downtime when saving IBM i objects to tape, or provides an environment for testing applications, which can be regularly updated by doing a FlashCopy of application data.

The setup for this solution consists of two IBM i partitions. Usually, both partitions are defined in the same physical system, but they could also reside in two separate units.

The two partitions are grouped in a cluster. The partition that runs your production application is called the production partition. The other partition is called the backup partition. The backup partition is typically used to run test or development of non-production applications, or to perform data backup operations from there without interrupting the production environment. Both partitions are connected to the same Storage system. The production application runs in an iASP that resides in the Storage system and contains FlashCopy source volumes.

To take a backup of the application using the backup partition:

1. Quiesce the application data in iASP and suspend database transactions.
2. Perform a FlashCopy of the iASP.
3. Resume database activity on the production iASP.
4. Vary on the iASP FlashCopy targets in the backup partition.
5. Use the backup partition to save the data to tape without impacting the production partition.

During the backup operation, the production application continues to run in the production partition.

See Figure 4-9 for an illustration of this setup.

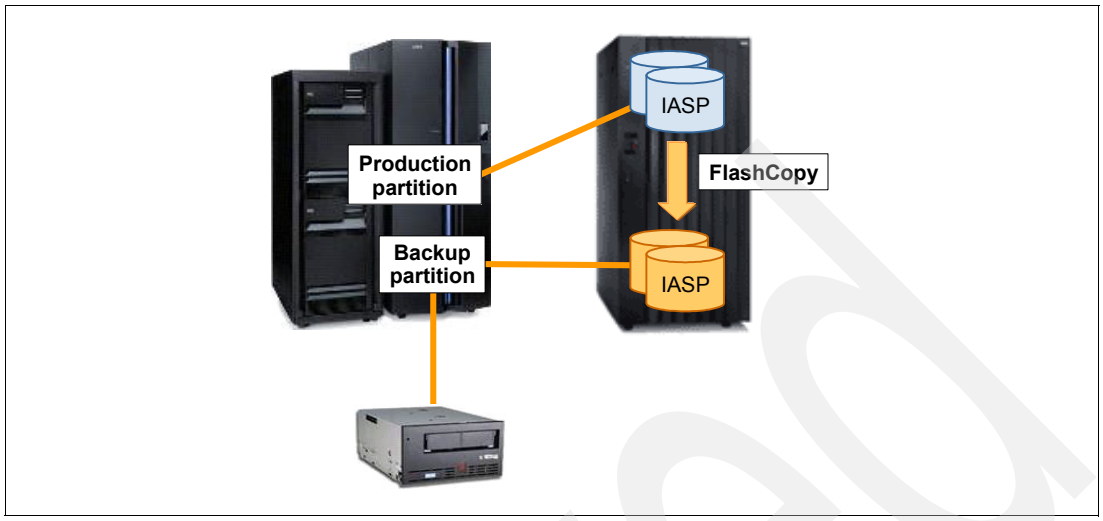


Figure 4-9 FlashCopy of iASP

4.4.3 Planning and requirements

The following points should be taken into account when planning for this solution:

- The application must be established in an iASP before implementing this solution. You can engage IBM services to set up your application for iASP.
- If the implementation is done using PowerHA for i, IOPs and Fibre Channel attachment cards can be shared between iASP and SYSBAS. However, you should be aware that the IOP will be reset/reloaded when performing the STRASPSSN command. Reset/reload of an IOP can prevent I/O operations from SYSBAS for some minutes.
- For both PowerHA on i and the IBM Copy Services toolkit, each iASP must have its own IOPs and Fibre Channel attachment cards.
- A FlashCopy license must be purchased for the Storage system using this solution.
- IBM i Option 41 HA Switchable Resources must be installed in both the production and the backup partition.
- You need to install IBM i 6.1 for implementing this solution with POWER™ HA for i. If you are using a previous version of IBM i use the IBM i Copy Services Toolkit.

Important: When using Copy Services functions such as metro mirror, global mirror, or FlashCopy for the replication of the load source unit or other IBM i disk units within the same Storage system or between two or more IBM Storage systems, the source volume and the target volume characteristics must be identical. The target and source must be of matching capacities and matching protection types.

4.4.4 Combining geographic mirroring and FlashCopy

There is also the possibility to combine geographic mirroring with FlashCopy, as shown in Figure 4-10. Here you would use geographic mirroring for either high availability or disaster recovery and the FlashCopy to decrease planned downtime for saves. All three nodes are in the cluster, but only node 1 and node 2 are in the recovery domain. Node 1 is the primary node and node 2 is the backup node. Node 1 is the source for FlashCopy and node 3 is the target for the FlashCopy.

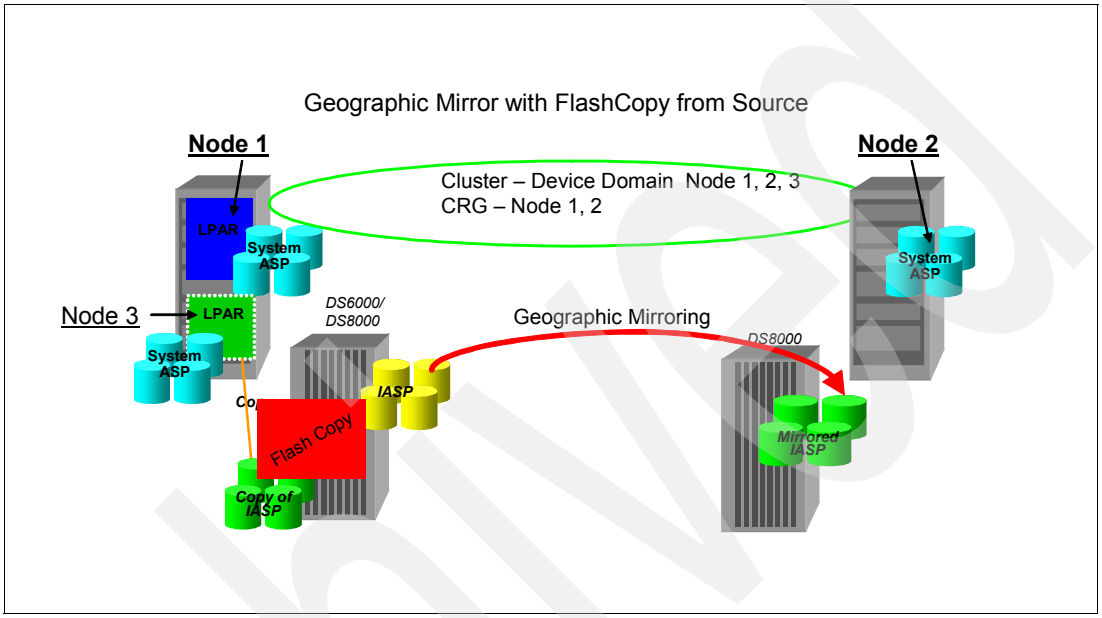


Figure 4-10 Geographic mirror with FlashCopy from source

The steps to take a backup of the production node are the same as for standalone FlashCopy:

1. Quiesce the application data in iASP via the CHGASPACT *SUSPEND and suspend database transactions on node 1.
2. Perform a FlashCopy of the iASP from node 3.
3. Resume database activity on the production iASP from node 1.
4. Vary on the iASP FlashCopy targets in the target partition from node 3.
5. Use the backup partition to save the data to tape without impacting the production partition from node 3.

Note: The Copy Services Toolkit can be used to automate the process so all steps could be from node 3.

See Figure 4-10 for an illustration of this setup.

The steps are different when taking a flash of the mirror copy target system. You first need to detach with tracking the target system from the production systems. See Figure 4-11.

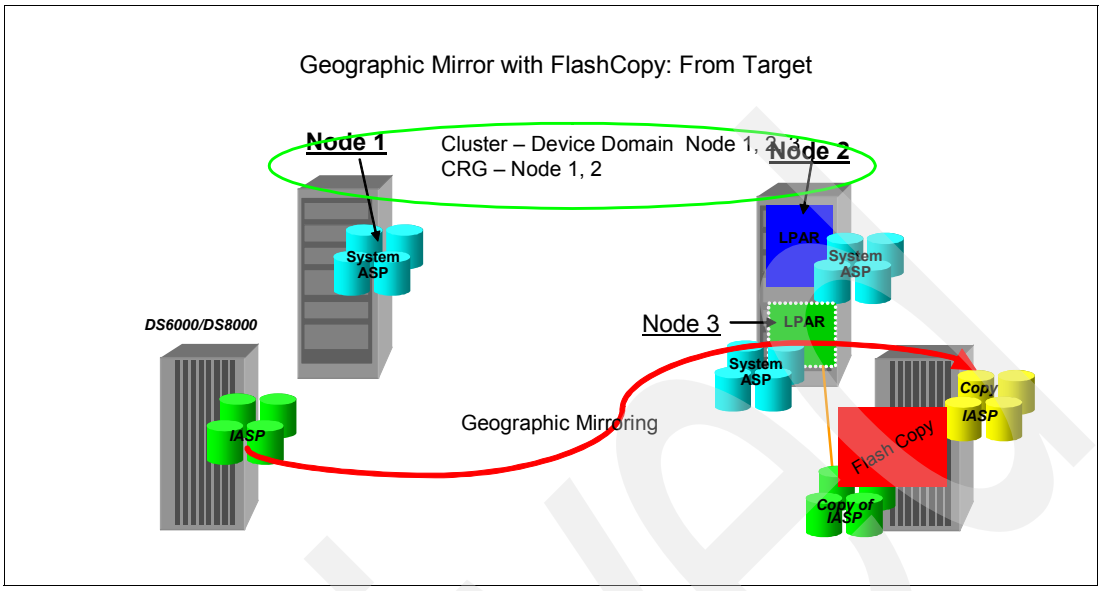


Figure 4-11 FlashCopy of the mirror copy on the target system

To do a FlashCopy of the mirror copy on the target system:

1. Quiesce the application data in iASP and suspend database transactions on node 1.
2. Detach with tracking mirrored copy node 2 from node 1.
3. Perform a FlashCopy of the iASP attached to node 2 from node 3.
4. Resume database activity on the production iASP from node 1 (automatically performs a resych to sync up node 1changes to node 2).
5. Reattach node 2 iASP to the geographic mirror.
6. Vary on flashed iASP to node 3.
7. Vary on the iASP FlashCopy targets in the backup partition from node 2.
8. Use the backup partition to save the data to tape from node 3 without impacting the production partition.

Note: The Copy Services toolkit can be used to automate the process so that all steps could be done from node 3.

V6R1M0 PTF MF45308 is required to support a combined geographic mirror and FlashCopy environment.

4.5 Metro mirror

In this section we describe high availability solutions with metro mirror-synchronous replication between local and remote external IBM Storage systems (Storage systems) connected to IBM i partitions in a SAN. Metro mirror is entirely managed by IBM Storage systems. Data is replicated on the disk level using a Fibre Channel connection between the local and remote sites.

Note: For more information about external storage with i5/OS refer to *IBM i and IBM System Storage: A Guide to Implementing External Disks on IBM i*, SG24-7120.

For more information about metro mirror refer to:

- *IBM System Storage DS8000: Copy Services in Open Environments*, SG24-6788
- *IBM System Storage Copy Services and IBM i: A Guide to Planning and Implementation*, SG24-7103

4.5.1 Metro mirror overview

Metro mirror provides real-time replication of logical volumes between two Storage systems that can be located up to 300 km from each other. It is a synchronous copy solution where write operations have to be completed on both copies (local and remote site) before they are reported back to the operating system as being finished. Metro mirror is typically used for applications that cannot suffer any data loss in the event of a failure.

As data is synchronously transferred, the distance between the local and the remote Storage systems will determine the effect on application response time. Figure 4-12 illustrates the sequence of a write update with metro mirror.

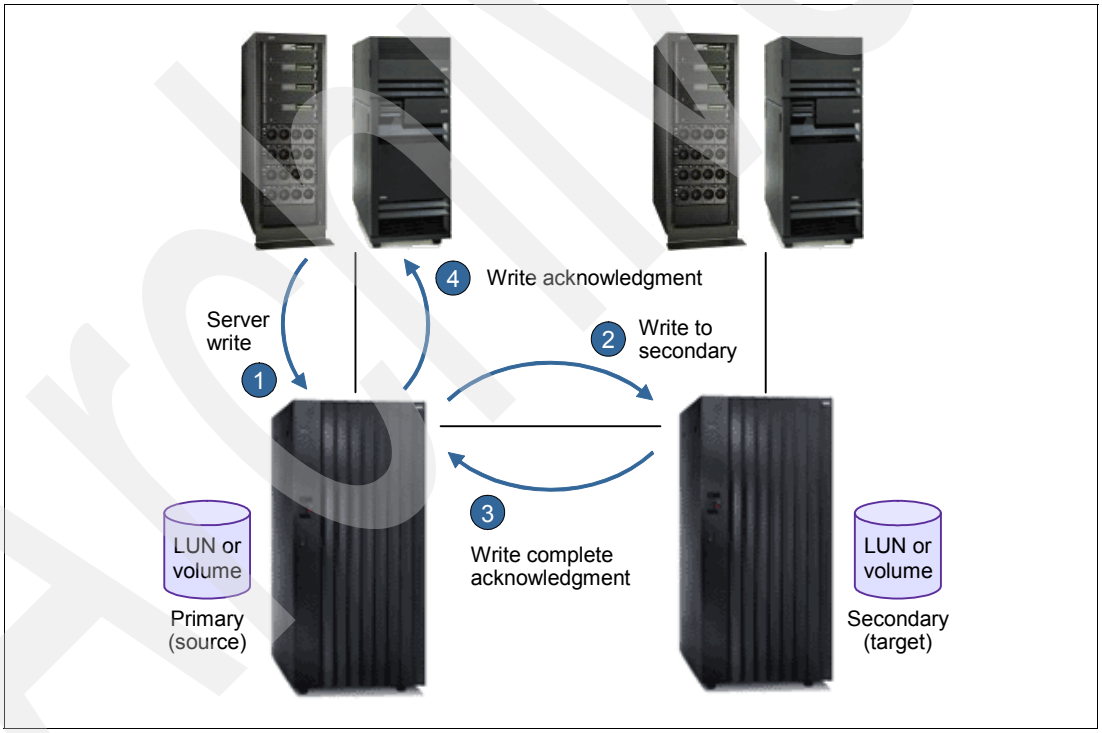


Figure 4-12 Metro mirror

Note: A disk unit residing on a Storage system is referred to as logical volume or logical unit number.

When the application performs a write update operation to a source volume, this is what happens:

1. Write to source volume (DS8000 cache and NVS).
2. Write to target volume (DS8000 cache and NVS).
3. Signal write complete from the remote target DS8000.
4. Post I/O complete to host server.

The Fibre Channel connection between the local and the remote Storage systems can be direct, through a switch, or through other supported distance solutions (for example, Dense Wave Division Multiplexor or DWDM).

Note: To achieve the best performance we recommend no more than 50 km distance between local and remote IBM Storage systems when implementing metro mirror with IBM i.

4.5.2 Basic metro mirror operation and options

In this section we briefly discuss basic metro mirror operations.

Establish a metro mirror pair

This operation establishes the remote copy relationship between a pair of volumes, the source (or local), and the target (or remote) that normally reside on different IBM Storage systems. Initially, the volumes will be in *simplex* state, and immediately after the pair is established they transition to the *copy pending* state. After the data on the pair has been synchronized (both volumes have the same data), the state of the pair becomes *full duplex*.

Suspend metro mirror pair

This operation stops copying data to the target and the pair transitions to the suspended state. Because the source Storage system keeps track of all changed tracks on the source volume, you can resume the copy operations at a later time.

Resume metro mirror pair

This operation resumes a metro mirror relationship for a volume pair that was suspended and restarts transferring data. Only modified tracks are sent to the target volume because the Storage system keeps track of all changed tracks on the source volume after the volume becomes suspended.

Terminate metro mirror pair

This operation ends the metro mirror relationship between the source and target volumes.

Failover and failback

The metro mirror failover and failback modes are designed to help reduce the time required to synchronize metro mirror volumes after switching between the production and the recovery sites.

In a typical metro mirror environment, processing will temporarily switch over to the metro mirror remote site upon an outage at the local site. When the local site is capable of resuming production, processing will switch back from the remote site to the local site.

At the recovery site, the metro mirror failover function combines into a single task the following steps: Terminate the original metro mirror relationship and turn the target volume to *source suspended* state. The state of the original source volume at the normal production site

is preserved. The state of the target volume (source suspended state) permits I/O operations to the volumes. At the same time a bitmap is established on each production and remote site to record changed tracks so that only the changed tracks will be propagated back to the production site after the outage is over.

To initiate the switchback to the production site, the metro mirror *Failback* function, at the recovery site, checks the preserved state of the original source volume at the production site to determine how much data to copy back. Then either all tracks or only out-of-sync tracks are copied, with the original source volume becoming a target volume.

4.5.3 Metro mirror with PowerHA for i

Within the PowerHA for i product, metro mirror can be used to replicate the data within an iASP to a remote site.

The setup scenario for this solution consists of a local IBM i partition and a remote IBM i partition. Both partitions are grouped in a cluster. Each partition is typically connected to its own Storage systems. This solution requires that volumes belonging to the iASPs in both the local and remote partition reside on external storage, while the system ASP in each partition can contain internal or external disk units.

A metro mirror relationship is established between the logical volumes (disk units) contained in the iASP on the local Storage system and logical volumes in another, remote Storage system to which the remote partition has access. Figure 4-13 illustrates the setup for metro mirror of an iASP.

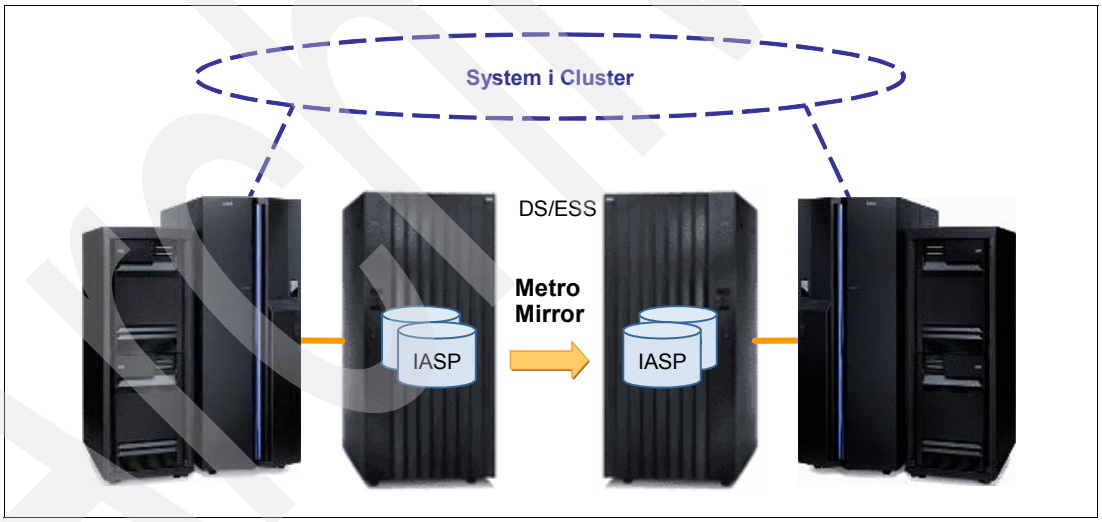


Figure 4-13 Metro mirror of iASP

This solution provides High Availability in case of planned and unplanned outages.

For each planned or unplanned outage, the metro mirror copy of the local iASP is made available to the remote partition, which continues to run production applications from the data in the remote copy of iASP.

When the planned outage is finished or the cause of failure in unplanned outage has been fixed, a reverse of the metro mirror direction (from remote logical volumes to local logical volumes) is performed to transfer updated data in the remote iASP back to original production partition. When the transfer is complete, the metro mirror relationship can be established

again in the original direction (from local to remote volumes), and production can resume at the local site.

4.5.4 Planning

We recommend that you take into account the following points when planning for this solution:

- This solution requires that critical application data reside in an iASP. Therefore, the application must be implemented in an iASP before establishing metro mirror.
- Plan for the Storage system on both local and remote sites. A Fibre Channel connection for metro mirror should be established between the two IBM Storage systems.

  Important: When using Copy Services functions such as metro mirror, global mirror, or FlashCopy for the replication of the load source unit or other IBM i disk units within the same Storage system or between two or more IBM Storage systems, the source volume and the target volume characteristics must be identical. The target and source must be of matching capacities and matching protection types.

- Careful sizing of local Storage system, remote Storage system, and Fibre Channel connection is required to achieve the best performance.

  For sizing guidelines refer to:

  – *IBM i and IBM System Storage: A Guide to Implementing External Disks on IBM i*, SG24-7120

  – *IBM System Storage Copy Services and IBM i: A Guide to Planning and Implementation*, SG24-7103
- The solution requires both the local and remote IBM i partitions to be grouped in a cluster. Therefore, plan for a cluster connection between the two systems and for IBM i software for clustering (5761-SS1 Option 41).
- If you plan to implement the solution with IBM i Copy Services Toolkit consider the following:

  – Solutions that use IBM i Copy Services Toolkit cannot be sold to the customer without prior approval by IBM Systems and Technology Group Lab Services. You can contact them via the following internet page:

    `http://www-03.ibm.com/systems/services/labservices/`

  – Take into account software requirements for IBM i Copy Services Toolkit, which are listed in the toolkit documentation.

4.6 Global mirror

In this section we describe a high availability solution with global mirror-asynchronous replication between a local and a remote external Storage system connected to IBM i partitions in SAN. Similar to metro mirror, global mirror is also managed by Storage systems and data is replicated on the disk level. Global mirror is typically used for replication on long distances in the network based on an FCP transport technology or on an IP-based network.

Note: For more information about external storage with i5/OS refer to *IBM i and IBM System Storage: A Guide to Implementing External Disks on IBM i*, SG24-7120.

For more information about global mirror refer to:

- *IBM System Storage DS8000: Copy Services in Open Environments*, SG24-6788
- *IBM System Storage Copy Services and IBM i: A Guide to Planning and Implementation*, SG24-7103

4.6.1 Functions used in global mirror

Global mirror is based on the following copy services functions of a Storage system:

- Global Copy
- FlashCopy
- FlashCopy consistency group

This preserves data consistency. In this section we shortly describe each of these.

Global Copy

Global Copy is a non-synchronous remote copy function for longer distances than are possible with metro mirror. Global Copy is appropriate for remote data migration, and in some server environments also for off-site backups and transmission of inactive database logs, at virtually unlimited distances.

With Global Copy, write operations complete on the source (local) Storage system before they are received by the target (remote) Storage system. This capability is designed to prevent the local system's performance from being affected by wait time from writes to the remote system. Therefore, the source and target copies can be separated by any distance.

FlashCopy

For more information about FlashCopy refer to 4.4.1, "FlashCopy overview" on page 52.

Dependant writes and data consistency

In a server application environment, we define *dependant writes* as follows: If the start of one write operation is dependent upon the completion of a previous write, the writes are dependent. An example of dependant writes in IBM i is writing journals and database records. A journal must be written to disk prior to the relevant updated database record.

Many applications require that their data are in a consistent state in order to begin or continue processing. In general, consistency implies that the order of dependent writes is preserved in the copy of data. In the metro mirror environment, keeping *data consistency* means that the order of dependent writes is preserved in all the metro mirror target volumes. But with Global Copy sequence of dependent writes may not be respected at the recovery site. We refer to such data at a recovery site as fuzzy data.

FlashCopy consistency group

Note: Distinguish between the FlashCopy consistency group and metro mirror consistency group. Here we describe the FlashCopy consistency group since it is employed in global mirror.

When FlashCopy is established with a FlashCopy consistency group, the following activities are performed when the FlashCopy relationship is created: As soon as FlashCopy is established for a source volume, the DS8000 holds off I/O activity to that volume for a time period by putting the volume in a *queue full* state. I/O activity resumes when FlashCopies of all volumes are established. Therefore, a time slot can be created during which dependent write updates do not occur, and FlashCopy uses that time slot to obtain a consistent point-in-time copy of the related volumes.

4.6.2 How global mirror 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.

Note: In a global mirror environment we usually refer to volumes on a local site as A volumes, Global Copy target volumes on a remote site as B volumes, and FlashCopy target volumes on a remote site as C volumes.

Establish Global Copy

A Global Copy relationship is established between the source volume on the local site and the target volume on the remote site. At this time data starts to be copied from the source volume to the target volume. After a first complete pass through the entire A volume, Global Copy will constantly scan through the out-of-sync bit map. This bitmap indicates changed data as it arrives from the applications to the source disk subsystem. Global Copy replicates the data from the A volume to the B volume based on this out-of-sync bit map.

Global Copy is an asynchronous process and does not immediately copy the data as it arrives from the host system to the A volume. Instead, as soon as a track is changed by an application write I/O, it is reflected in the out-of-sync bitmap as with all the other changed tracks. There can be several concurrent replication processes that work through this bitmap, thus maximizing the utilization of the high bandwidth Fibre Channel links.

At this point data consistency does not yet exist at the remote site.

FlashCopy

FlashCopy is an integral part of the global mirror solution, and now it follows as the next step in the course of establishing a global mirror session (Figure 4-14).

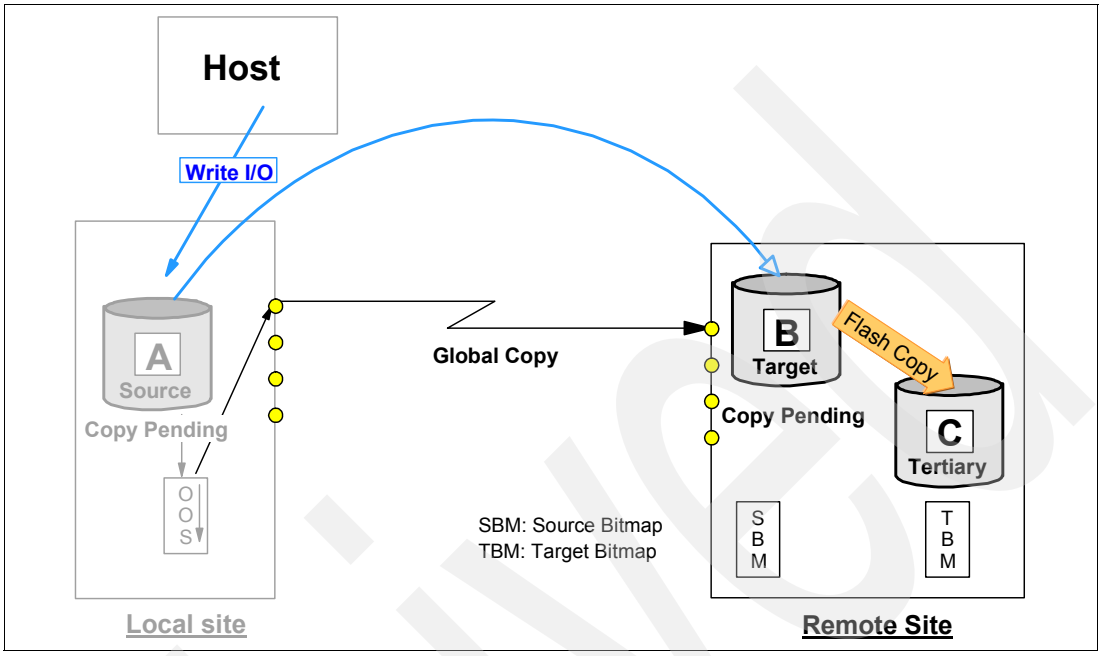


Figure 4-14 FlashCopy within global mirror

Figure 4-14 shows a FlashCopy relationship with a Global Copy target volume as the FlashCopy source volume. Volume B is now both at the same time a Global Copy target volume and a FlashCopy source volume. In the same disk subsystem is the corresponding FlashCopy target volume.

Global mirror session

Defining a *global mirror session* creates a kind of token, which is a number between 1 and 255. This number represents the global mirror session.

When the global mirror session is started, it triggers events that involve all the volumes within the session. This includes very fast bitmap management at the local storage disk subsystem, issuing inband FlashCopy commands from the local site to the remote site, and verifying that the corresponding FlashCopy operations successfully finished. This all happens at the microcode level of the related Storage system that are part of the session, fully transparently, and in an autonomic fashion from the users' perspective.

All C volumes that belong to the global mirror session comprise the consistency group.

To achieve the goal of creating a set of volumes at a remote site that contains consistent data, asynchronous data replication alone is not enough. It must be complemented with either a kind of journal or a tertiary copy of the target volume. With global mirror this third copy is naturally created through the use of FlashCopy.

The microcode automatically triggers a sequence of autonomic events to create a set of consistent data volumes at the remote site. We call this set of consistent data volumes a consistency group.

Once the consistency group is created, then a FlashCopy is triggered with volume B as the FlashCopy source and volume C as the FlashCopy target. The FlashCopy only needs to copy

data that was changed since the last FlashCopy operation. Immediately after the FlashCopy process is logically complete, the Global Copy process continues until the next consistency group creation process is started.

4.6.3 Global mirror with PowerHA for i

The scenario for this solution consists of a local IBM i partition and a remote IBM i partition grouped in a cluster. Critical application data reside in an iASP in the local partition, the iASP containing volumes on a Storage system. A global mirror relation for the iASP volumes is established with a remote Storage system to which the remote partition has access. In other words, the global mirror secondary volumes (copy of the production iASP volumes) can be varied on for the remote partition.

Note that it is only necessary that the iASPs on both partitions (local and remote) reside on external storage. Still, the system ASP for each partition can be on internal disks or on external disk systems. The solution is illustrated on Figure 4-15.

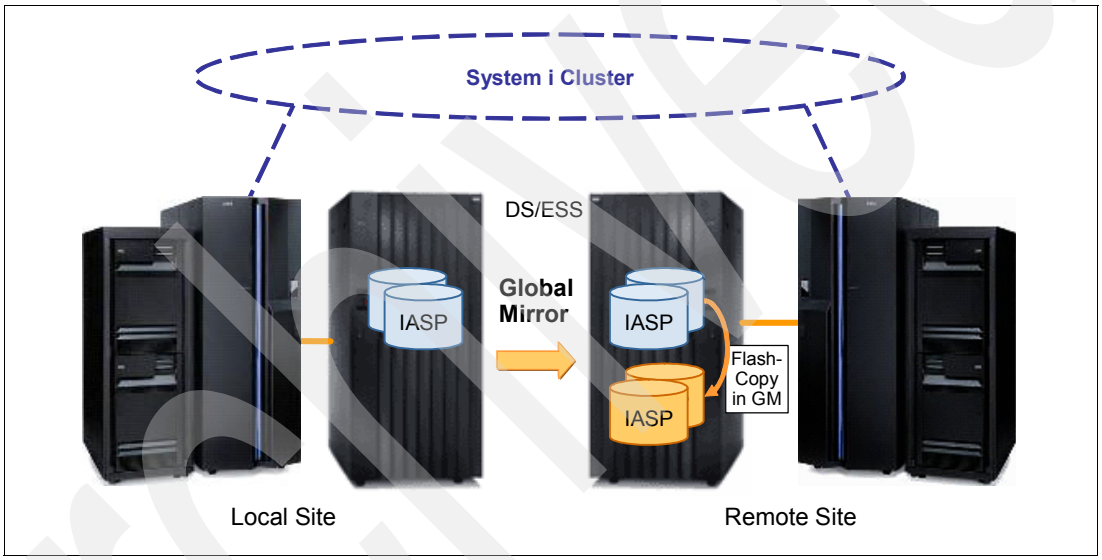


Figure 4-15 Global mirror of an iASP

The high availability solution can be managed via the PowerHA for i product, using either a command interface or the Cluster Resource Services GUI.

4.6.4 Planning and requirements

When clients plan for global mirror, their principal considerations are recovery point objective and needed bandwidth between local and remote Storage systems. Expectations about RPO should be well understood, and careful sizing is needed to provide enough bandwidth to achieve the expected RPO.

For sizing global mirror links in an IBM i environment, follow the steps described in *IBM System Storage Copy Services and IBM i: A Guide to Planning and Implementation*, SG24-7103.

Other planning considerations and requirements for this solution are the same as for the solution for metro mirror of an iASP described in 4.5.4, “Planning” on page 62, with the following change: For this solution a global mirror license is needed for the local and remote Storage systems, and FlashCopy license is needed for the remote Storage system. Planning

includes the need for the same characteristics of source and target volumes as described in 4.4.3, “Planning and requirements” on page 56.

Part 2

PowerHA for i setup and user interfaces

In this part we describe the different user graphical interfaces that PowerHA for i has and we also discuss migration and sizing considerations.

This part includes the following chapters:

5

Getting started: PowerHA for i

In this chapter we provide you with all the prerequisites that you need to successfully start using IBM PowerHA for i in your environment.

5.1 PowerHA for i installation requirements

Before installing IBM PowerHA for i license product (5761-HAS) check whether:

- IBM i 6.1 is installed on all system nodes (servers or logical partitions) that will be part of your high availability or disaster recovery solution.
- IBM i Option 41 HA Switchable Resources is also installed on these nodes.

The licensed program 5761-HAS must be installed on all nodes that you want to be part of your cluster.

5.2 Current fixes

To use the many new functions of high availability you must install the licensed product PowerHA for i (5761-HAS). If this is being installed later than the rest of the system, you will want to reapply your latest CUME and Hiper PTFs in order to install PTFs for the PowerHA for i license product.

You will also want install the recommended fixes for high availability. To find this list of PTFs, go to the following Web site:

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

The first section, called “Recommended for all systems,” provides information about the latest cumulative PTF package, HIPER PTFs, and database group PTFs. The second section is called “Recommended for specific products or functions.” Select **High Availability: Cluster, IASP, XSM, and Journal**.

At the time of this publishing, the direct link was:

http://www-912.ibm.com/s_dir/slkbase.nsf/ibmscdirect/8EB5B4B734F7B89D8625742500757CEA

5.3 Tips on the different GUI interfaces

In the following sections you can find tips on the different GUI interfaces provided by the IBM Systems Director Navigator for i5/OS. These are the Cluster Resource Services GUI, the disk management GUI and the high-availability solutions Manager GUI.

5.3.1 Connectivity

For enhanced security IBM Systems Director Navigator for i5/OS provides an inactivity time out of about 20 minutes. If you do time out, you will have to sign back on, and then you may see the invalid session message, as in Figure 5-1

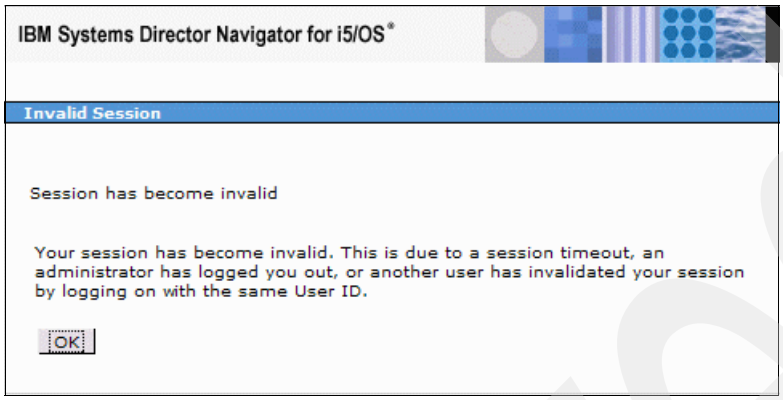


Figure 5-1 Example of a session time out

Using the refresh button will avoid this situation. Using refresh does not mean using the Web browser's refresh button, but the refresh button integrated into the interface. See Figure 5-2 for an example.

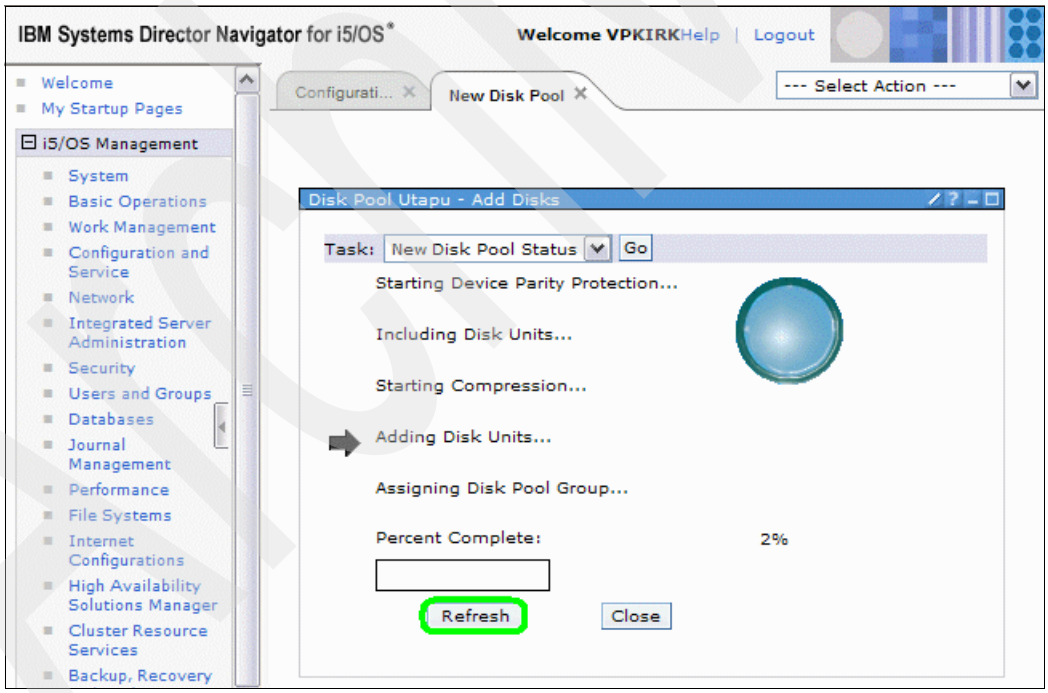


Figure 5-2 Refresh button example

5.3.2 Use the system name, not the IP address

Development recommends using the system name when using IBM Systems Director Navigator for i5/OS, not the IP address.

5.3.3 IBM Systems Director Navigator for i5/OS loops/hangs

If IBM Systems Director Navigator for i5/OS becomes nonresponsive or appears to hang, it may need to be restarted.

If a hang/loop condition occurs, the following steps will restart the environment:

1. ENDTCPSVR SERVER(*HTTP) HTTPSVR(*ADMIN)
2. STRTCPSVR SERVER(*HTTP) HTTPSVR(*ADMIN)
3. WRKACTJOB SBS(QHTTPSVR)
4. Wait for the jobs to quit taking CPU for the most part in the QHTTPSVR subsystem.

5.3.4 Cluster Resource Services GUI

Here you can find tips specific to the Cluster Resource Services GUI.

IP addresses

Development recommends using the default IP address when setting up the cluster through this interface.

If you plan to use dedicated IP addresses that are not the default for this system, it is best to use a different interface to set up the cluster, such as command line.

5.3.5 DASD GUI

In this section we provide tips about the DASD GUI.

Setting up the SST user ID connection

When enabling and accessing disk units from IBM Systems Director Navigator for i5/OS, as shown in Figure 5-3, the error shown in Figure 5-4 may occur if you have not set the proper password level for SST user profiles.

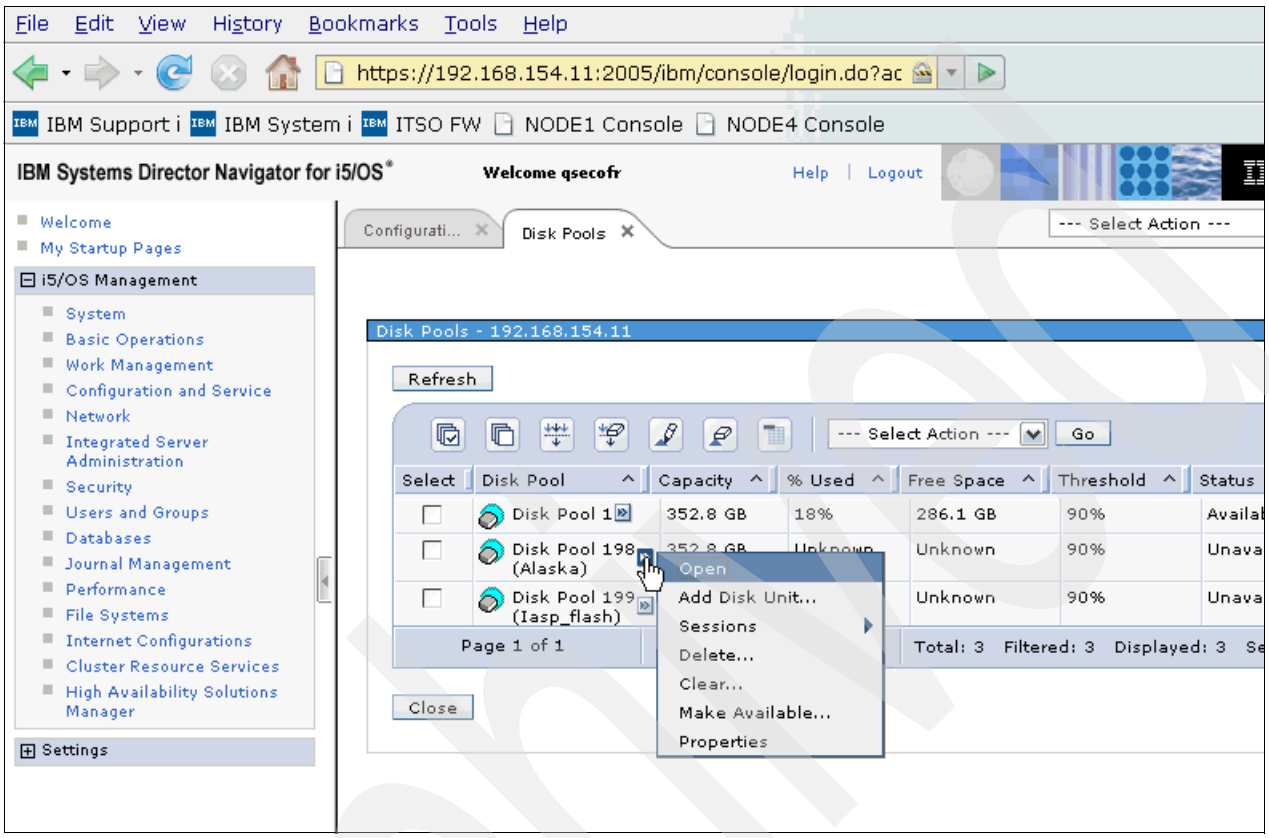


Figure 5-3 Accessing disk configuration from IBM Systems Director Navigator for i5/OS

This is because the SST enter user profile pop window is no longer displayed.

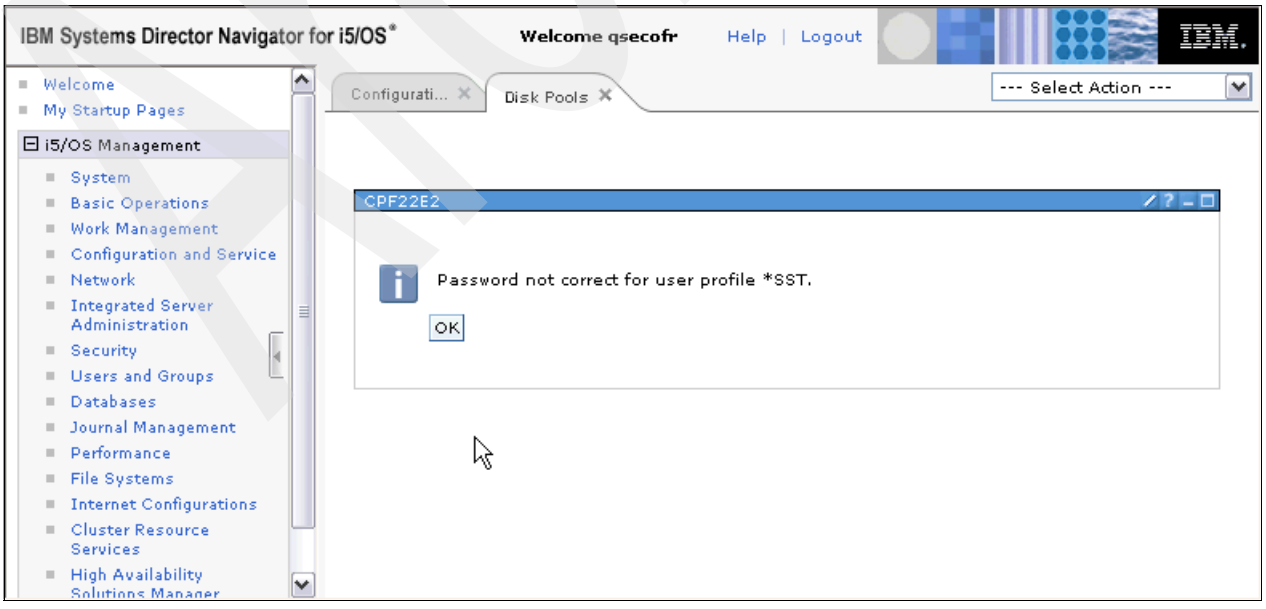


Figure 5-4 Possible error code if user IDs are not set up correctly

In this case you must follow these procedures before you can perform any disk management tasks using IBM Systems Director Navigator for i5/OS or System i Navigator in order to set up the proper authorizations for dedicated service tools (DST).

1. Ensure that the user profile that will be used to access disk units in IBM Systems Director Navigator for i5/OS has at least these authorities:
   - *ALLOBJ: All object authority
   - *SERVICE
2. Start DST from your HMC, as shown in Figure 5-5.

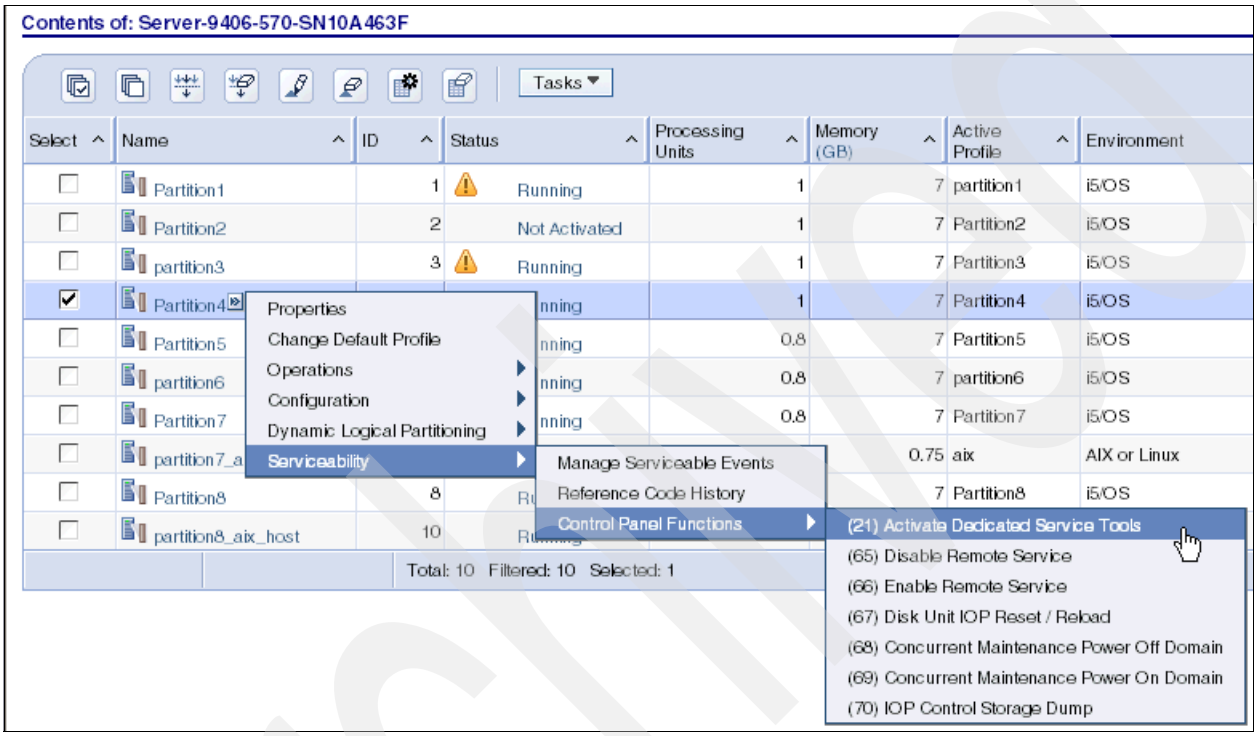


Figure 5-5 Activating DST from HMC

3. Sign on to DST using your service tools user ID and password (Figure 5-6).

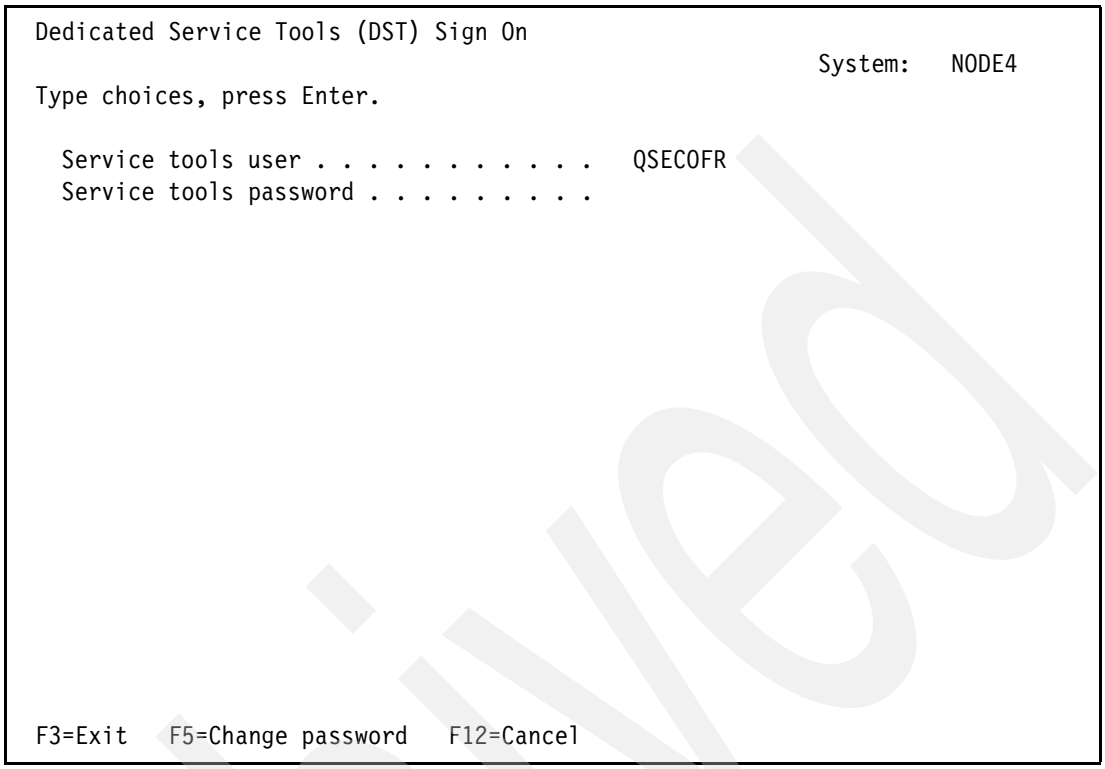


Figure 5-6 DST sign-on

4. When the Use Dedicated Service Tools (DST) display is shown, select option 5 (Work with DST environment), as shown in Figure 5-7, and press Enter.

```
Use Dedicated Service Tools (DST)
                                                          System:   NODE4
Select one of the following:

     1. Perform an IPL
     2. Install the operating system
     3. Work with Licensed Internal Code
     4. Work with disk units
     5. Work with DST environment
     6. Select DST console mode
     7. Start a service tool
     8. Perform automatic installation of the operating system

    10. Work with remote service support

    12. Work with system capacity
    13. Work with system security
    14. End batch restricted state

Selection
     5

F3=Exit   F12=Cancel
```

Figure 5-7 DST main menu

5. The Work with DST Environment menu is displayed, as shown in Figure 5-8. From the Work with DST Environment menu, select option 6 (Service tools security data).

```
Work with DST Environment
                                                        System:   NODE4
Select one of the following:

     1. Active service tools
     2. System devices
     3. Service tools user IDs

     5. Service tools device IDs
     6. Service tools security data
     7. Change service tools language









Selection
     6

F3=Exit    F12=Cancel
```

Figure 5-8 Work with DST environment

6. From the Work with Service Tools Security Data menu, select option 6 (Change password level), as shown in Figure 5-9. Ensure that the password level is set to SHA (Secure Hash Algorithm) encryption or password level 2, and press F12.

```
Work with Service Tools Security Data
                                                      System:   NODE4
Select one of the following:

     1. Reset operating system default password
     2. Change operating system install security
     3. Work with service tools security log
     4. Restore service tools security data
     5. Save service tools security data
     6. Change password level                        PWLVL 1
     7. Work with lock for device IDs from SST       Disabled
     8. Password expiration interval in days          180
     9. Maximum sign-on attempts allowed               3
    10. Duplicate password control                    18
    11. Autocreate service tools device IDs           10




Selection
     6

F3=Exit   F12=Cancel
```

Figure 5-9 Work with Service Tools Security Data

A warning message is displayed about the version of System i Navigator that has to be used with this password level, as shown in Figure 5-10.

```
Confirmation to Set Password Level
                                                        System:   NODE4
The user has selected to change the password level.

To use system service tools on workstations running System i
Navigator, the workstation needs to have Version 5 Release 1
or later of System i Access for Windows.  Before you press
Enter to set the password level on the server, ensure that
all workstations that will use system service tools are
updated.  Functions that require this update include, but
are not limited to, the scheduling functions for LPAR. For
information about how to install System i Access for Windows,
see Install System i Access for Windows topic in the System i
Information Center (www.ibm.com/systems/i/infocenter). For
information about service tools, see the service tools topic,
which is located under the security topic in the System i
Information Center.

Press Enter to confirm your choice to set password level 2.
Press F12 to return to change your choice.


F12=Cancel
```

Figure 5-10 Confirmation of set password level

7. Press F12 to return to the Work with DST Environment panel.

8. From there select option 3 (Service tools user IDs) to work with service tools user IDs, as shown in Figure 5-11.

```
Work with DST Environment
                                                        System:   NODE4
Select one of the following:

     1. Active service tools
     2. System devices
     3. Service tools user IDs

     5. Service tools device IDs
     6. Service tools security data
     7. Change service tools language









Selection
     3

F3=Exit    F12=Cancel
```

Figure 5-11 Work with DST Environment

9. 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 who is using IBM Systems Director Navigator for i5/OS. For example, if the user profile and password combination is JOSE and my1pass, then the DST user ID and password combination must be JOSE and MY1PASS, as shown in Figure 5-12.

```
Work with Service Tools User IDs
                                                          System:   NODE4
Type option, press Enter.
  1=Create             2=Change password      3=Delete
  4=Display            5=Enable               6=Disable
  7=Change privileges  8=Change description   9=Link/Remove link

Opt User ID     Description                                  Status
 1  JOSE
    DEGROFF                                                 Enabled
    KINGS        KENT KINGSLEY                              Enabled
    QSECOFR      QSECOFR                                    Enabled
    QSRV         QSRV                                       Enabled
    RONPETE      RON PETERSON                               Enabled
    VPKIRK                                                  Enabled
    11111111     11111111                                   Enabled
    22222222     22222222                                   Enabled




F3=Exit    F5=Refresh    F12=Cancel
```

Figure 5-12 Work with Service Tools User IDs

10. Use function key PF5 on the panel shown Figure 5-14 on page 85 to the change user's privilege.

```
 Create Service Tools User ID
                                                          System:   NODE4
 Service tools user ID name  . . . . . :  JOSE

 Type choices, press Enter
   Password  . . . . . .

   Set password to expire  . . . . . .  2   1=Yes, 2=No
   Description . . . . . .  Jose Goncalves - IBM STG-LS
Europe
   Linked user profile . . . . . .  JOSE

 F3=Exit   F5=Change privilege   F12=Cancel
```

Figure 5-13 Create Service Tool User ID

11. Give this service tools user ID at least these authorities:
    - Disk units – operation
    - Disk units – administration

    See Figure 5-14.

```
Change Service Tools User Privileges
                                                          System:   NODE4
Service tools user ID name  . . . . . . . . :  JOSE

Type option, press Enter.
  1=Revoke   2=Grant

Option   Functions                                          Status
  2      Disk units - operations                            Revoked
  2      Disk units - administration                        Revoked
         Disk units - read only                             Revoked
         System partitions - operations                     Revoked
         System partitions - administration                 Revoked
         Partition remote panel key                         Revoked
         Operator panel functions                           Revoked
         Operating system initial program load(IPL)         Revoked
         Install                                            Revoked
         Performance data collector                         Granted
         Hardware service manager                           Revoked
         Display/Alter/Dump                                 Revoked
         Main storage dump                                  Granted

More...
 F3=Exit   F5=Reset   F9=Defaults   F12=Cancel
```

Figure 5-14 Change Service Tools User Privileges

12. Press Enter to enable these changes.

```
Change Service Tools User Privileges
                                                          System:   NODE4
Service tools user ID name  . . . . . . . . :  JOSE

Type option, press Enter.
  1=Revoke   2=Grant

Option   Functions                                          Status
         Disk units - operations                            Granted
         Disk units - administration                        Granted
         Disk units - read only                             Revoked
         System partitions - operations                     Revoked
         System partitions - administration                 Revoked
         Partition remote panel key                         Revoked
         Operator panel functions                           Revoked
         Operating system initial program load(IPL)         Revoked
         Install                                            Revoked
         Performance data collector                         Revoked
         Hardware service manager                           Revoked
         Display/Alter/Dump                                 Revoked
         Main storage dump                                  Granted

More...
 F3=Exit   F5=Reset   F9=Defaults   F12=Cancel
```

Figure 5-15 Change Service Tools User Privileges - changed

13. Exit DST and start i5/OS.

5.4 Requirements for setting up a cluster

For setting up a cluster you must have:

- ALWADDCLU *ANY or *RQSAUT
- STRTCPSVR *INETD on both nodes in the cluster (We recommend that the *INETD server is set to autostart.)

5.5 Cluster administrative domain

If you plan to use the cluster administrative domain to synchronize user ID passwords between your nodes these are the prerequisites:

- The system value QRETSVRSEC (Retain server security data) must be set to 1 (retain data) on all nodes in the administrative domain.
- After changing the system value a user has to sign on to the system before you can add his user profile to the administrative domain. If this is not done, you will receive an error message that states that the password is not available for the user ID (CPDAA01) and that the MRE has not been added (HAE0001).
- If that profile already exists on more than one node in the administrative domain, then after changing the system value, the user has to sign on to all the nodes in the administrative domain before you can add this user profile as an MRE. If the user did not log on to all nodes after the system value was changed you will receive a message indicating an unexpected return code (CPDBB11) and the MRE will not be added.

5.6 Metro/global mirror or FlashCopy

In this section we discuss prerequisites for using metro mirror, global mirror, or FlashCopy together with PowerHA for i.

Installing DSCLI

For the PowerHA for i license program to be able to communicate with the external Storage system you need to install the DS command-level Interface (DSCLI) on all nodes in the cluster. The DSCLI software can be found at:

`http://www-304.ibm.com/systems/support/supportsite.wss/brandmain?brandind=5345868`

From this page choose **Download**, choose your type of Storage system, and choose **Downloadable Files** → **DSCLI Command Level Interface**. From the page that you are guided to you can download different releases of the DSCLI. This download contains an ISO file. You can either burn this ISO file to a CD or use a virtual CD player on your PC to access data inside the ISO image.

The ISO image contains a readme file that tells you how to install the DSCLI onto your IBM i environment. Alternatively, you can create a shortcut to the setupWIN32.EXE file on your PC. You can then edit the properties of this shortcut to the EXE file as shown in Figure 5-16. By adding " -os400" to the target you redirect this program to IBM i.

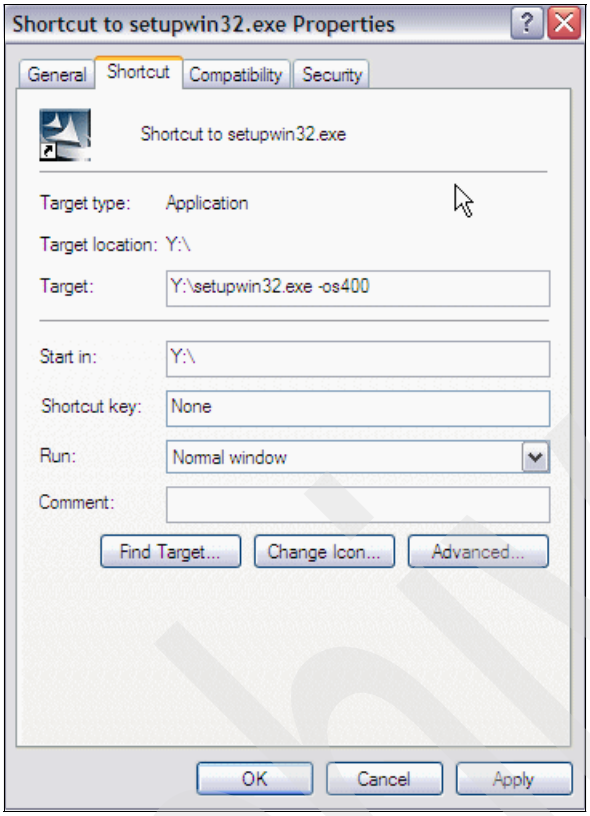


Figure 5-16 Edit shortcut to DSCLI Installer to route to IBM i

Make sure that the *FILE-host server is active on the system on which you want to install DSCLI. This can be done by issuing a NETSTAT *CNN command and looking for an as-file entry in the local port column.

Once you have done the change to the shortcut properties you can now start the shortcut by double-clicking it. This presents you with a sign-on panel, as shown in Figure 5-17.

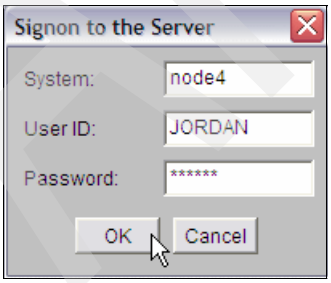


Figure 5-17 Signon panel for installing DSCLII

Clicking **OK** after filling in the required information takes you to an install wizard. Several panels appear. Accept the license agreement and all default settings. If nothing happens after you fill in the Signon panel make sure that the name that you entered for the system on which you want the DSCLI to be installed can be resolved and that the as-file is active on that system.

6

High Availability Solutions Manager GUI

This chapter describes how the PowerHA for i High Availability Solutions Manager graphical interface provides solution-based cluster management through a simplified Web-based browser.

6.1 High Availability Solution Manager GUI

The IBM PowerHA for i (previously known as HASM) is a new licensed program that provides two graphical interfaces, command-line interface, and APIs to assist administrators and programmers in configuring and managing high availability solutions. A resource trained on PowerHA for i can implement your high availability with a solution-based approach that is a graphical interface that guides you through verifying your environment, setting up, and managing your chosen solution.

Requirements

Before you can use the product make sure that the following requirements are met:

- To set up your solution using the GUI interface verify that the licensed product 5761-HAS and 5761-SS1 Option 41 (HA Switchable Resources) are installed and a valid license key exists an all systems that will be part of your high availability solution.
- Verify that the INETD server is active on all nodes in the cluster that you are going to implement. This can be verified by the presence of a QTOGINTD (User QTCP) job in the Active Jobs list of subsystem Qsyswrk. To start it run the STRTCPSVR (Start TCP/IP Server) command and specify the *INETD parameter.
- You must be signed on using the QSECOFR user profile. The password for QSECOFR must be the same across all the nodes to be added to the high availability solution. The password for the dedicated service tools (DSTs) QSECOFR must be the same as for the QSECOFR user profile, but in upper case.
- Before starting the implementation of your HA solution we recommend installing the high availability cluster, independent auxiliary storage pool (iASP), XSM, and journal recommended fixes. For further information or to order these PTFs, follow this link:

  `http://www.ibm.com/eserver/support/fixes`

6.2 HASM GUI

The High Availability Solutions Manager graphical interface provides several predefined solutions. Each solution provides a different level of high-availability coverage and has specific advantages, restrictions, and requirements.

The GUI operations are deployed via IBM Systems Director for i5/OS, a Web-based console that allows you to:

- Select a high availability solution.
- Verify requirements for your high availability solutions.
- Set up a high availability solution.
- Manage a high availability solution.

To open the interface:

1. Open a Web browser and enter the address `http://system:2001`, where *system* is the host name of the system. An alternative is `http://system:2005/ibm/console`, where *system* is the host name of the system. We have found at times that we cannot connect at one address, but can connect at the other address (Figure 6-1).
2. Log on to the system with your user profile and password.

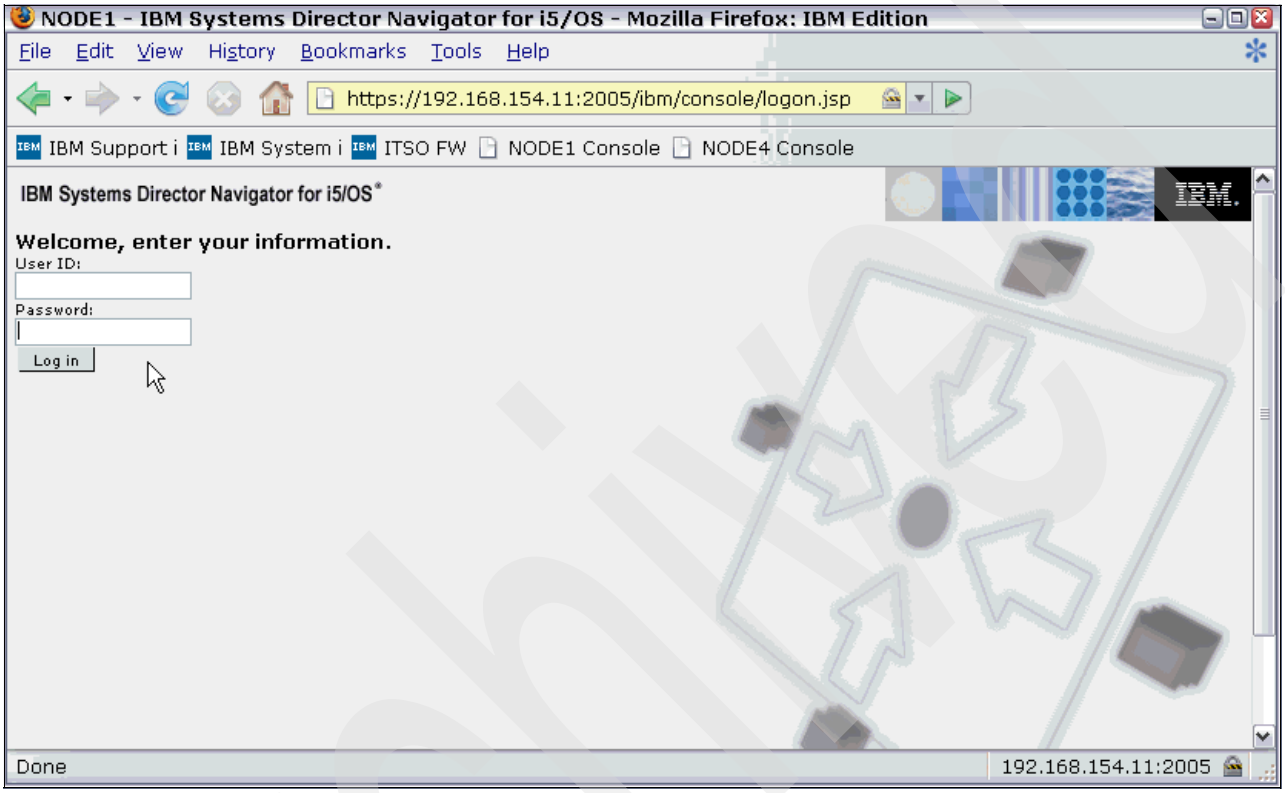


Figure 6-1 IBM Systems Director Navigator login

3. Expand **i5/OS Management** and click **High Availability Solutions Manager**, as shown in Figure 6-2.

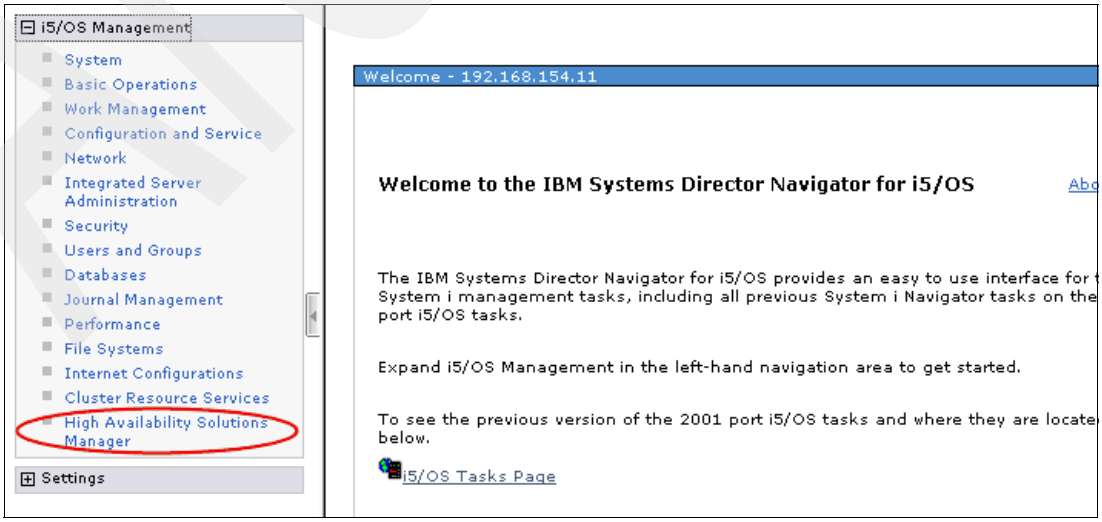


Figure 6-2 HASM Welcome page

4. Click **See How High Availability Solutions works** (Figure 6-3).

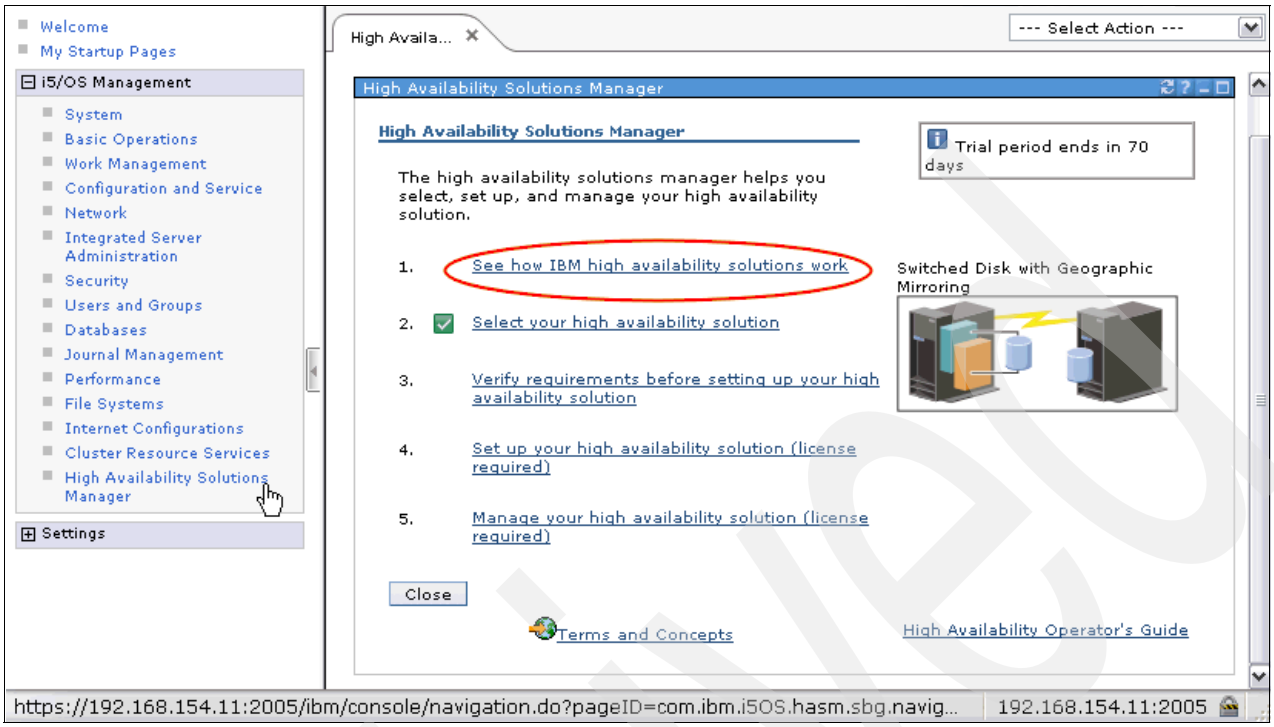


Figure 6-3 HASM GUI

A Flash demo provides an overview of the high availability supported solutions that you can set up using the High Availability Solutions Manager graphical interface. They are:

- Switched disk between logical partitions
- Switched disk between systems
- Switched disk with geographic mirroring
- Cross-site mirroring with geographic mirroring

6.3 Choosing your high availability solution

Once you have seen the Flash demos you are ready to select your high availability solution.

1. On this panel click **Select your High Availability Solutions** (Figure 6-4).

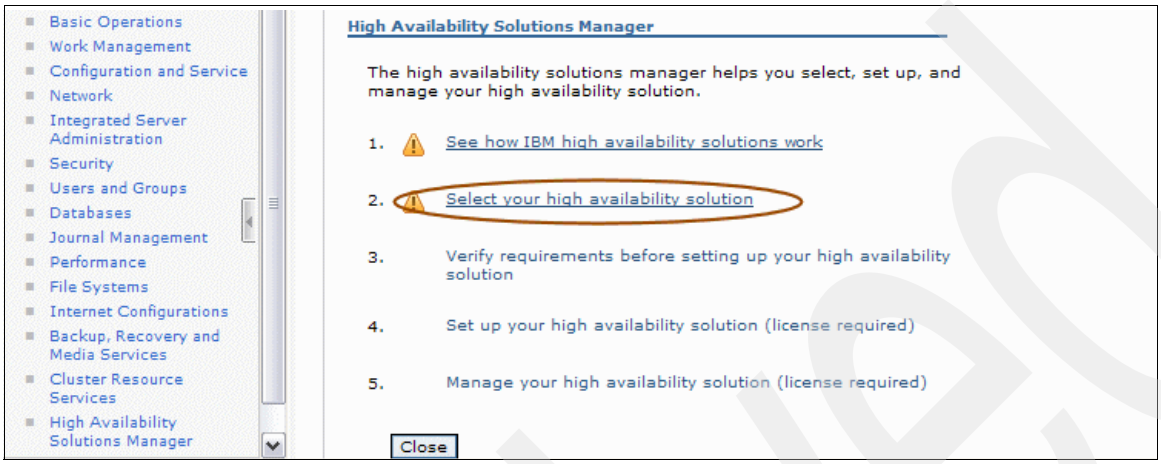


Figure 6-4 Select your high availability solution

2. Choose your solution and click **Select** (Figure 6-5). The four solutions to choose from are:
   - Switched disk between logical partitions

     This high availability solution uses disk pools that are switched between two logical partitions configured on the same system.

     The High Availability Solutions Manager GUI configures the clusters, the administrative domain, and the cluster resource group during the set up of your availability solutions step.

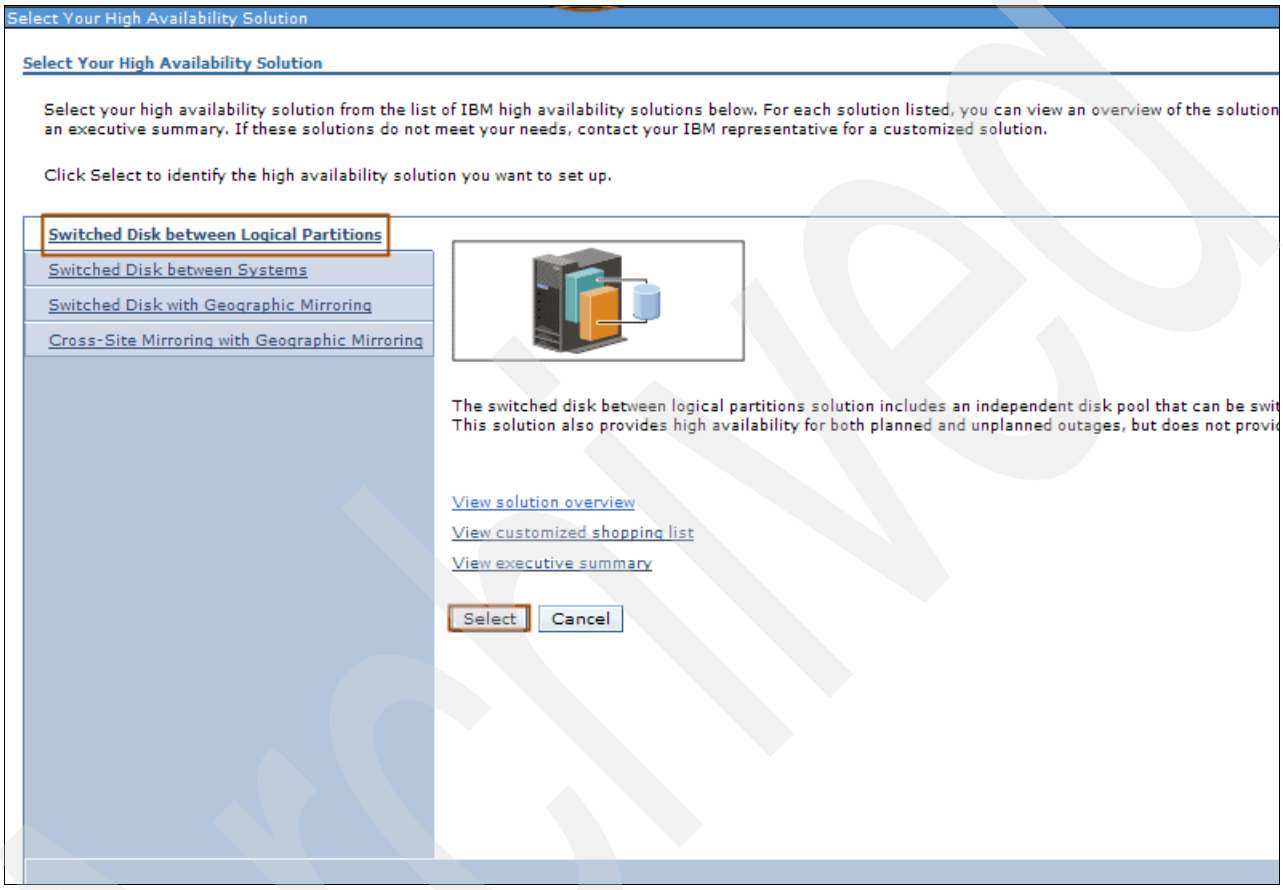


Figure 6-5 Select Your High Availability Solution: Switched Disk between Logical Partitions

– Switched disk between systems

This high availability solution uses switched disks between two systems (Figure 6-6).

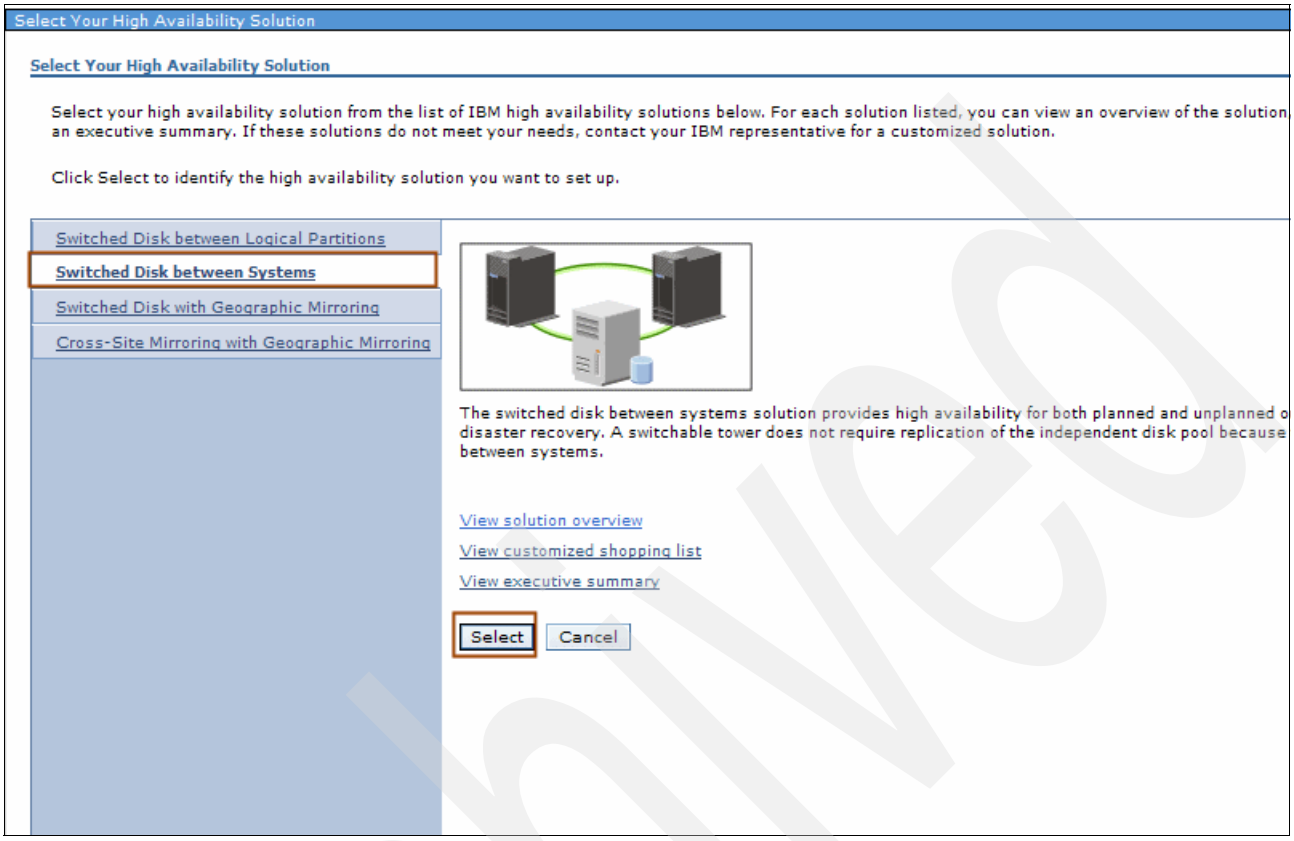


Figure 6-6 Switched Disk between Systems

– Switched disk with geographic mirroring

This solution uses three systems in a cross-site mirroring environment to provide both disaster recovery and high availability (Figure 6-7).

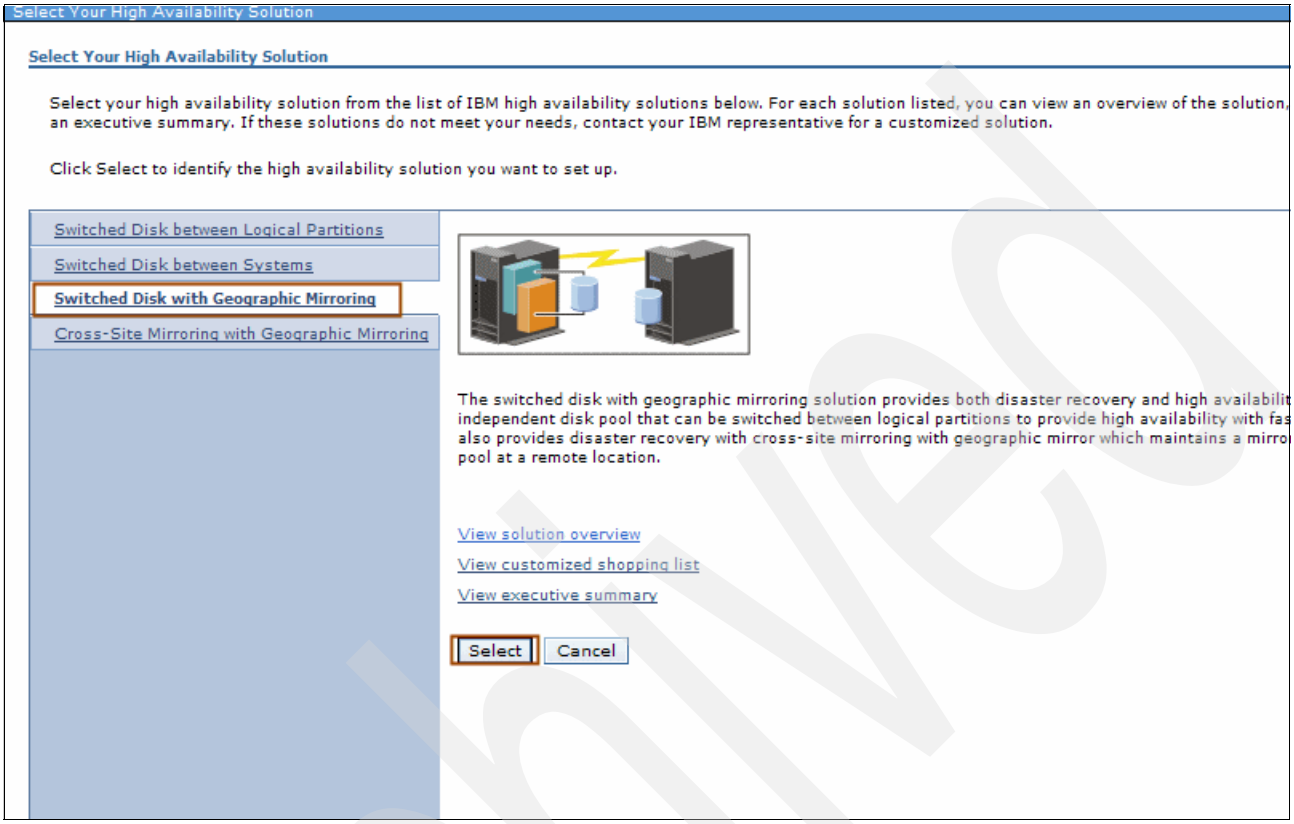


Figure 6-7 Switched Disk with Geographic Mirroring

– Cross-site mirroring with geographic mirroring

This solution provides high availability and disaster recovery by maintaining identical copies of the disk pool at two sites, local and remote, that are geographically separated (Figure 6-8).

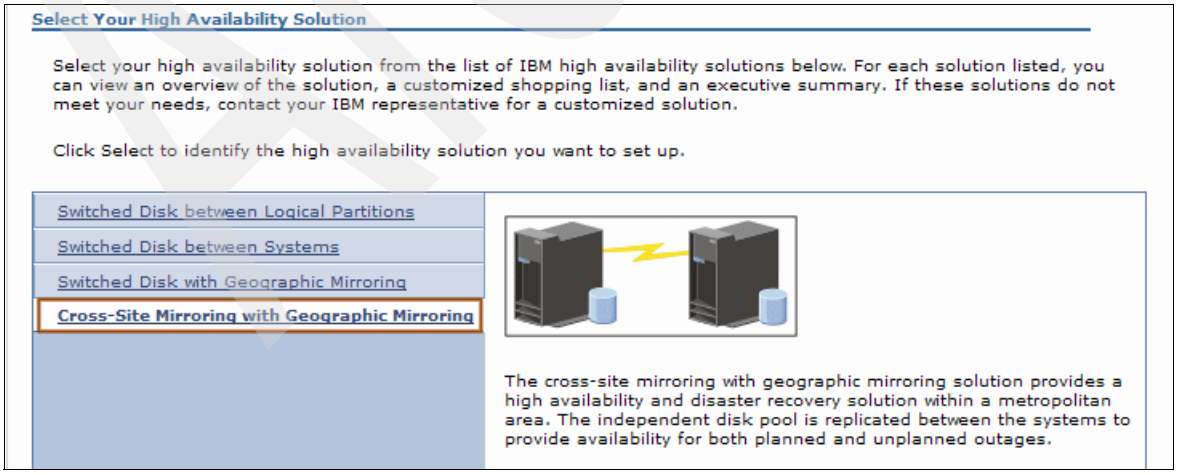


Figure 6-8 Cross-site Mirroring with Geographic Mirroring

6.4 Viewing a customized shopping list

For each solution you can view an overview of the solution, a customized shopping list, and an executive summary.

Start on the Select your High Availability Solution panel (see 6.2, “HASM GUI” on page 90) to see how to get there:

1. Click **View solution overview** to see a Flash demo of the selected high availability solution (Figure 6-9).

Figure 6-9 View solution overview

2. Click **View customized shopping list** (Figure 6-10).

Figure 6-10 View customized shopping list

On the next panel a customized shopping list example shows the list of information requested to implement your high availability solution (Figure 6-11).

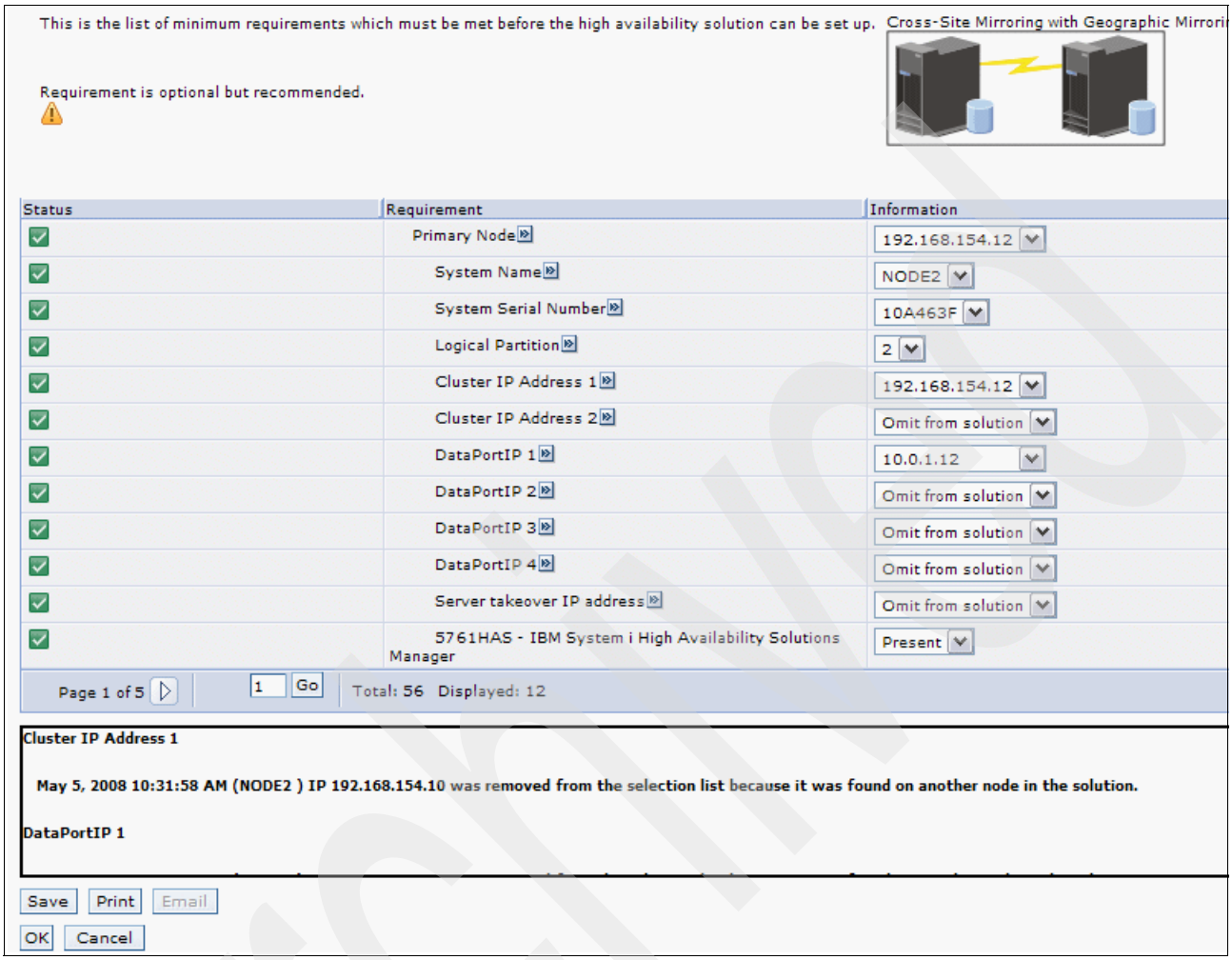


Figure 6-11 Customized shopping list

3. Click **OK** and then click **View executive summary** to view a summarized list of the advantages and restrictions of the selected solution, as shown in Figure 6-12.

Figure 6-12 View executive summary

4. An executive summary is shown and the user has the option to save or print it, as shown in Figure 6-13. Then click **Close** to return to the High Availability Solution Manager main panel.

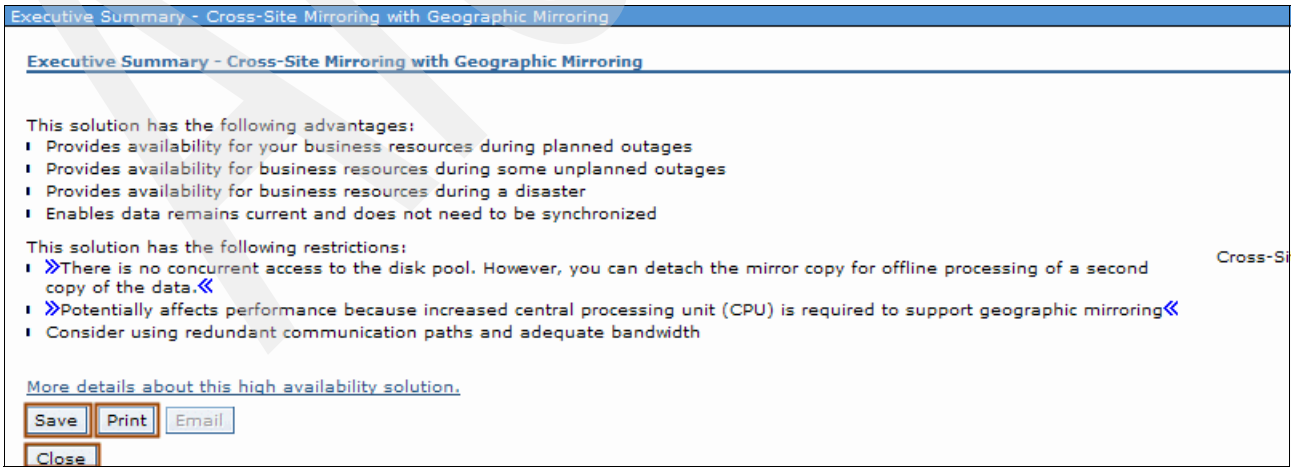


Figure 6-13 Executive Summary

6.5 Verifying requirements for your high availability solution

Before setting up your high availability solution you must ensure that all the hardware and software requirements are met. In this section we show you how to do this using the solution-based GUI.

1. Starting from the High Availability Solutions Manager panel click **Verify requirements before setting up your high availability solution** (Figure 6-14).

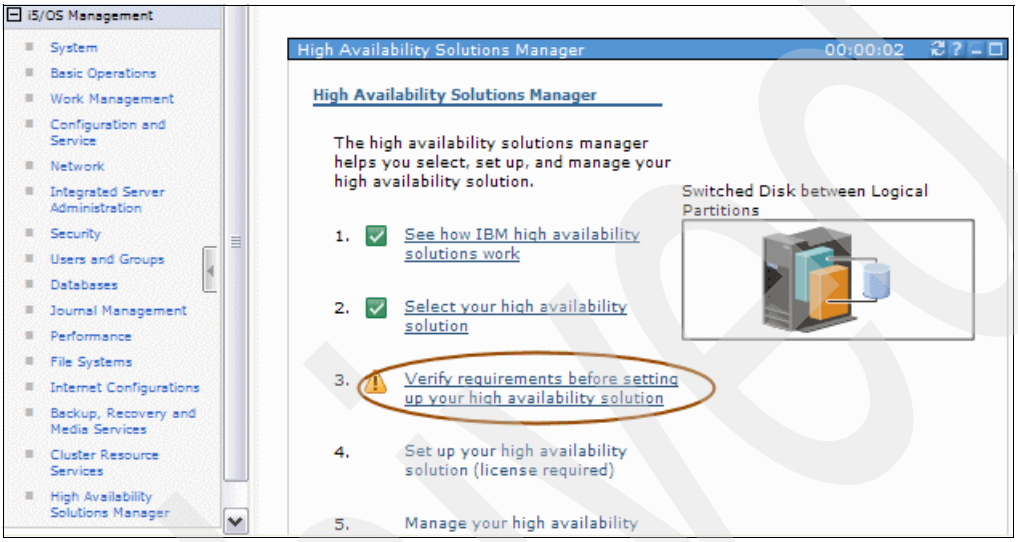


Figure 6-14 Verify requirements before setting up your high availability solution

The next window shows the running status of the verify requirements list task (Figure 6-15).

Figure 6-15 Verify requirements running status window

After few minutes the next window comes up displaying the minimum requirements that **must** be met before the high availability can be set up, as shown in Figure 6-16.

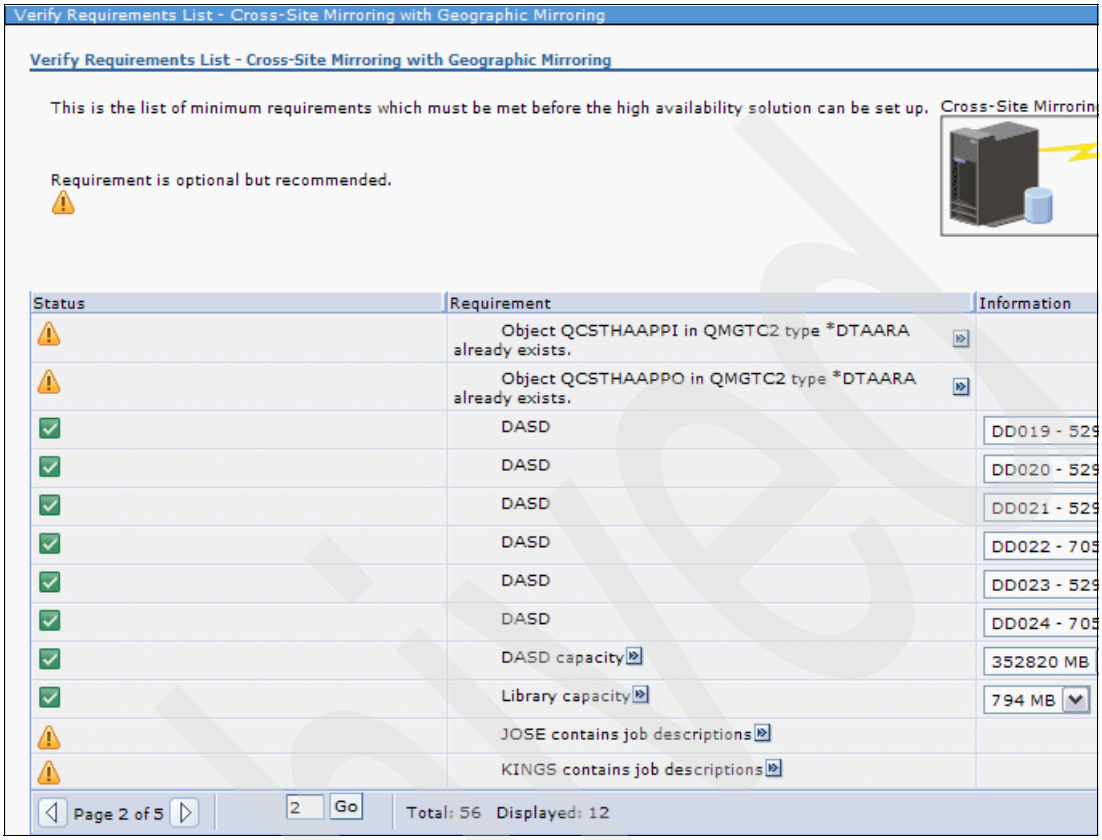


Figure 6-16 Verify requirements example

For each requirement the graphical interface provides a status value. See Figure 6-17 for the status explanation.

Status	Description
	The requirement must be met to set up your high availability solution. You can meet a requirement by supplying missing identification information for resources which will be used later to set up your high-availability solution or by installing missing hardware and software. After entering or selecting missing information click **OK**. After installing missing hardware or software, click **Refresh** so the requirements list can be updated to reflect the changes.
	The requirement is optional but might be recommended based on your specific business requirements. For example, a minimum of one disk drive is required to configure an independent disk pool but one disk might not be sufficient to store all of your data. You can meet a requirement by supplying missing identification information for resources which will be used later to set up your high-availability solution or by installing missing hardware and software. After entering or selecting missing information click **OK**. After installing missing hardware or software click **Refresh** so the requirements list can be updated to reflect the changes.
	The requirement has been met.

After all required hardware, software and information has been identified and successfully verified, you can set up your solution.

Figure 6-17 Verify requirements status

If a requirement is not met, as in the case shown in the next window (Figure 6-18), scroll down in the message area at the bottom of the panel to find the message stating the reason why the requirement is not met.

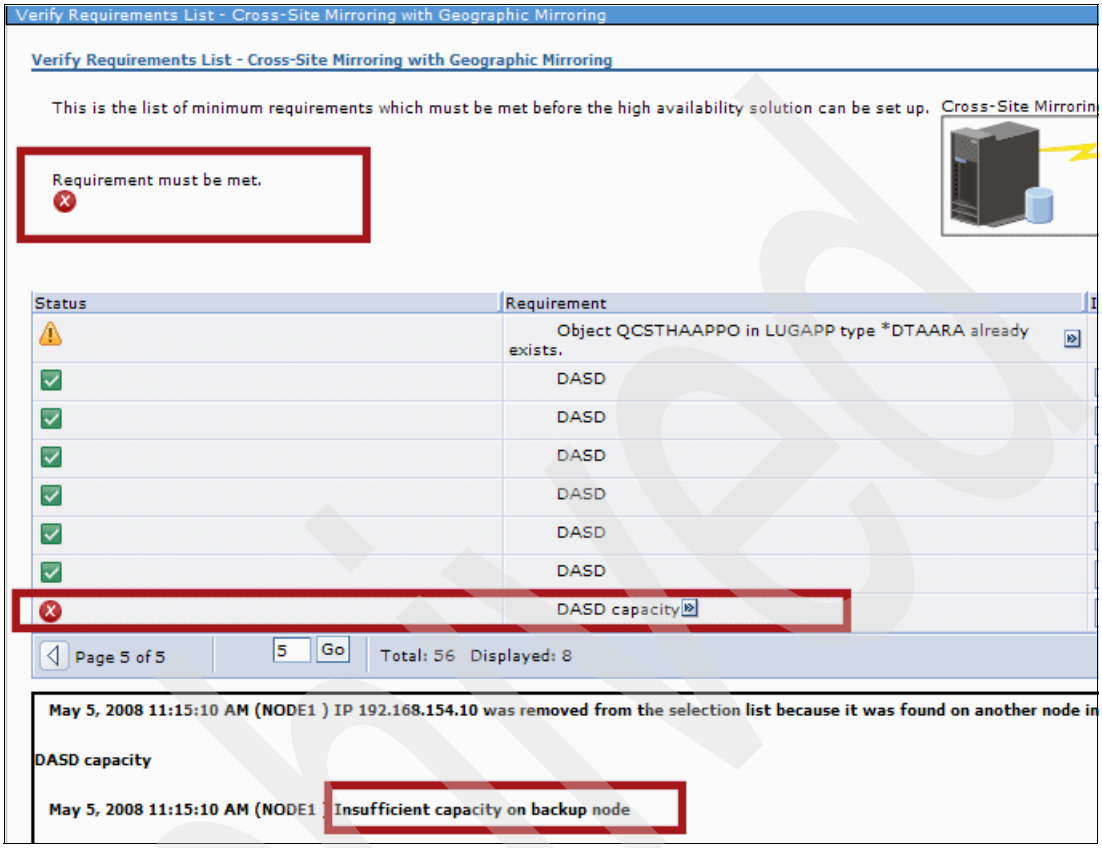


Figure 6-18 Verify requirements

2. Enter all the information requested and click **Refresh** so that the requirements list is updated to reflect the changes, as shown in Figure 6-19.

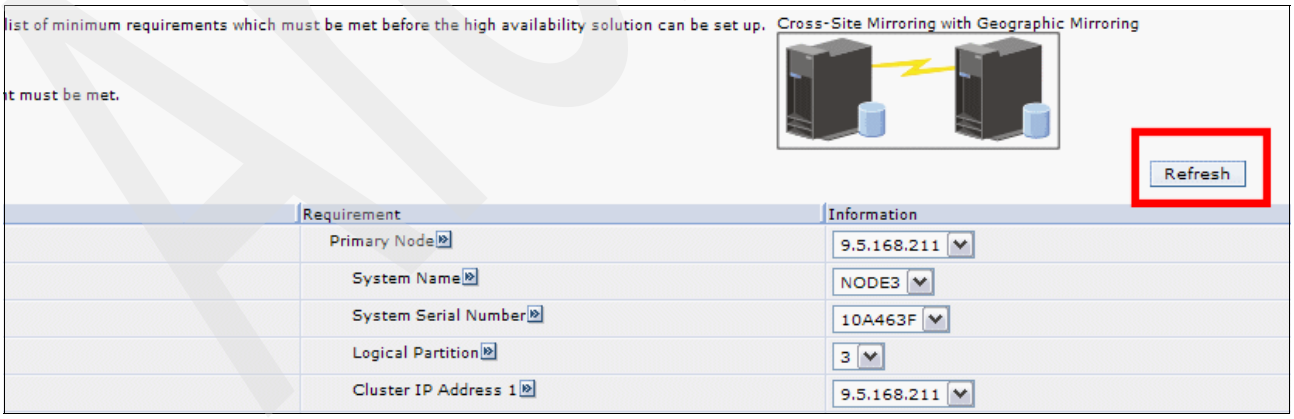


Figure 6-19 Verify requirements window

3. Click **OK** to finish, as shown in Figure 6-20.

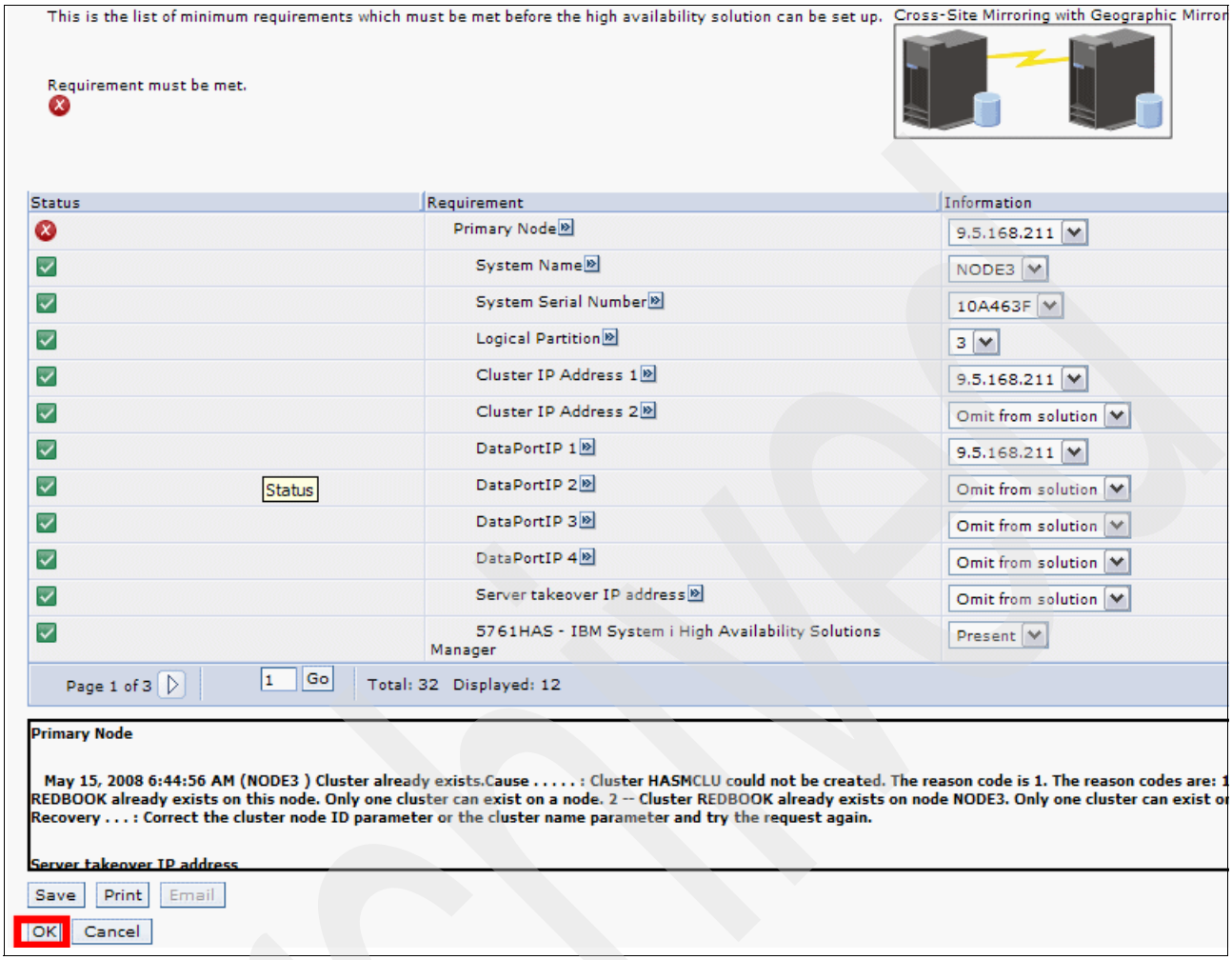


Figure 6-20 Verify requirements complete

6.6 Configuring cross-site mirroring with geographic mirroring

In this section we describe how to configure cross-site mirroring with geographic mirroring using the HASM GUI.

The requirements to implement the solution are:

- 5761-SS1 V6R1M0 HA Switchable Resources
- 5761-HAS V6R1M0 iHASM
- INETD server in active status on both nodes in the cluster
- Recommended last cumulative ptf level and latest fixes available for HA

6.6.1 Getting started with the setup of your high availability solution

The following is a summary of what is done in the Set Up High Availability Environment step of the deployment using the HASM GUI:

1. Create a cluster.
2. Add a node to the cluster.
3. Start the cluster nodes.
4. Add all the nodes to the device domain.
5. Create an independent ASP device description.
6. Grant authorities to QPGMR.
7. Create and start a cluster admin domain.
8. Add monitored resource entries.
9. Create a device CRG.
10. Configure the independent ASP.
11. Configure the mirror copy if there is to be one.
12. Start the device CRG.
13. Vary on the independent ASP.
14. Activate the user profile policies (when a profile is created or deleted).

Now we explain the steps to do this:

1. Starting from the High Availability Solution Manager panel, select **Select your high availability solution**, as shown in Figure 6-21.

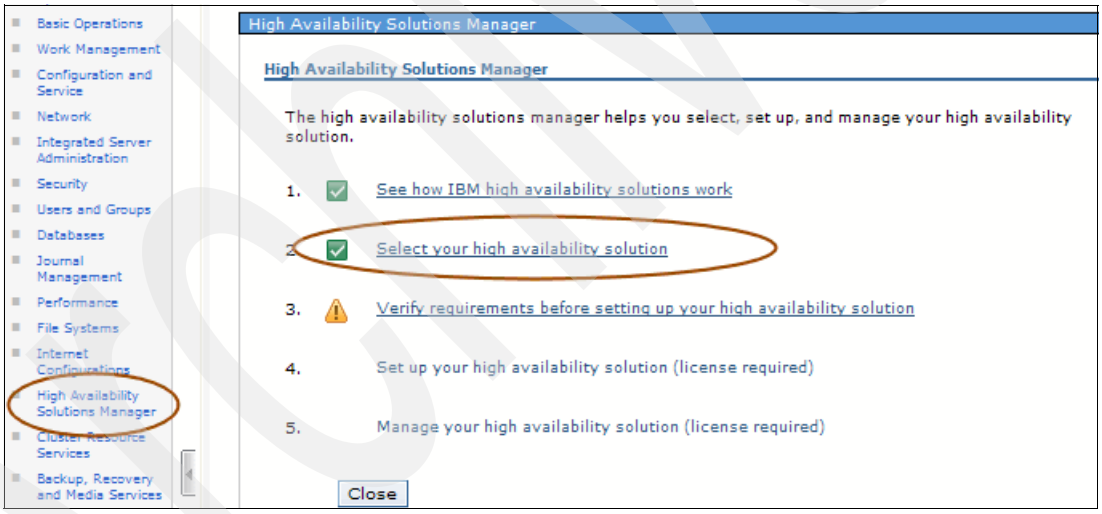


Figure 6-21 Select your high availability solution

2. Choose **Cross-Site® Mirroring with Geographic Mirroring** and then click **Select** to continue, as shown in Figure 6-22.

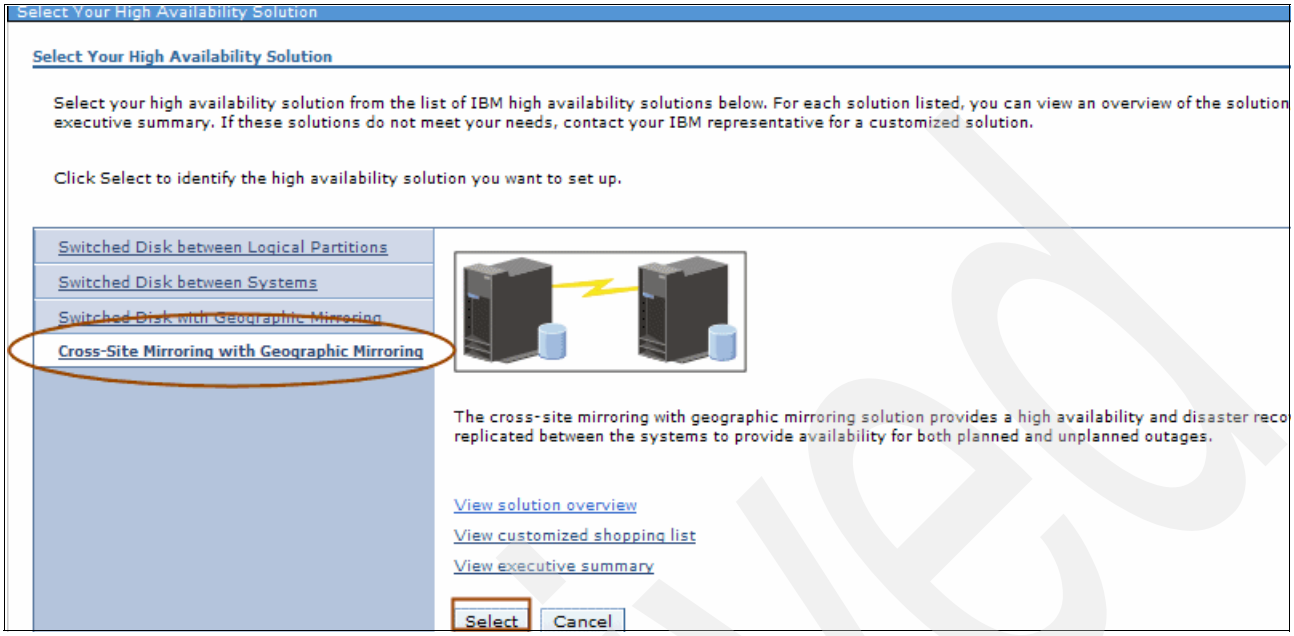


Figure 6-22 Select your High Availability Solution

3. Click **Verify requirements before setting up your high availability solution**, as shown in Figure 6-23

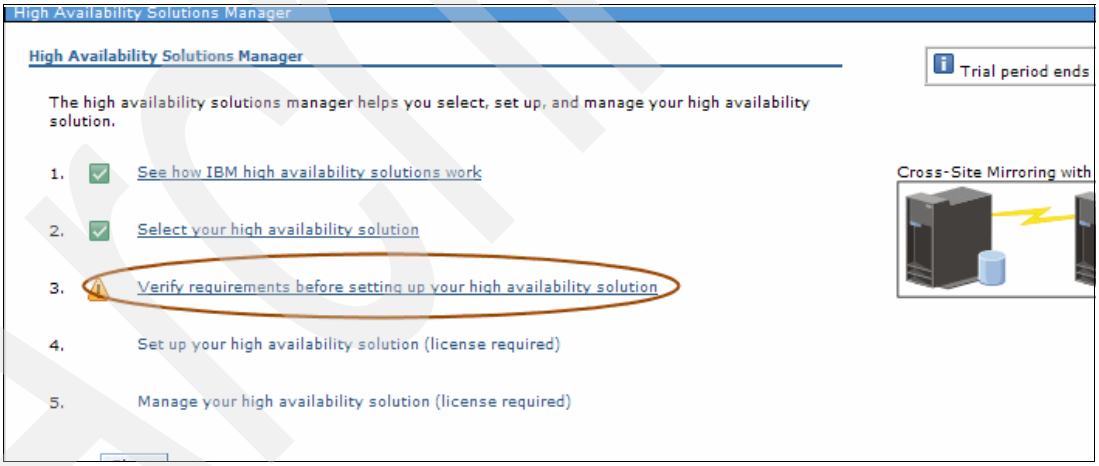


Figure 6-23 Setting up your high availability

Fill in all the requirements listed in the information column for both nodes. For the primary and backup, then click **Refresh** to refresh the list of requirements with the latest data entered, as shown Figure 6-24.

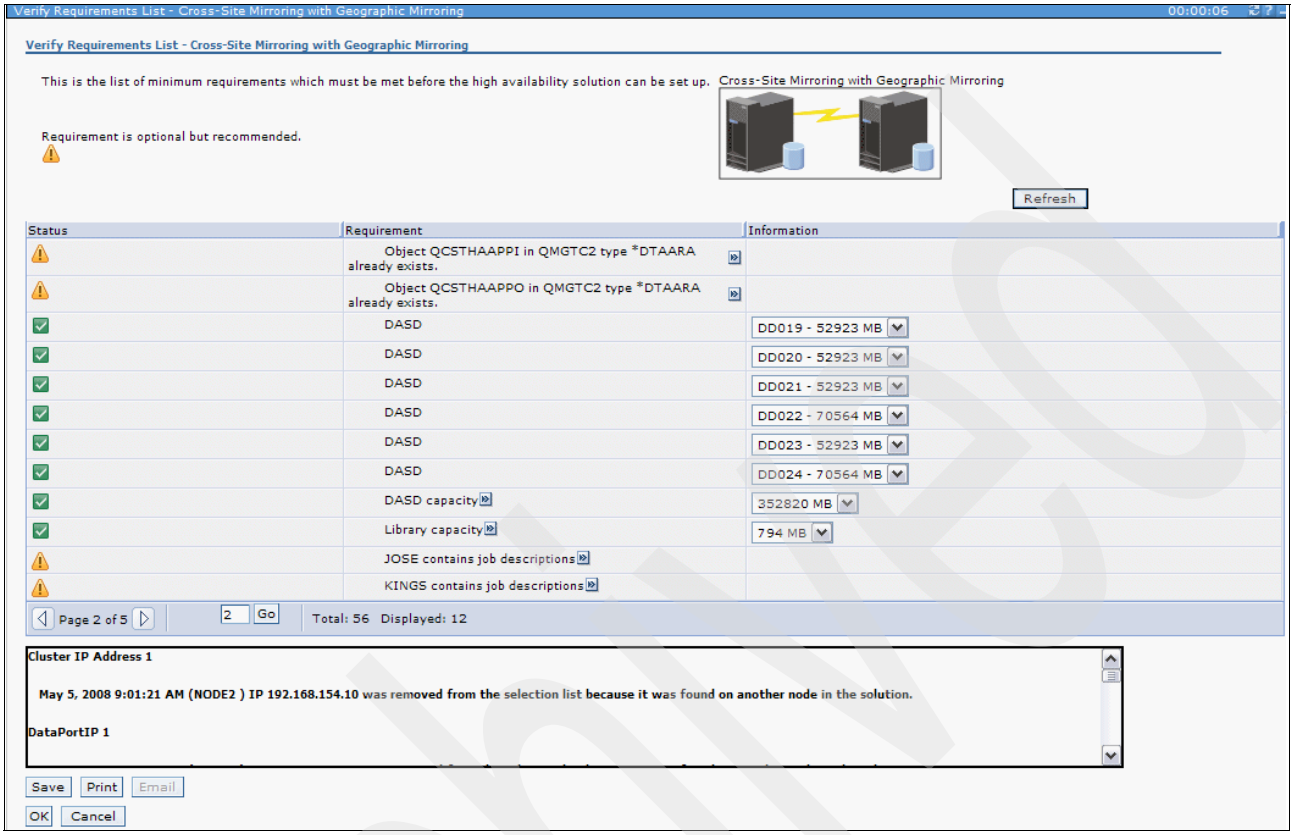


Figure 6-24 Verify Requirements List: Cross-Site Mirroring with Geographic Mirroring

4. Click **OK** to return to the previous panel.

> **Attention:** We strongly recommend that you test this on a non-production system that your applications can be moved to in an independent disk pool environment before setting up your high availability solution with the High Availability Solutions Manager graphical interface.

6.6.2 Setting up your high availability solution

Next we must set up the high availability solution. To do this:

1. Click **Set up your high availability solution**, as shown in Figure 6-25.

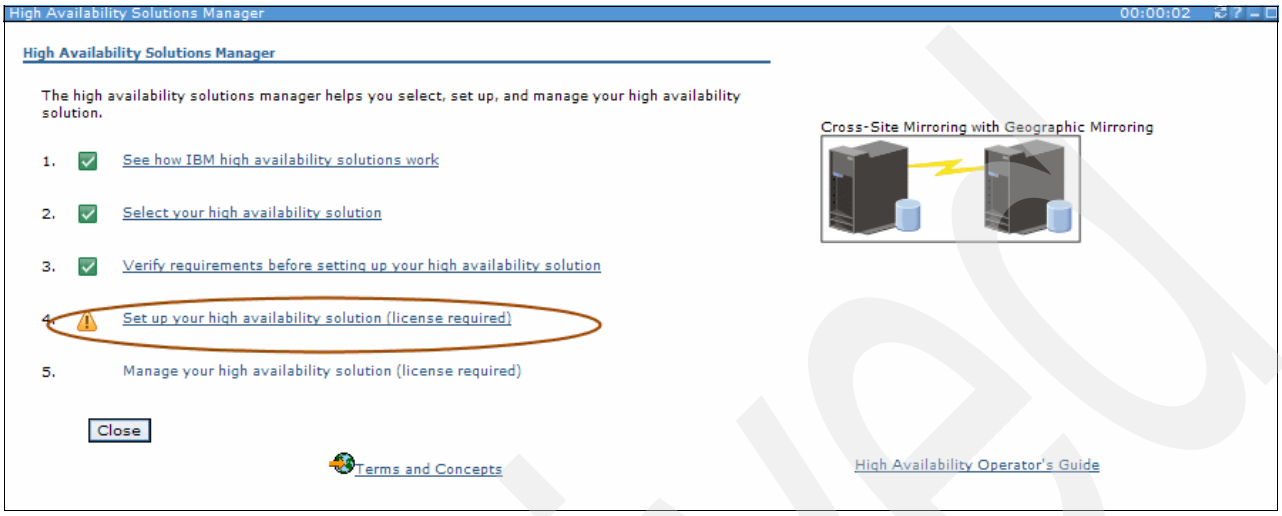


Figure 6-25 Setting up your high availability solution

Note: Setting up your high availability solution must be done when the systems in your high availability solution are in dedicated state.

2. On the next window click **Go** to run Set up high availability policies, as shown in Figure 6-26.

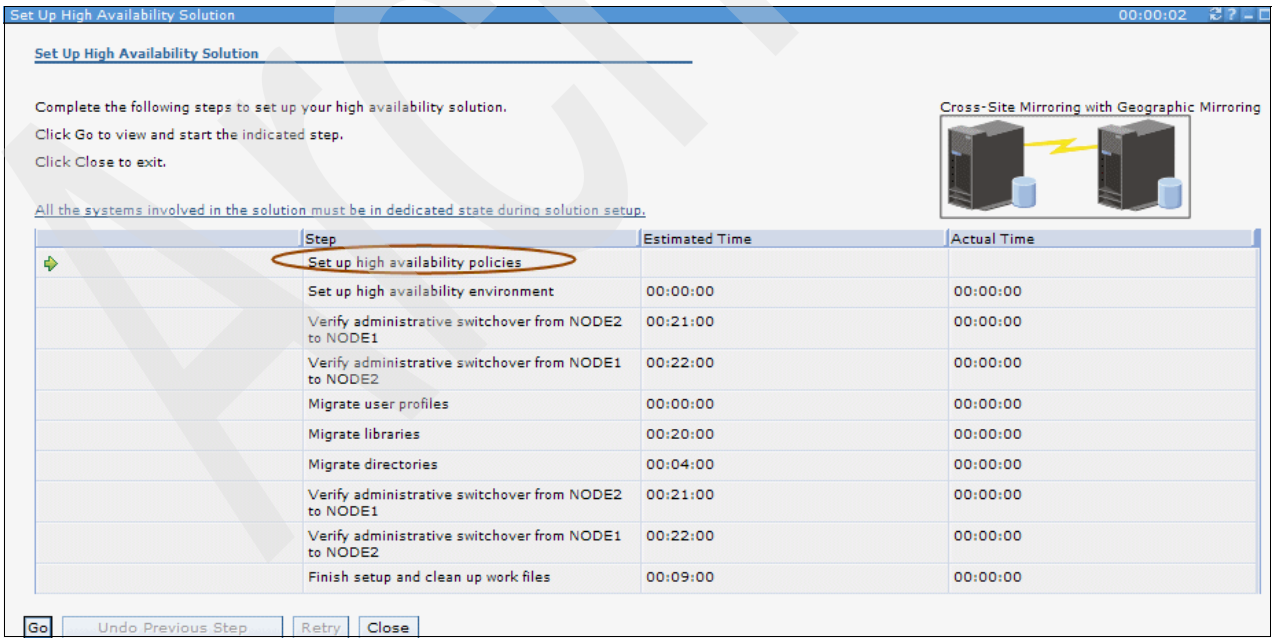


Figure 6-26 Set up your high availability solution policies

3. On the Dedicated state window click **Check State** to proceed, as shown in Figure 6-27.

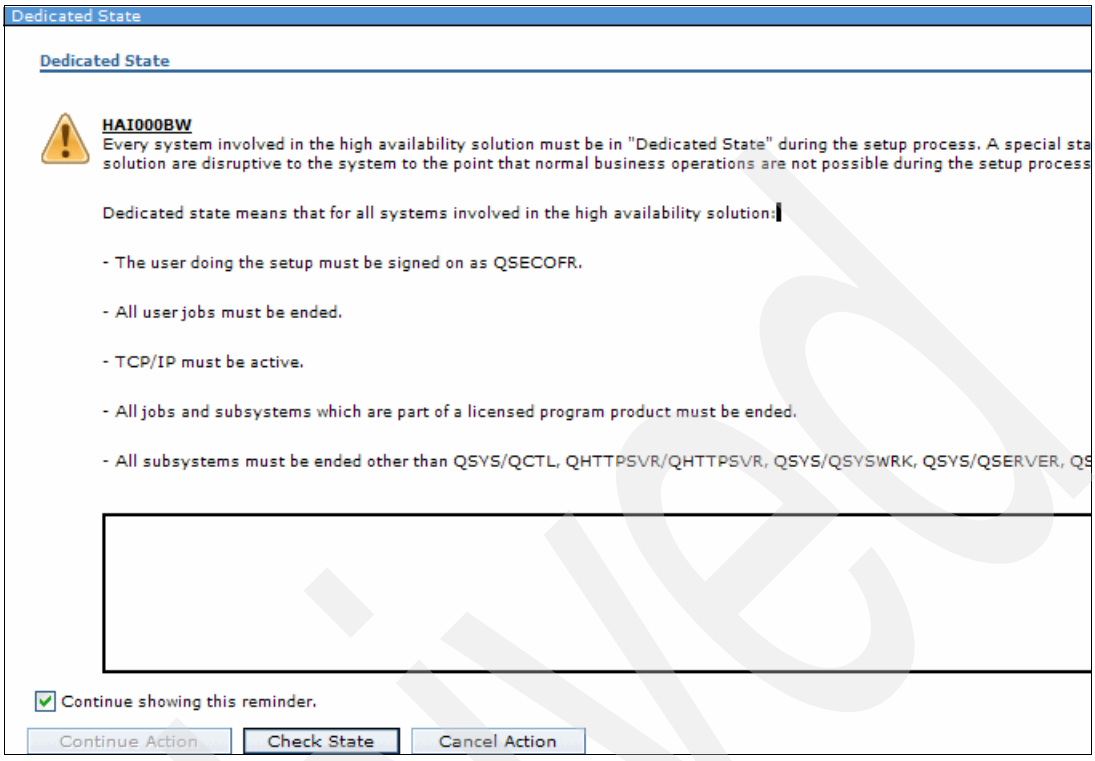


Figure 6-27 Dedicated State

After the check state operation is completed in the message area at the bottom of the panel you can see the jobs that should be ended on both nodes to put the systems in dedicated state, as shown in Figure 6-28.

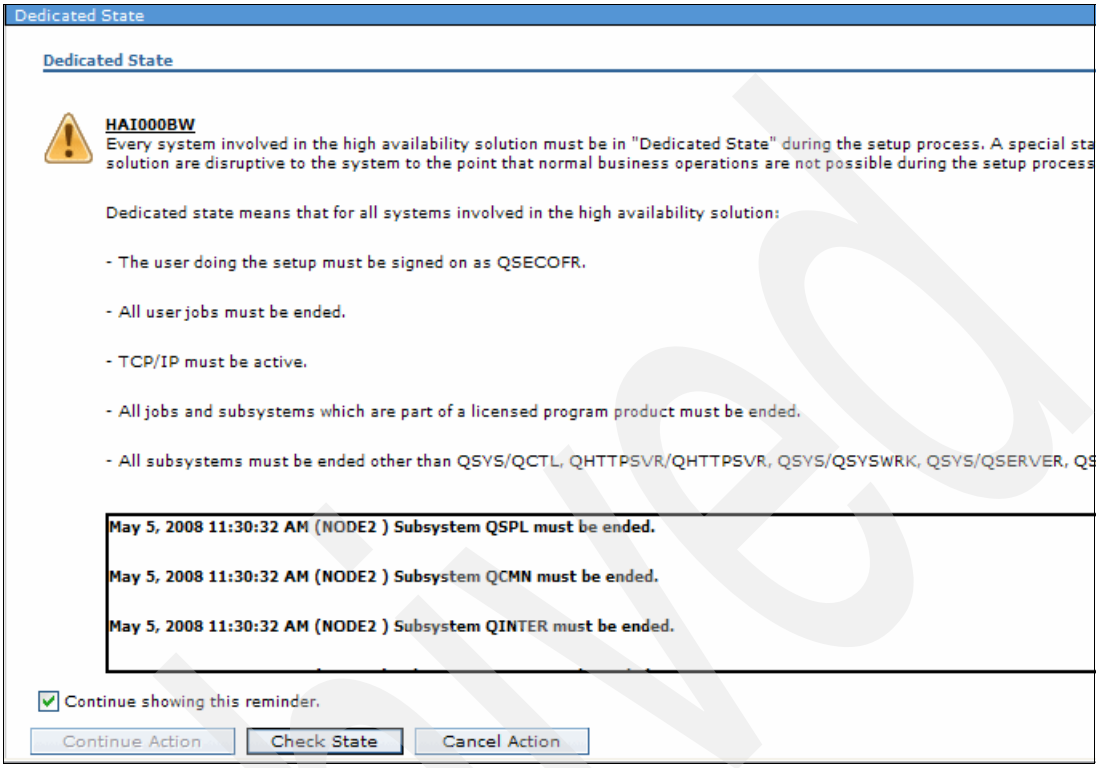


Figure 6-28 Dedicated State (check state)

4. Go to the command line and perform all the actions to end the jobs found active on your systems, then click **Continue Action** to proceed, as shown in Figure 6-29.

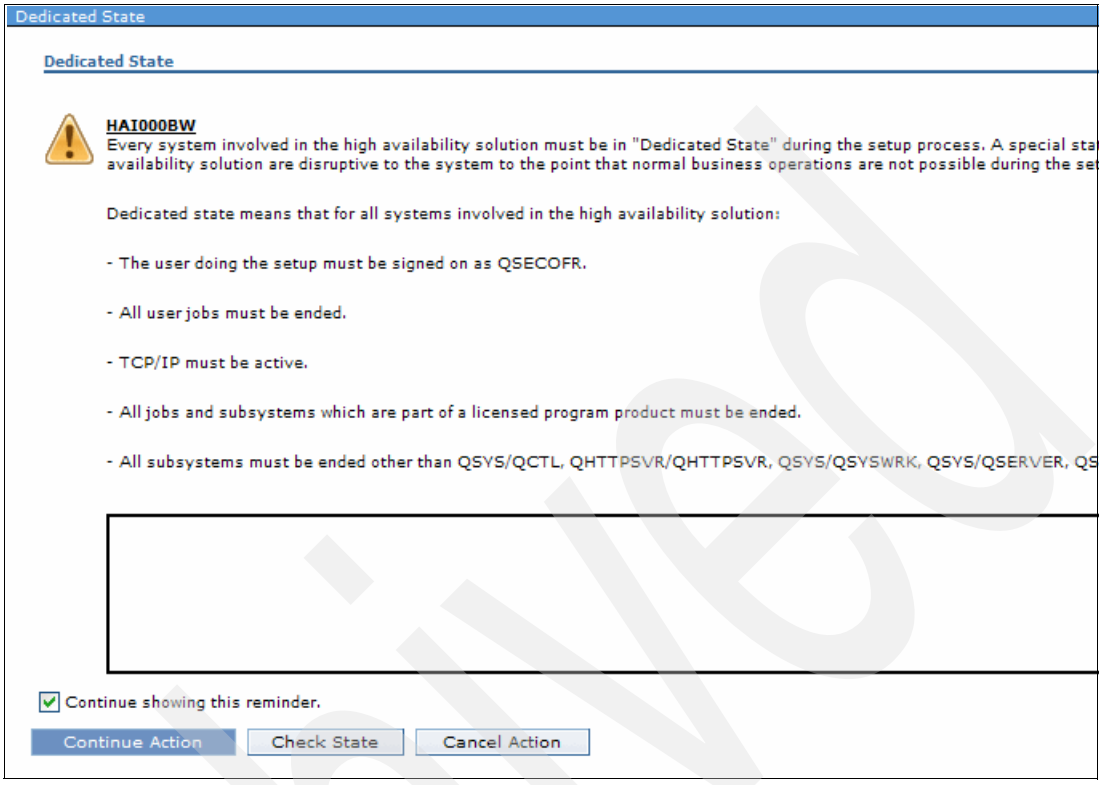


Figure 6-29 Dedicated State: Continue Action

5. Click **Close** to continue, as shown in Figure 6-30.

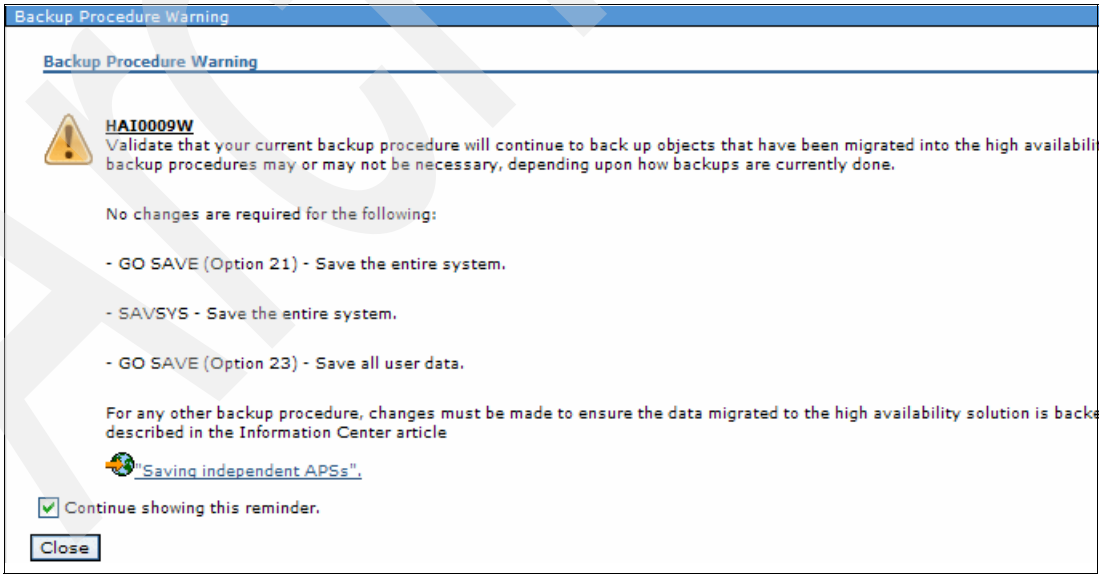


Figure 6-30 Backup Procedure Warning

6. Select the policies that you want to use in your high availability environment, then click **OK** to continue (Figure 6-31).

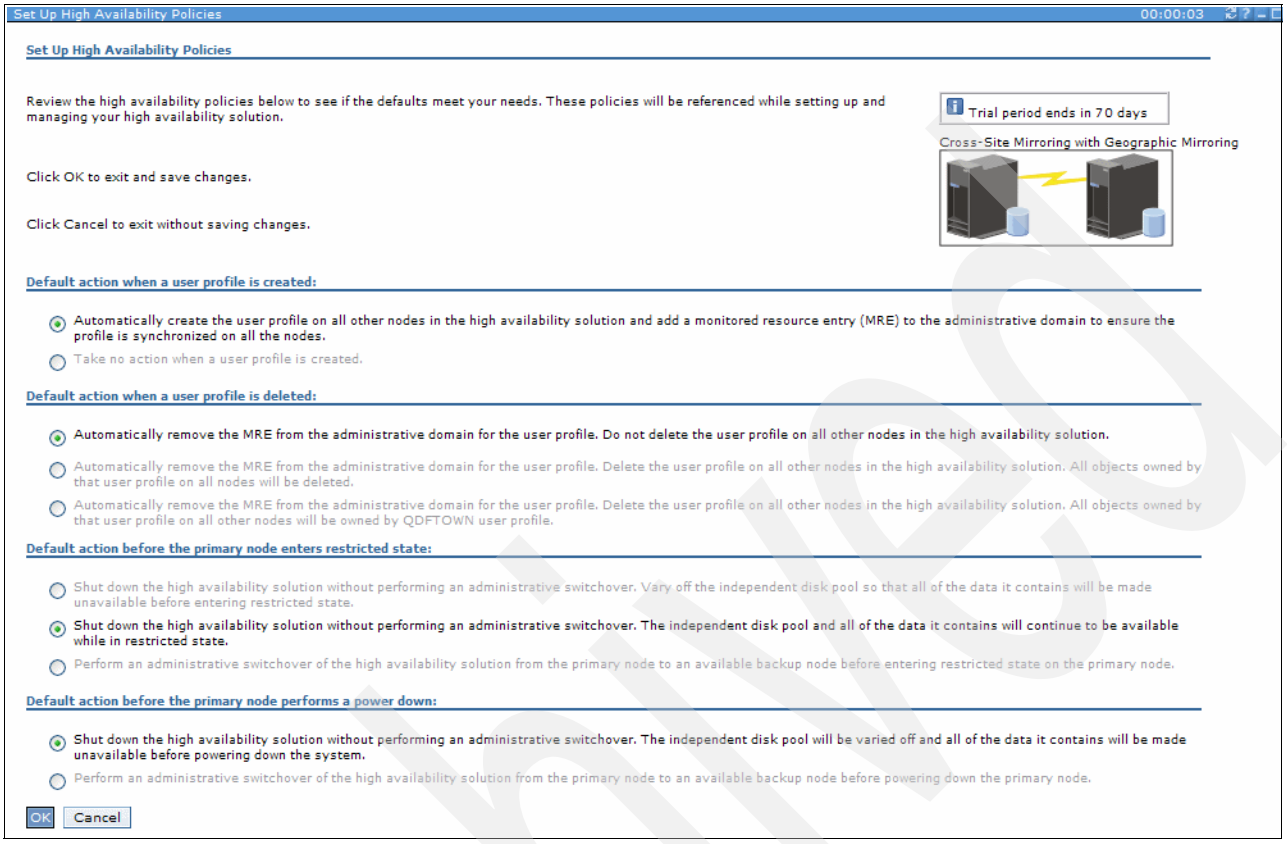


Figure 6-31 Set Up High Availability Policies

7. In the next window the green arrow indicates the task that the Set up high availability environment is going to perform. Click **Go** to run it (Figure 6-32).

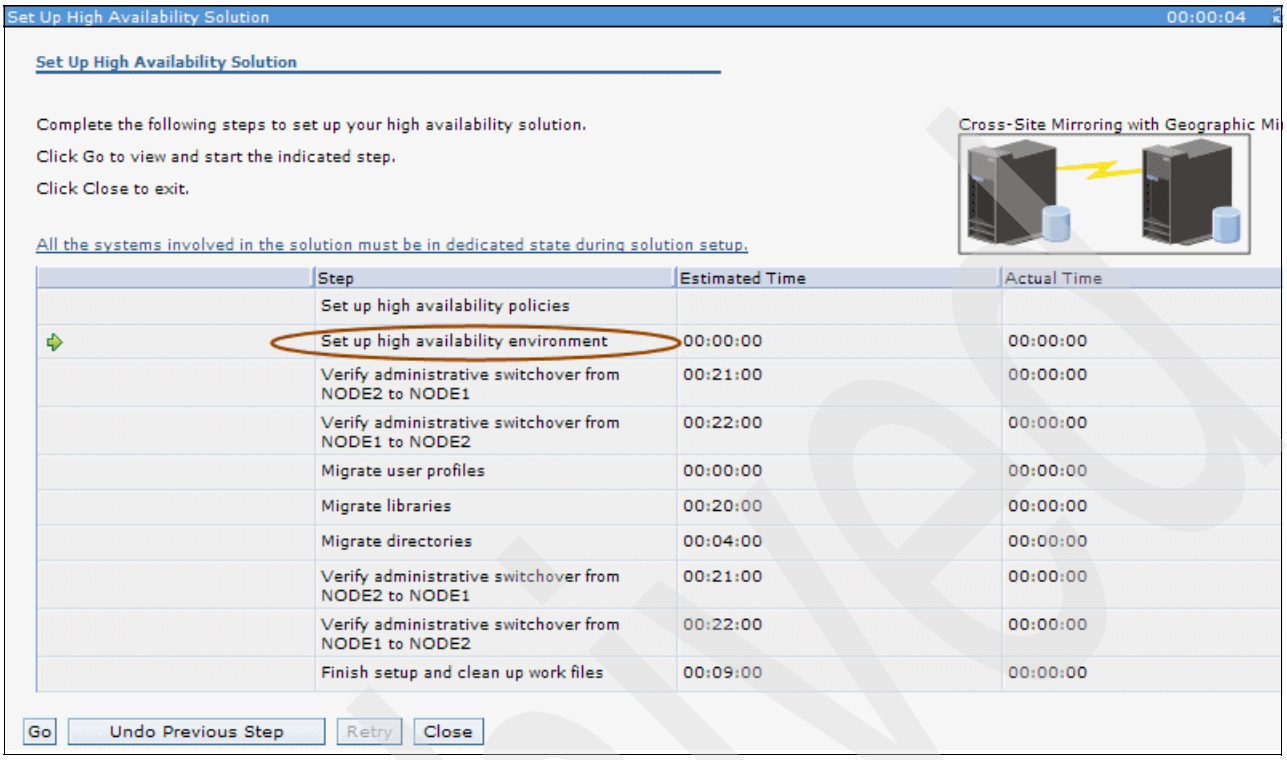


Figure 6-32 Set up high availability environment

8. Click **Run now** to perform the substeps that the green arrow is pointing to, as shown in Figure 6-33.

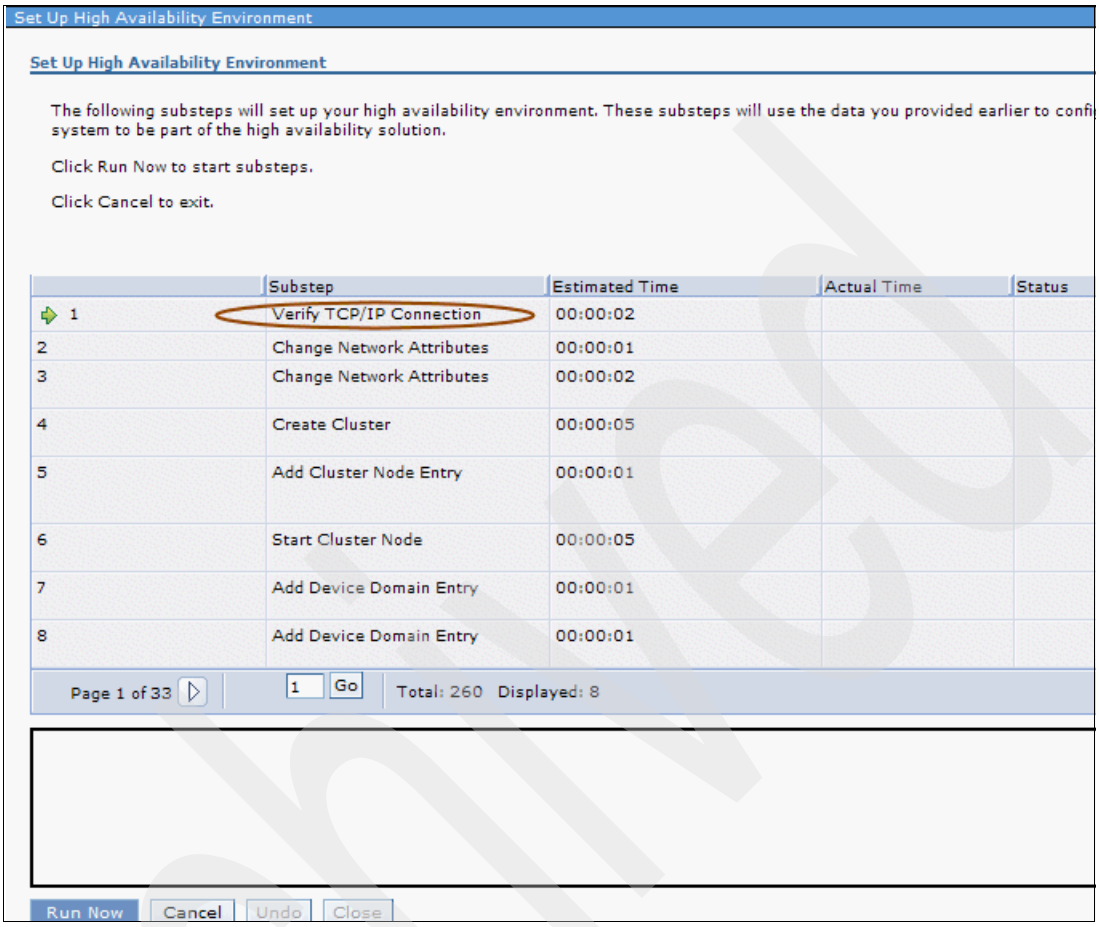
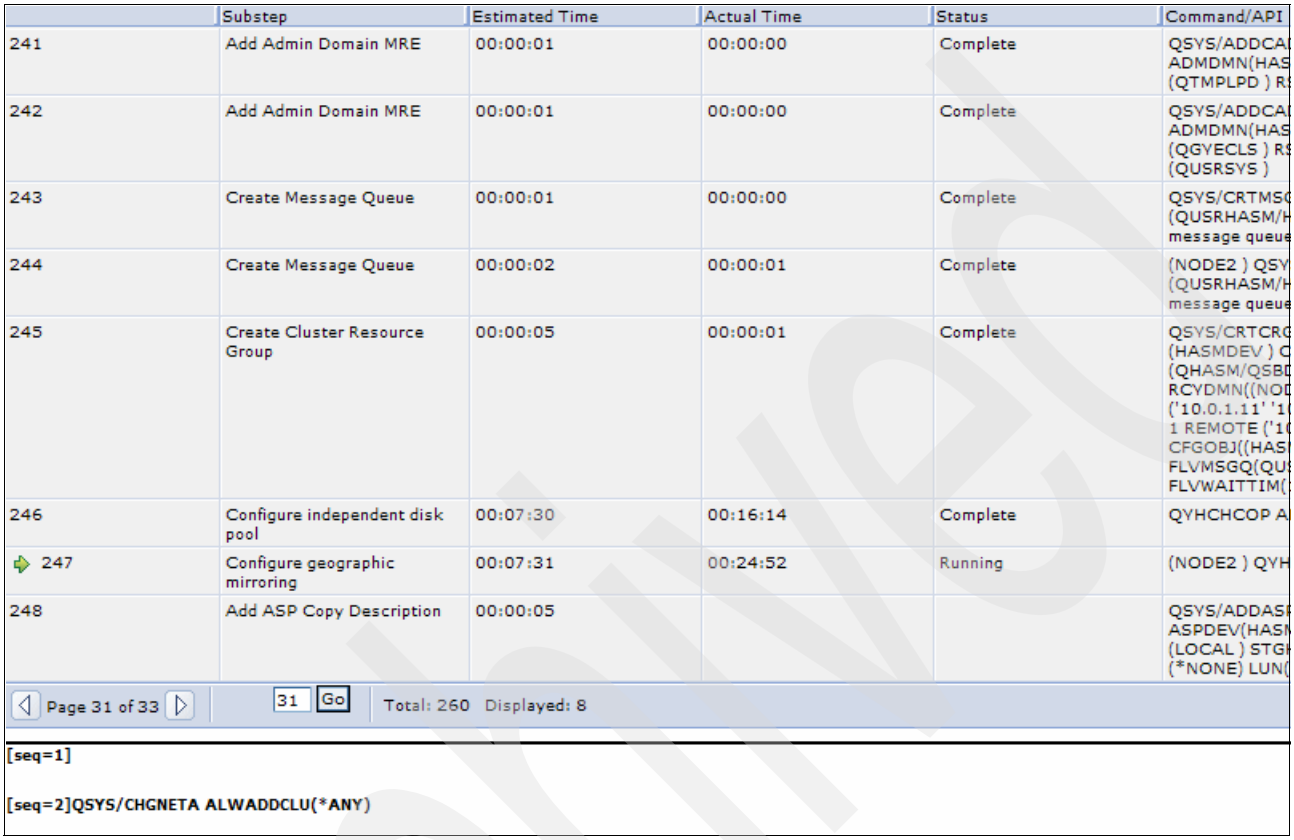


Figure 6-33 Substeps to set up the high availability environment

The High Availability Solution Manager performs all the substeps listed in the Set up high availability solution window. In the estimated time column you can see how much time each substep will take to complete. The status column is updated when the substeps complete, as shown in Figure 6-34.



	Substep	Estimated Time	Actual Time	Status	Command/API
241	Add Admin Domain MRE	00:00:01	00:00:00	Complete	QSYS/ADDCA ADMDMN(HAS (QTMPLPD) R
242	Add Admin Domain MRE	00:00:01	00:00:00	Complete	QSYS/ADDCA ADMDMN(HAS (QGYECLS) R (QUSRSYS)
243	Create Message Queue	00:00:01	00:00:00	Complete	QSYS/CRTMS (QUSRHASM/H message queue
244	Create Message Queue	00:00:02	00:00:01	Complete	(NODE2) QSY (QUSRHASM/H message queue
245	Create Cluster Resource Group	00:00:05	00:00:01	Complete	QSYS/CRTCRG (HASMDEV) C (QHASM/QSB RCYDMN((NOD ('10.0.1.11' '1 1 REMOTE ('1 CFGOBJ((HAS FLVMSGQ(QU FLVWAITTIM(
246	Configure independent disk pool	00:07:30	00:16:14	Complete	QYHCHCOP A
247	Configure geographic mirroring	00:07:31	00:24:52	Running	(NODE2) QYH
248	Add ASP Copy Description	00:00:05			QSYS/ADDAS ASPDEV(HAS (LOCAL) STG (*NONE) LUN(

Page 31 of 33 | 31 Go | Total: 260 Displayed: 8

```
[seq=1]

[seq=2]QSYS/CHGNETA ALWADDCLU(*ANY)
```

Figure 6-34 Set up in progress

9. Click **Close** to return to the Set Up High Availability window when all the substeps are completed (Figure 6-35).

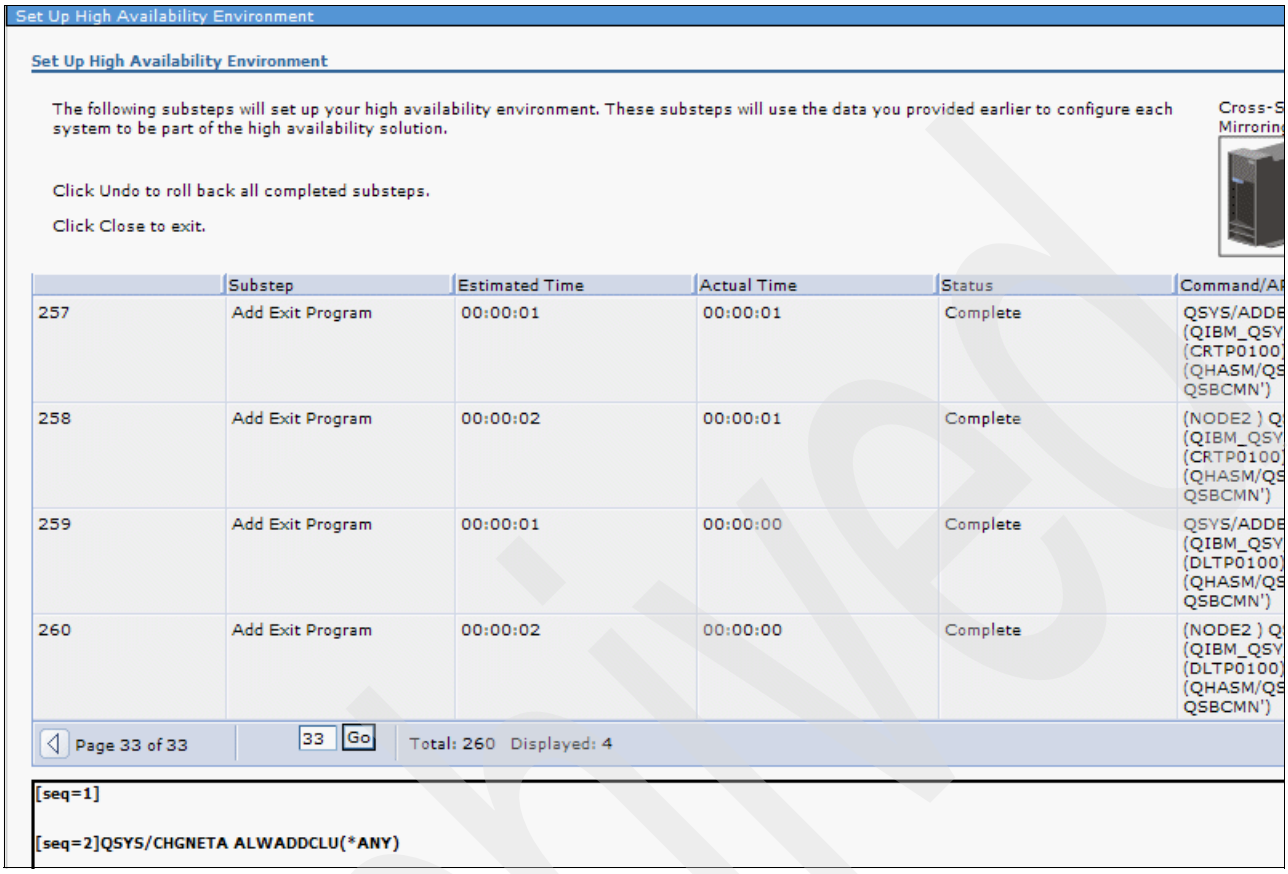


Figure 6-35 Set up high availability completed

10. ,Click **Close** to return to the Set up high availability solution window, as shown in Figure 6-36.

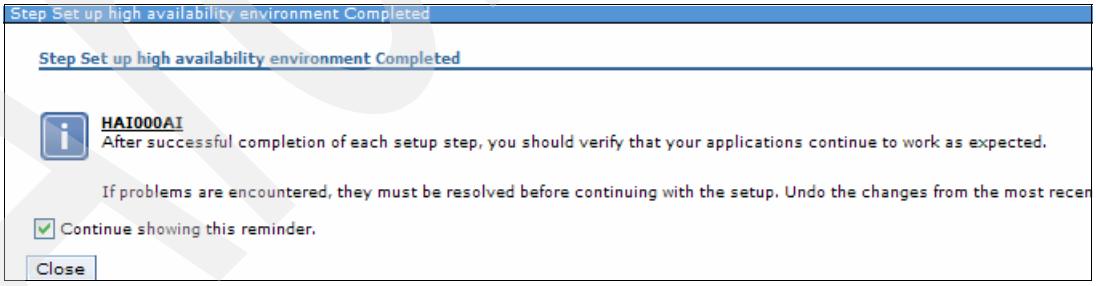


Figure 6-36 Step Set up high availability environment completed

From the command line run DSPCLUINF to verify that the cluster HASMCLU has been successfully created, as shown in Figure 6-37.

```
                          Display Cluster Information
        Cluster  . . . . . . . . . . . . . . . . . :   HASMCLU
        Consistent information in cluster  . . . . :   Yes
        Current cluster version  . . . . . . . . . :   6
        Current cluster modification level . . . . :   0
        Configuration tuning level . . . . . . . . :   *NORMAL
        Number of cluster nodes  . . . . . . . . . :   2
        Number of device domains . . . . . . . . . :   1
        Number of administrative domains . . . . . :   1
        Number of cluster resource groups  . . . . :   1
        Cluster message queue  . . . . . . . . . . :   *NONE
        Library  . . . . . . . . . . . . . . . . . :   *NONE
        Failover wait time . . . . . . . . . . . . :   *NOWAIT
        Failover default action  . . . . . . . . . :   *PROCEED
```

Figure 6-37 DSPCLUINF command

Then run the WRKCFGSTS *DEV HASMIASP (iASP name created by the solution-based GUI) on both the primary node and the backup node, as shown in Figure 6-38 and Figure 6-39.

```
                  Work with Configuration Status             NODE1
                                                          05/14/08  11:12:00
Position to  . . . . .              Starting characters

Type options, press Enter.
  1=Vary on   2=Vary off   5=Work with job   8=Work with description
  9=Display mode status    13=Work with APPN status...

Opt  Description      Status               -------------Job--------------
     HASMIASP         AVAILABLE
```

Figure 6-38 IASP status on primary node

```
                  Work with Configuration Status             NODE2
                                                          05/14/08  13:12:38
Position to  . . . . .              Starting characters

Type options, press Enter.
  1=Vary on   2=Vary off   5=Work with job   8=Work with description
  9=Display mode status    13=Work with APPN status...

Opt  Description      Status               -------------Job--------------
     HASMIASP         VARIED ON
```

Figure 6-39 iASP status on backup node

To verify the geographic mirroring configuration run the command DSPASPSSN specifying the session name parameter as HASMASPSSN, as shown in Figure 6-40.

```
                                   DSPASPSSN
  Session . . . . . :   HASMASPSSN
  Type  . . . . . :     *GEOMIR
  Mode  . . . . . :     SYNC
  Suspend timeout . . . . . :     120
  Synchronization priority  . . . . . :     *MEDIUM
  Track space . . . . . :     0

Copy Descriptions

ASP           ASP                                          Data
Device        Copy              Role         State        State    Node
HASMIASP      HASMLCLCPY     PRODUCTION     AVAILABLE      USABLE    NODE1
HASMIASP      HASMRMTCPY        MIRROR       ACTIVE           USABLE    NODE2
```

Figure 6-40 DSPASPSSN command.

After each step is complete you can review the display log. Click the context menu (red circle) to open it, as shown in Figure 6-41.

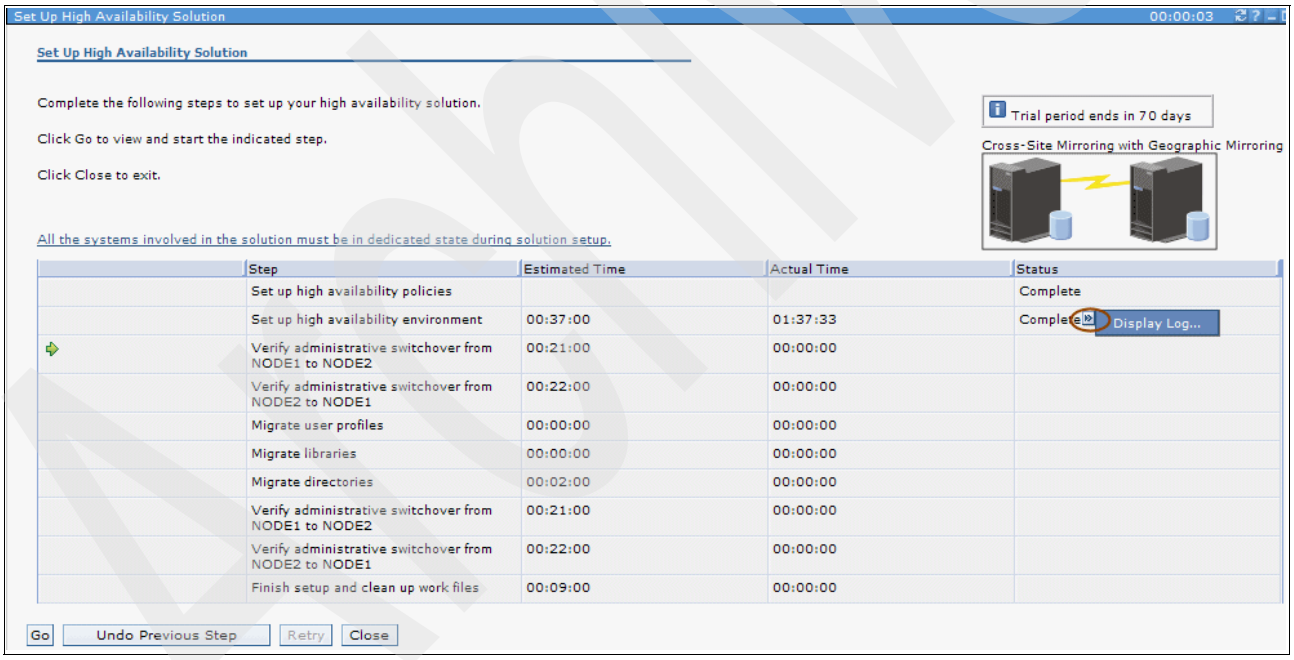


Figure 6-41 Display log

Before running the next steps (verify administrative switchover) you must take the systems involved in the high availability solution in a dedicated state, which means:

a. Log on to the systems with your QSECOFR user profile and password.

b. End all user jobs.

c. Ensure that TCP/IP is active.

d. End all jobs and subsystems associated with all licensed programs (all LPPs).

e. Ensure that all subsystem jobs are ended except QCTL, QBATCH, QSYSWRK, QUSRWRK,QSERVER. and QHTTPSVR.

11. Click **Go** to run the verify administrative switchover from NODE1(primary) to NODE2(backup), as shown in Figure 6-42.

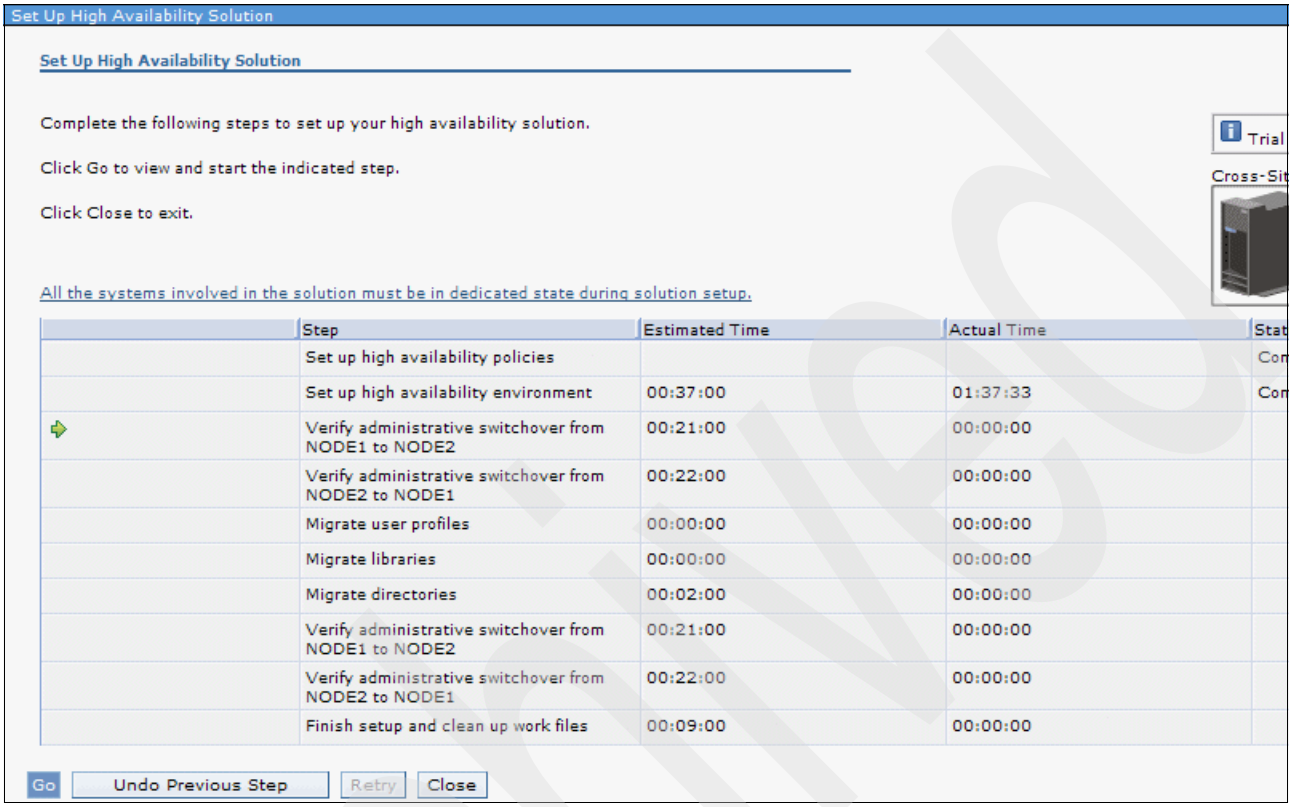


Figure 6-42 Verify administrative switchover

12. Click **Run now** to perform all the substeps displayed, as shown in Figure 6-43.

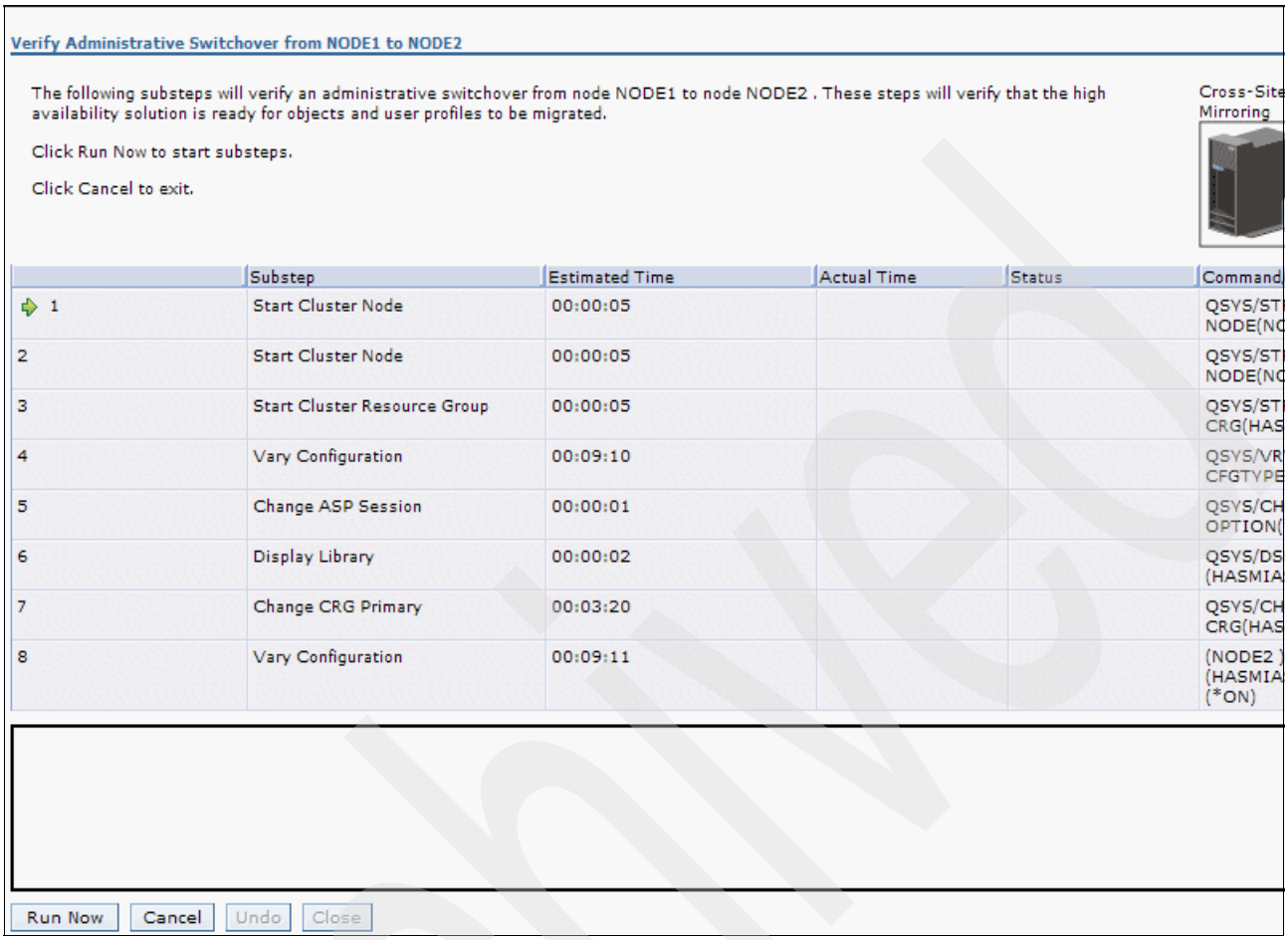


Figure 6-43 Verify administrative switchover execution

13. You are back again on the Verify administrative switchover window. Click **Run now** to perform this step, as shown in Figure 6-43.

14. The Verify administrative switchover step performs the Change CRG Primary task (CHGCRGPRI), substep 7, as shown in Figure 6-44.

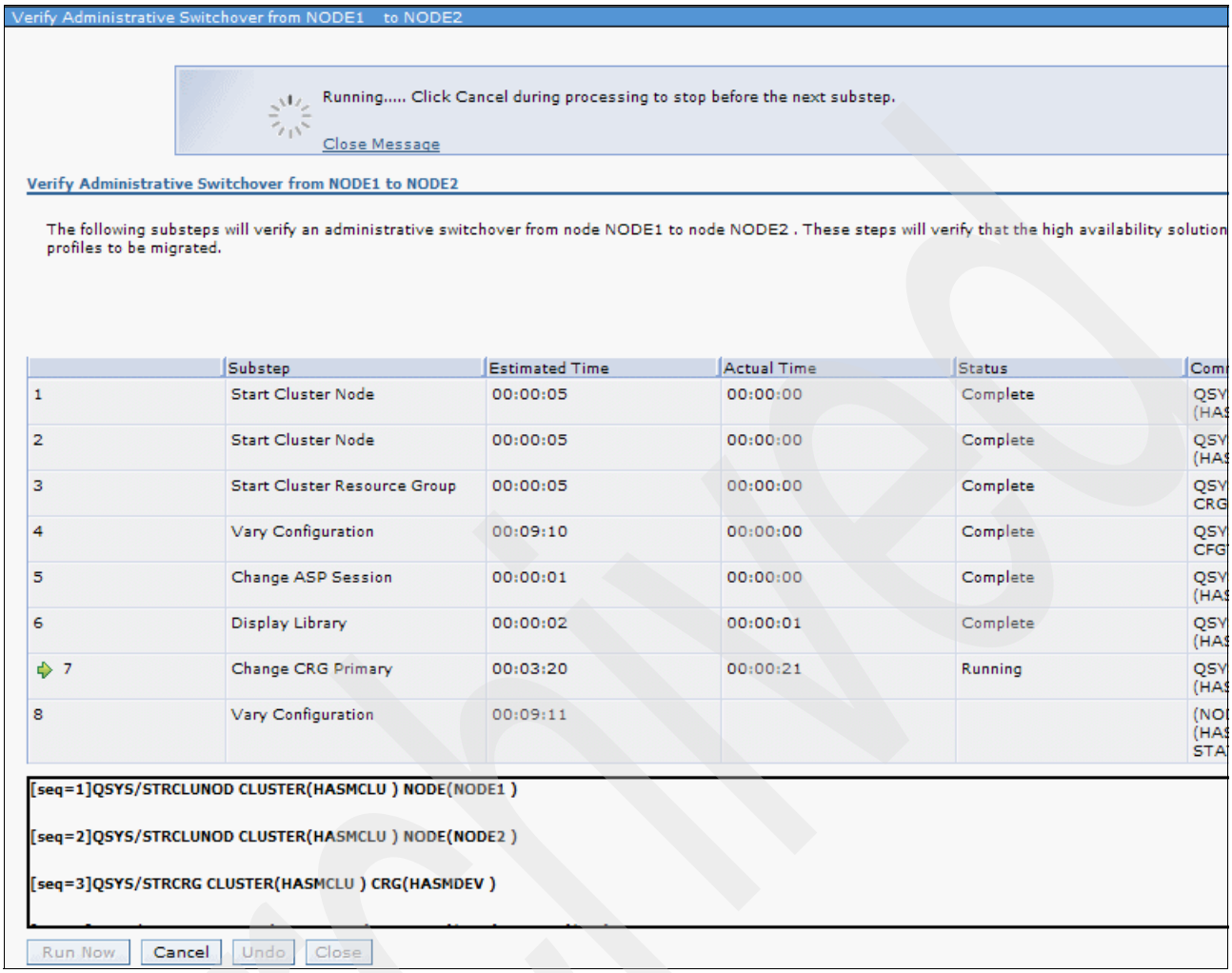


Figure 6-44 Change CRG Primary

15. When the switchover is completed the window shown in Figure 6-45 is displayed. Click **Close** to return to the Setup high availability solution window.

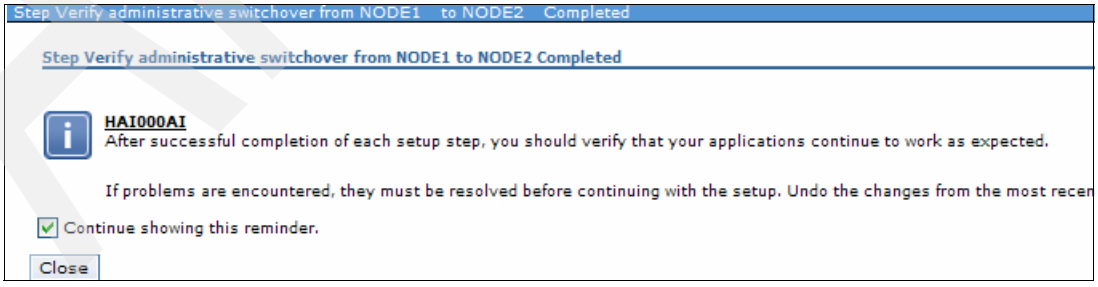


Figure 6-45 Switchover completed

16. From the command line run the DSPCRGINF command to check that the node's role has been changed. Now NODE1 is the backup system and NODE2 is the primary system, as shown Figure 6-46.

```
                        Display CRG Information

 Cluster  . . . . . . . . . . . . . :   HASMCLU
 Cluster resource group . . . . . :   HASMDEV
 Reporting node . . . . . . . . . :   NODE1
 Consistent information in cluster:   Yes

                         Recovery Domain Information

                        Current      Preferred    Site       IP
 Node       Status      Node Role    Node Role    Name       Address
 NODE2      Active      *PRIMARY     *PRIMARY     LOCAL      10.0.1.11
                                                              10.0.2.11
 NODE1      Active      *BACKUP  1   *BACKUP  1   REMOTE     10.0.1.12
10.0.2.12 Number of recovery domain nodes  :   2
```

Figure 6-46 DSPCRGINF command

17. On the Set up High Availability Solution window click **Go** to perform the Verify administrative switchover task from NODE2 (new primary) to NODE1(backup) (Figure 6-42 on page 119). On the Step Verify administrative switchover completed window click **Close** to return to the Set up High availability Solution window (Figure 6-45 on page 121).

6.6.3 Migrating user profile

When the user profile migration is done:

1. The job description used by the user profile is duplicated into library QUSRHASM:

   ```
   CRTDUPOBJ OBJ(JOBDname ) FROMLIB(LIBname ) OBJTYPE(*JOBD ) TOLIB(QUSRHASM)
   ```

2. The INLASPGRP parameter of the jobd is changed from *none to HASMIASP:

   ```
   CHGJOBD JOBD(QUSRHASM/JOBDname ) INLASPGRP(HASMIASP)
   ```

3. The jobd is added to the cluster administrative domain HASMADMDMN:

   ```
   ADDCADMRE CLUSTER(HASMCLU ) ADMDMN(HASMADMDMN) RESOURCE(JOBDname) RSCTYPE(*JOBD
   ) RSCLIB(QUSRHASM )
   ```

4. The jobd parameter in the user profile is changed to the new one in QUSRHASM:

   ```
   CHGUSRPRF USRPRF(profile) JOBD(QUSRHASM/JOBDname)
   ```

5. The *usrprf is added to the cluster administrative domain HASMADMDMN:

   ```
   ADDCADMRE CLUSTER(HASMCLU ) ADMDMN(HASMADMDMN) RESOURCE(JORDAN) RSCTYPE(*USRPRF )
   ```

Take the following steps:

1. On the Set up high availability solution window click **Go** to perform next step, Migrate user profiles, as shown in Figure 6-47.

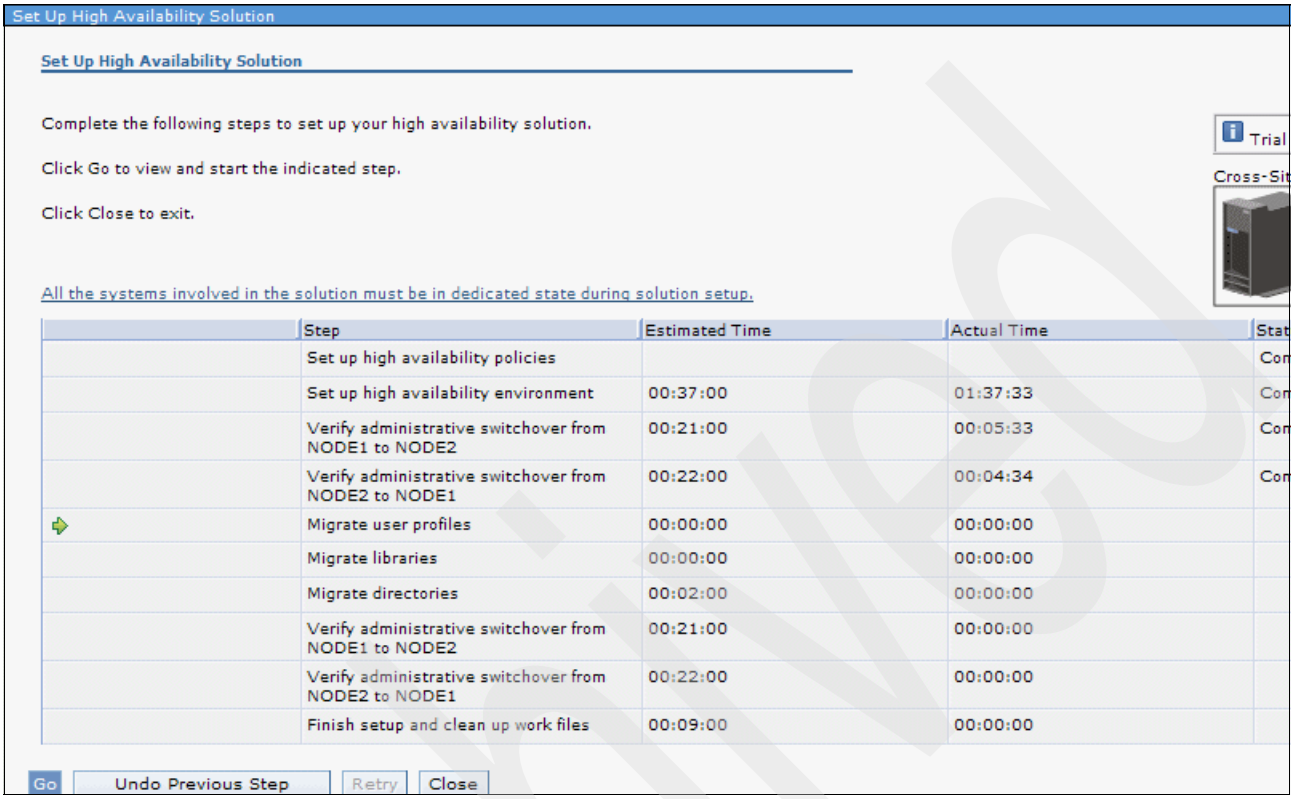


Figure 6-47 Migrate user profiles

2. On the Migrate user profiles window click the boxes in the Select column to select the user profile to migrate, as shown in Figure 6-48, then click **Migrate**.

   If a user profile cannot be migrated click the context menu next to the user profile name and select **view restriction**.

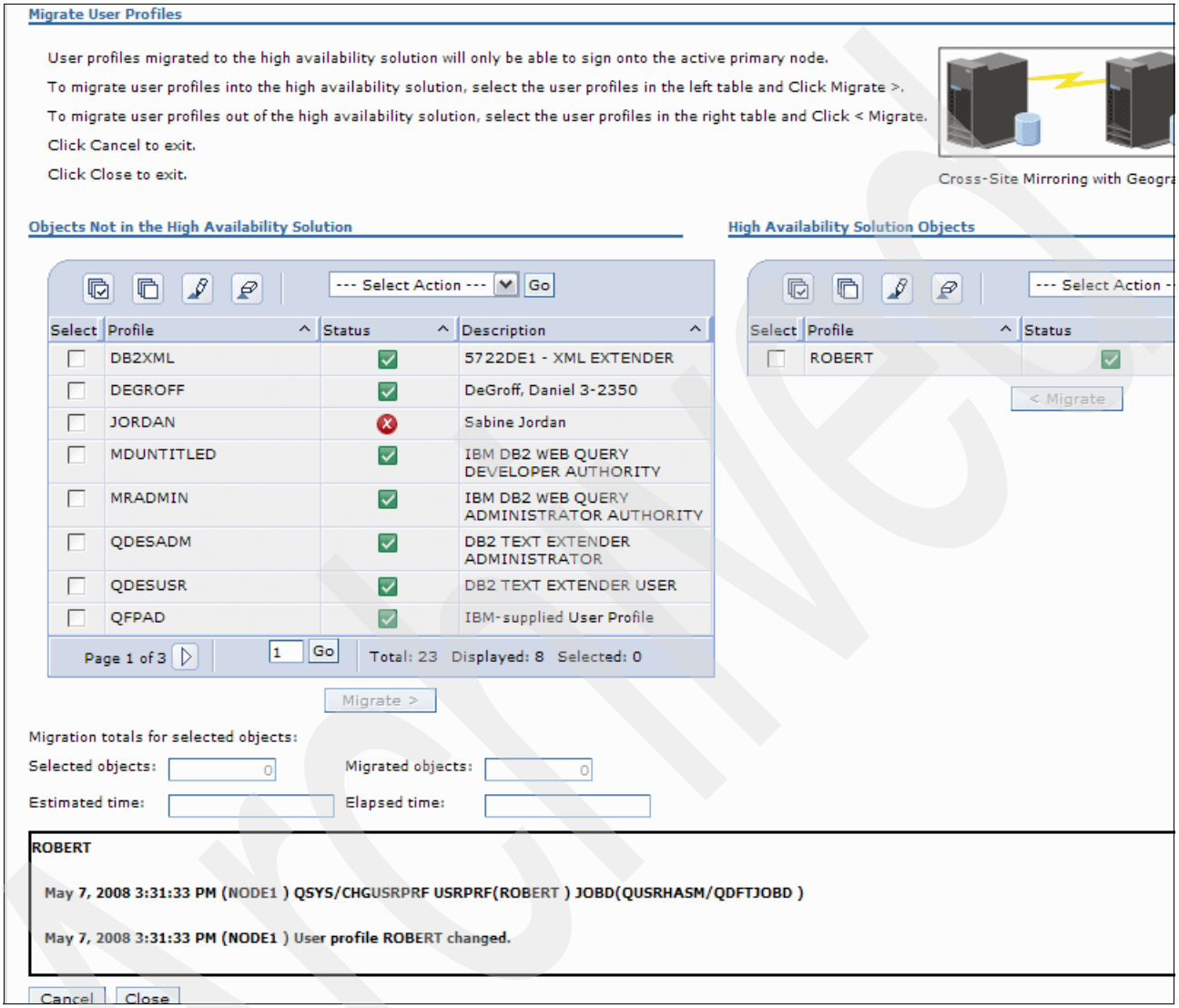


Figure 6-48 Migrate user profiles example

3. On Migrate user profiles Completed window click **Close** to return to the Set up high availability solution window (Figure 6-49).

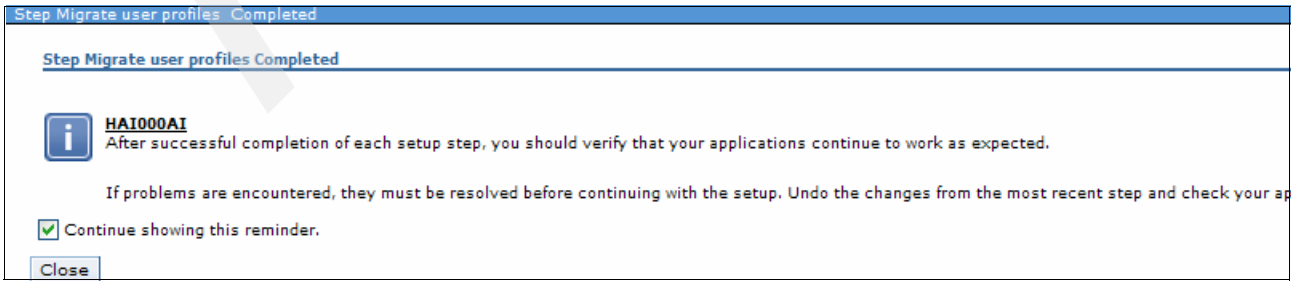


Figure 6-49 MIgrate user profiles completed

6.6.4 Migrating libraries

To migrate libraries:

1. Continue the setup of the high availability solution with the step Migrate libraries. Click **Go** to start it, as shown in Figure 6-50.

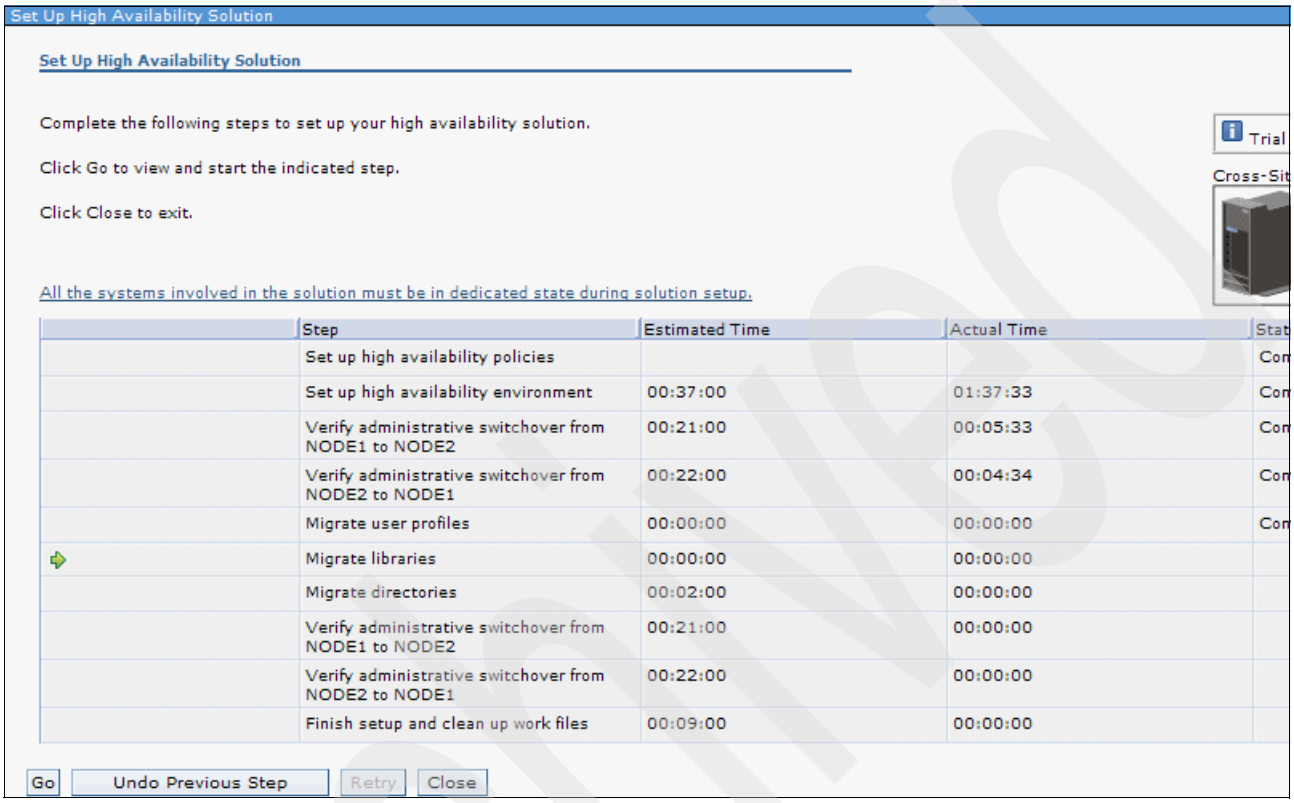


Figure 6-50 Set up Migrate libraries step

2. On the Migrate libraries window select the device to be used for the migration task from the Device used by migration drop-down menu, as shown in Figure 6-51.

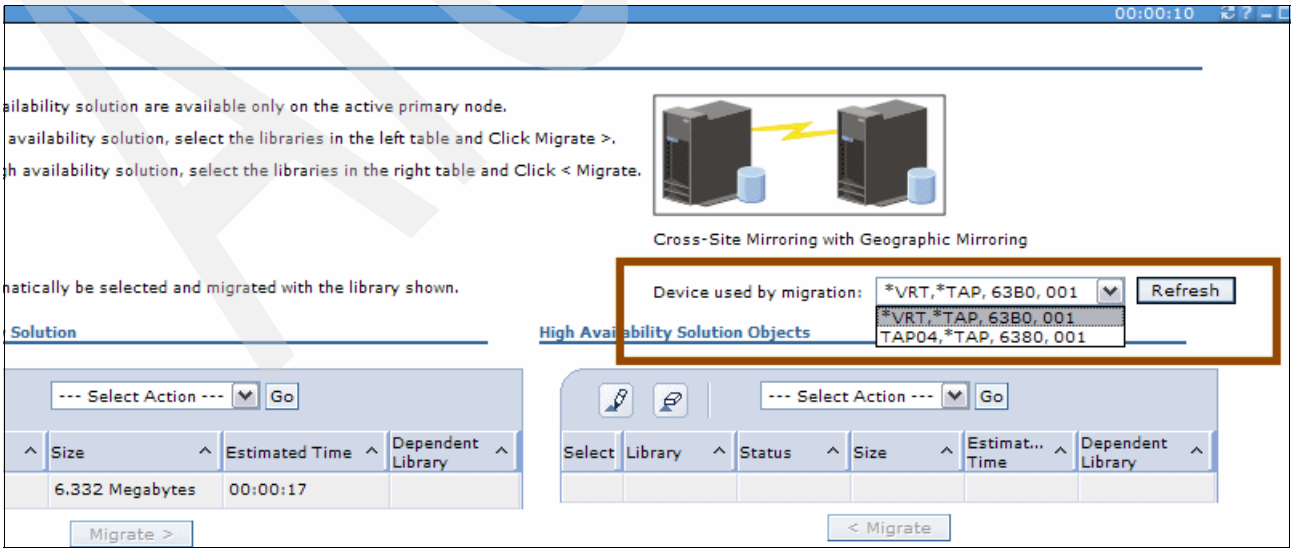


Figure 6-51 Tape device selection

3. Select the libraries to be migrated from the Objects Not in the High Availability Solution table. Click the box in the Select column, then click **Migrate** (Figure 6-52).

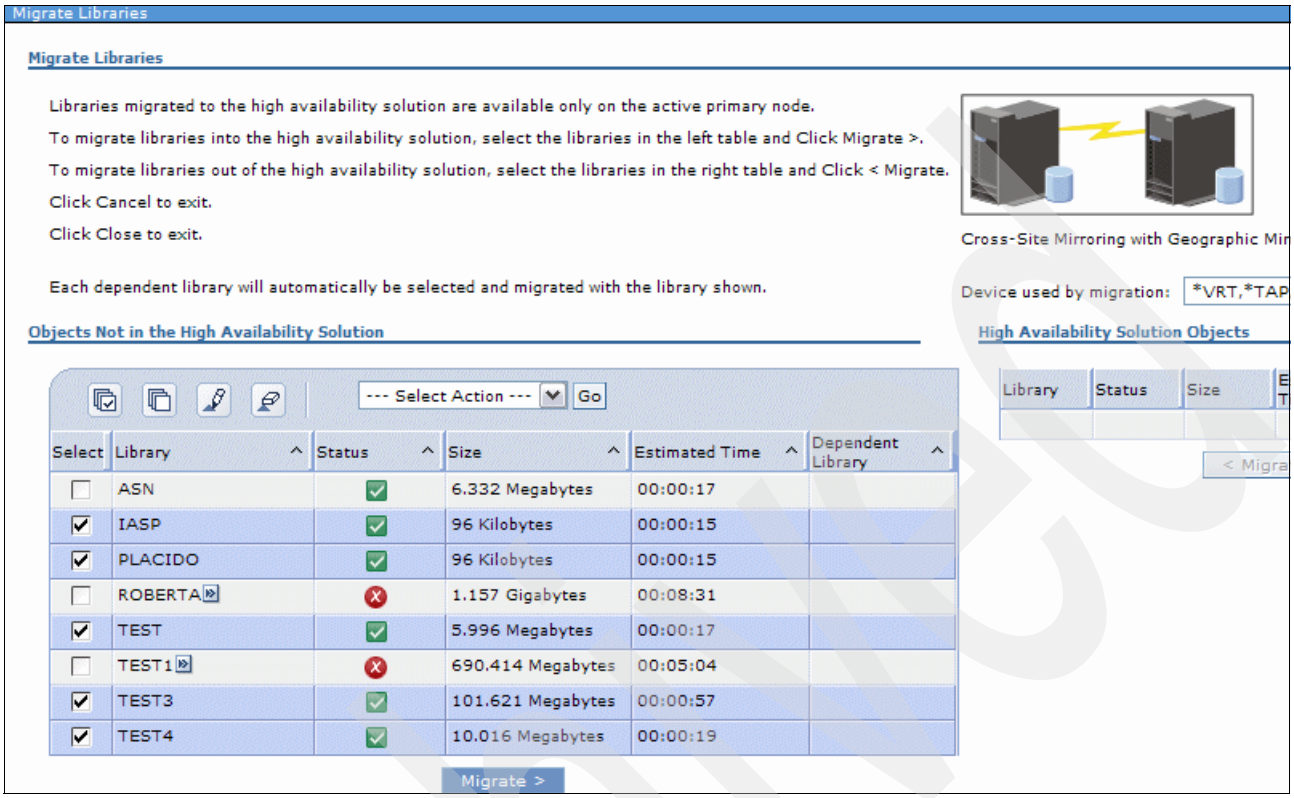


Figure 6-52 Migrate libraries selection

If a library shows a failed status it that means that it cannot be migrated. As you saw in the previous window, libraries ROBERTA and TEST cannot be migrated. See the context menu next to the library name and select **View restrictions,** as shown in Figure 6-53.

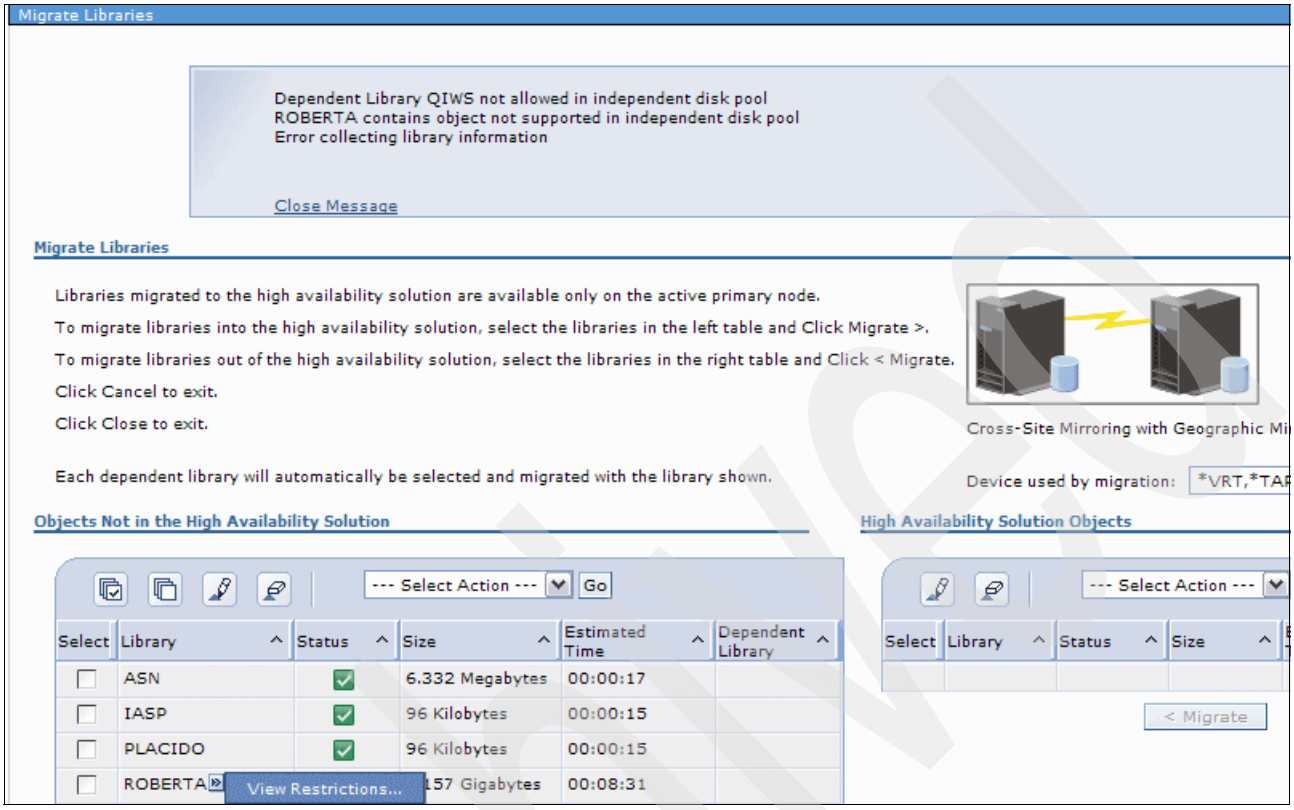


Figure 6-53 Migration library view restrictions

4. Once the migration is complete click **Close** to return to the Set Up High Availability Solution window, as shown in Figure 6-54.

   While the migration is running each library that has been migrated is moved from the table on the left (Objects Not in the High Availability Solution) to the table on the right (Objects in the High Availability Solution). See the message area at the bottom of the panel for completion messages and error messages logged during the migration step.

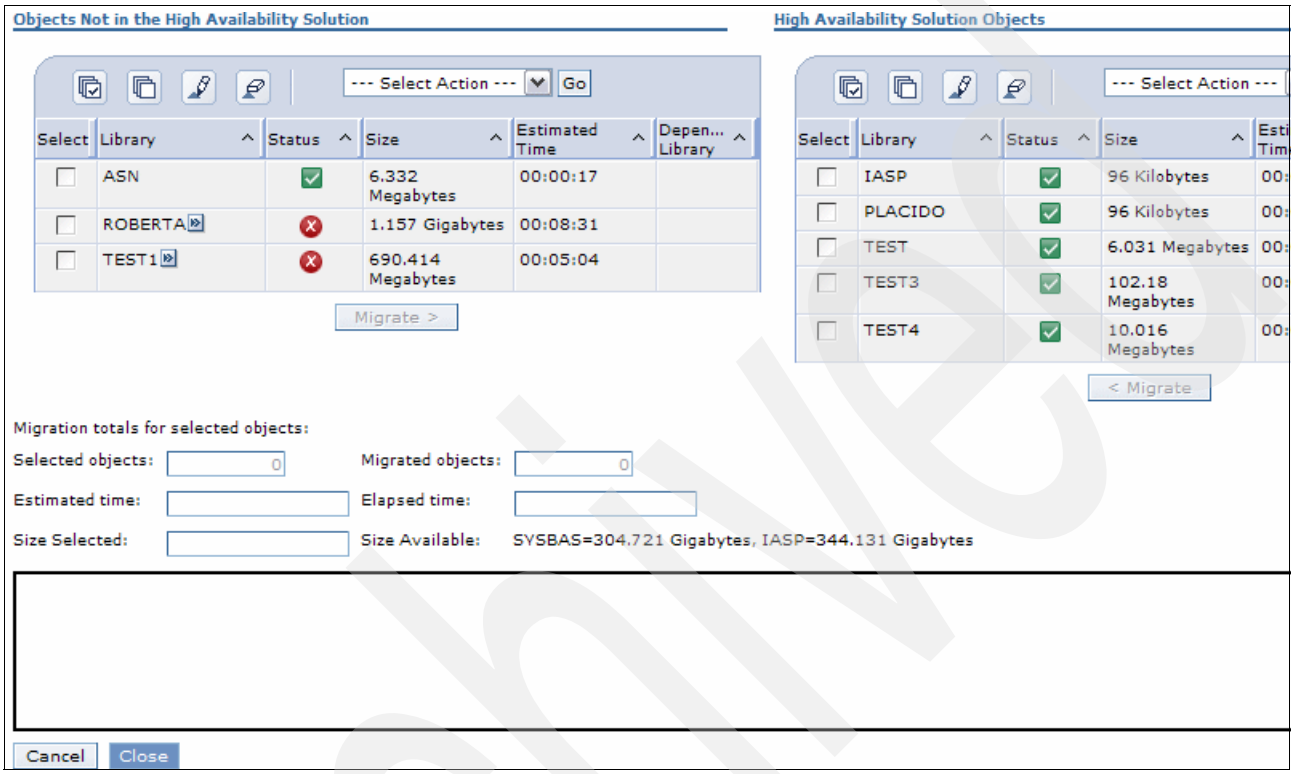


Figure 6-54 Migration libraries completed

5. On the Migrated library completion window click **Close** to return to the Set up High Availability Solution window (Figure 6-55).

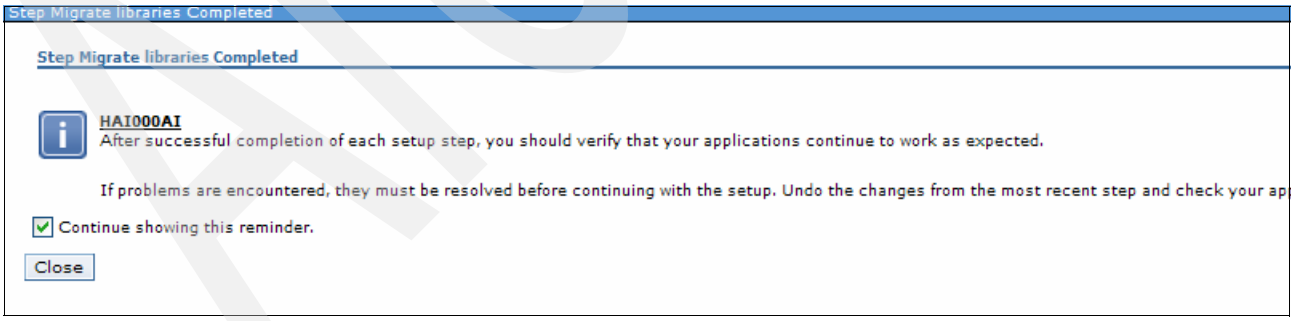


Figure 6-55 Migration library warning message

6.6.5 Migrating directories

To migrate directories:

1. On the Set up High Availability Solution window click **Go** to perform the Migrate Directories step (Figure 6-56).

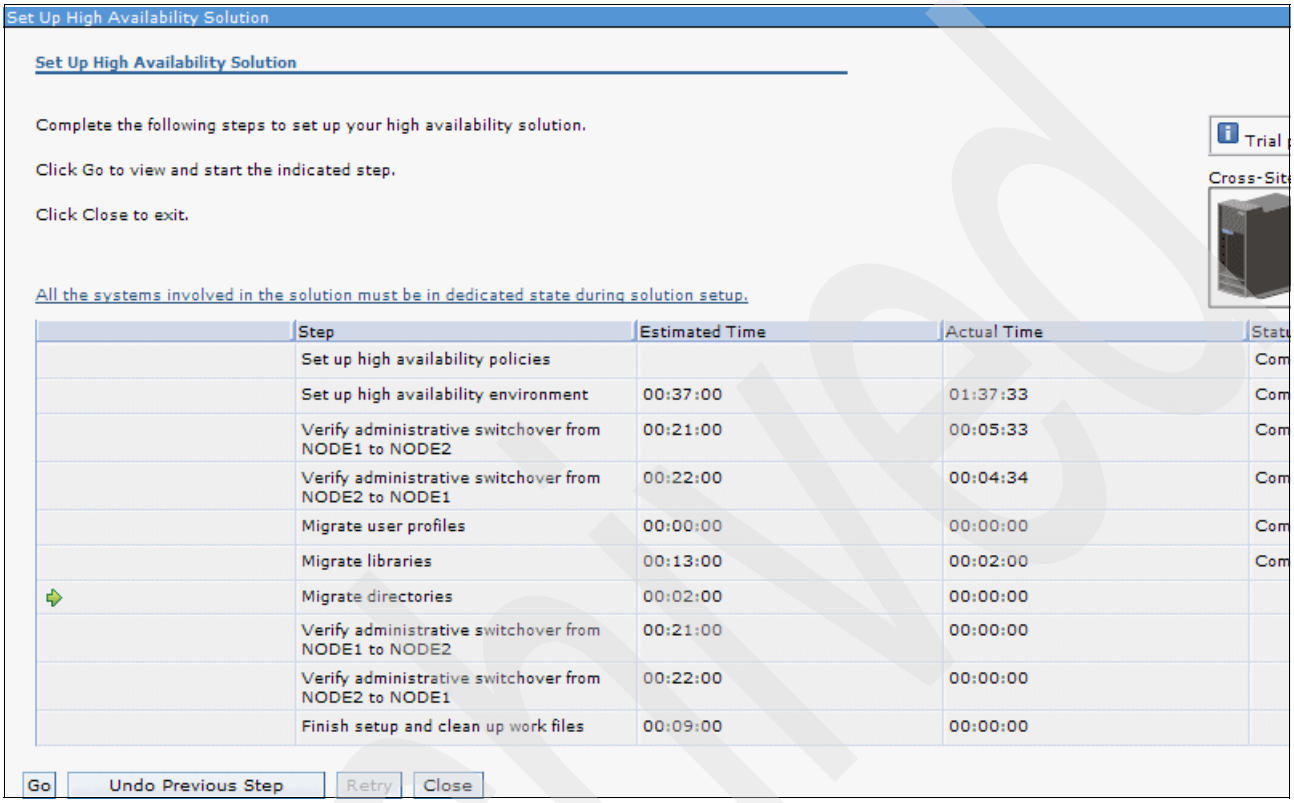


Figure 6-56 Migrate directories

2. On the Migrate directories window select the device from the Device used by migration drop-down menu, then click the box in the Select column of the Objects Not in the High Availability Solution table and choose the directories to migrate. Then click **Migrate**, as shown in Figure 6-57.

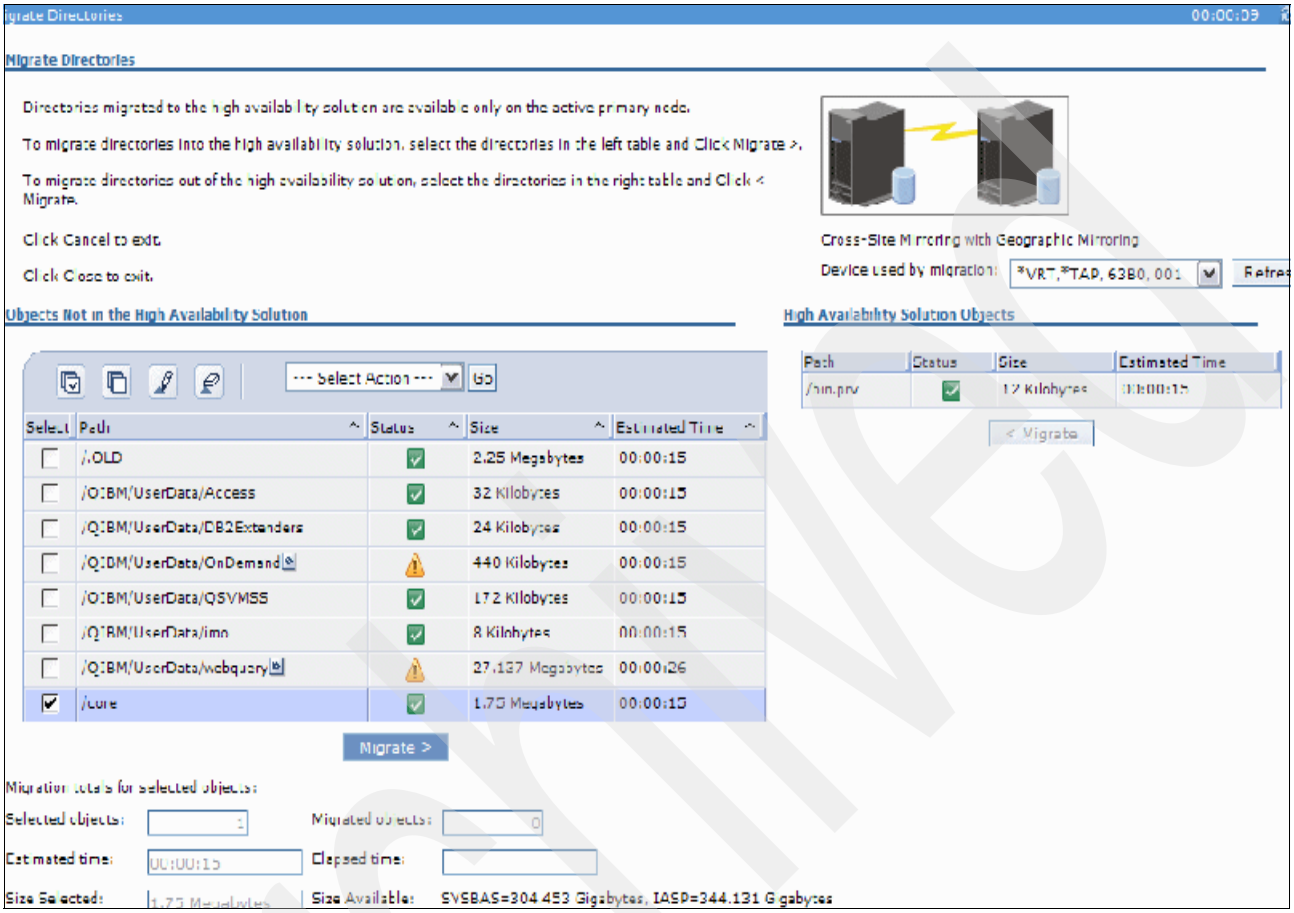


Figure 6-57 Directories migration

3. Click **Close** (Figure 6-58).

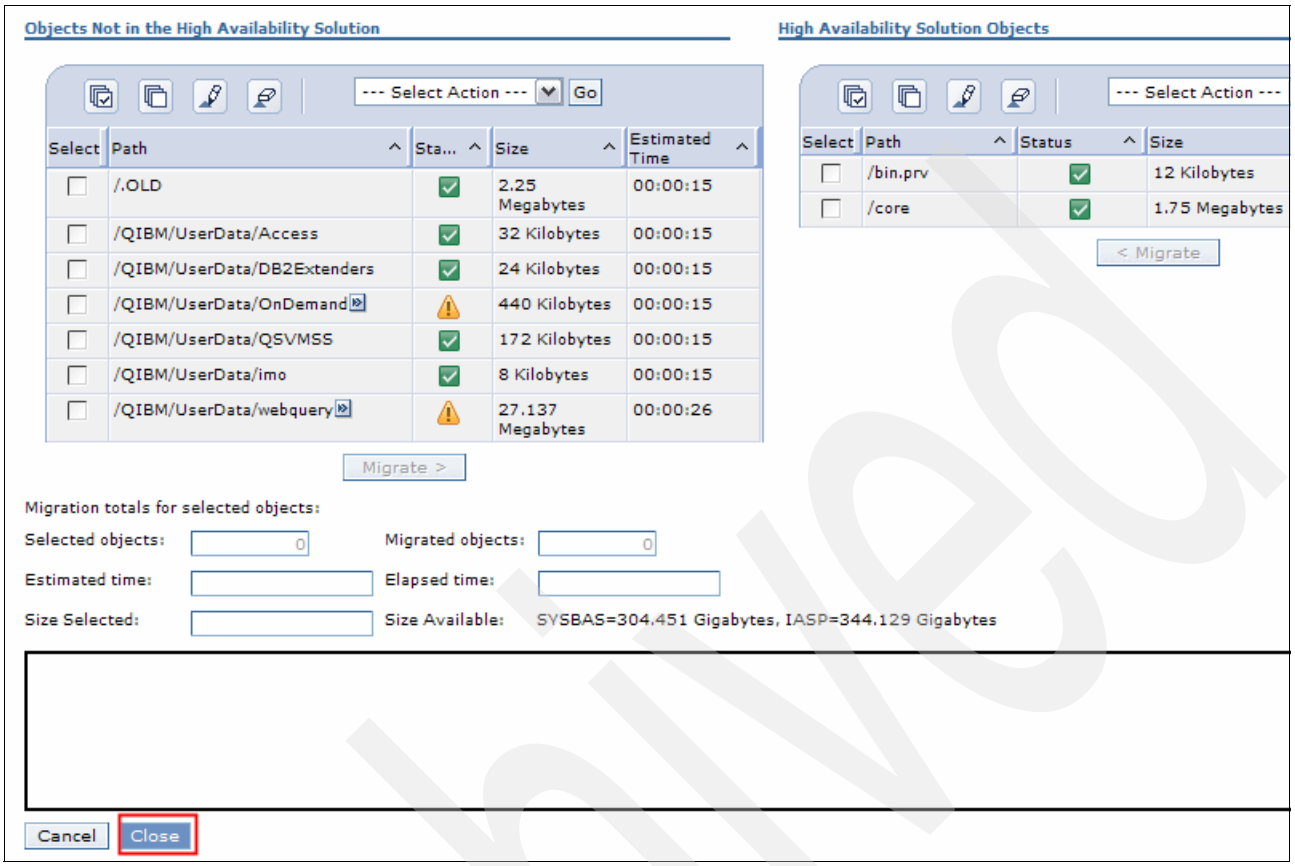


Figure 6-58 Migrate directory step

The directory /bin.prv has been migrated. As you saw in the previous window, it is now under the HA solutions objects table. The following command was run during this migration step:

```
RST DEV('/QSYS.LIB/HASMTAP.DEVD') OBJ(('/bin.prv' *INCLUDE
'/HASMIASP/bin.prv')) ALWOBJDIF(*ALL) PVTAUT(*YES) CRTPRNDIR(*YES)
PRNDIROWN(QSYS)
```

4. Click **Close** to return to the set up High availability Solution main panel, as shown in Figure 6-59.

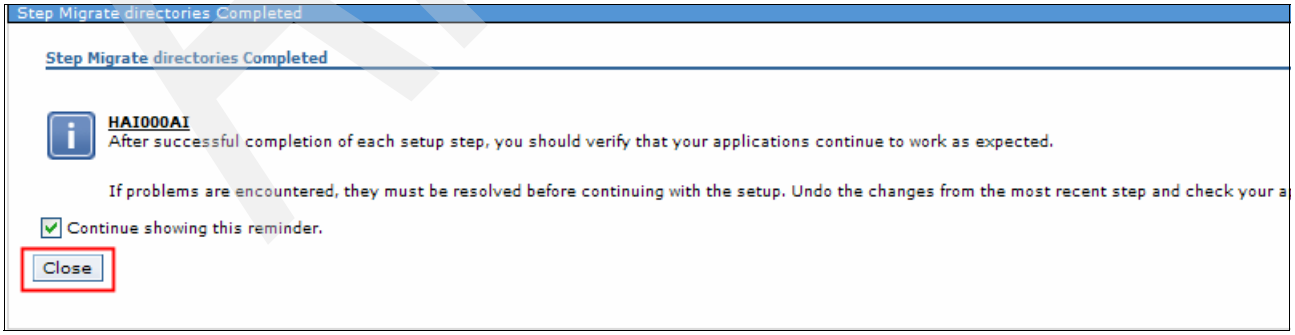


Figure 6-59 Migrate directory completed

Run WRKLNK from the command line to verify that the symbolic links have been created for the directories that have been migrated into the iASP, as shown in Figure 6-59 on page 131.

```
  Work with Object Links

  Directory  . . . . :   /

  Type options, press Enter.
    2=Edit   3=Copy   4=Remove   5=Display   7=Rename   8=Display attributes
    11=Change current directory ...

  Opt   Object link            Type     Attribute    Text
        bin                    DIR
        bin.prv                SYMLNK
        core                   SYMLNK
        dev                    DIR
```

Figure 6-60 wrklnk command

6.6.6 Switching

To switch:

1. On the Set up High Availability main menu click **GO** to run the Verify administrative switchover from NODE1 to NODE2, as shown in Figure 6-61.

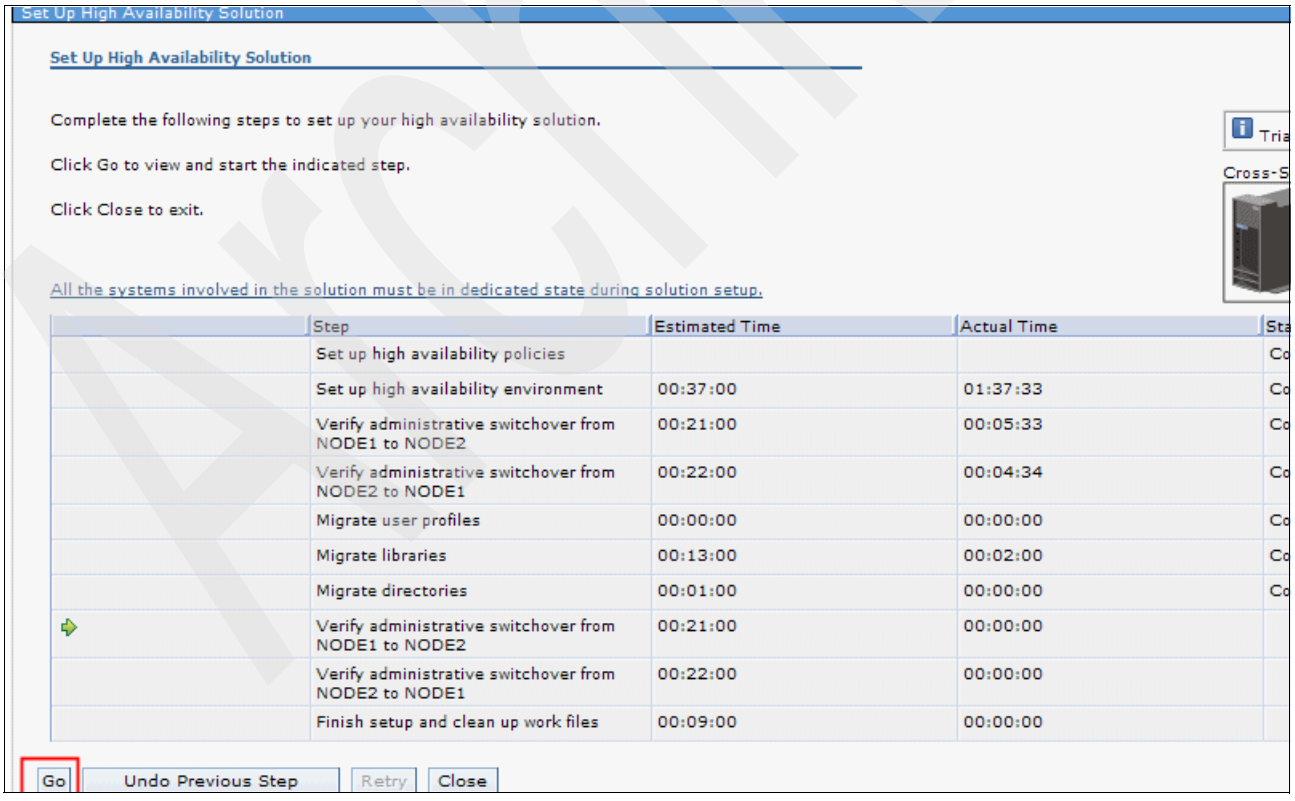


Figure 6-61 Verify administrative switchover

2. Click **Run now**, as shown in Figure 6-62.

   The following substeps verify an administrative switchover from NODE1 to node NODE2. This will verify that the high availability solution is ready to become operational on NODE2.

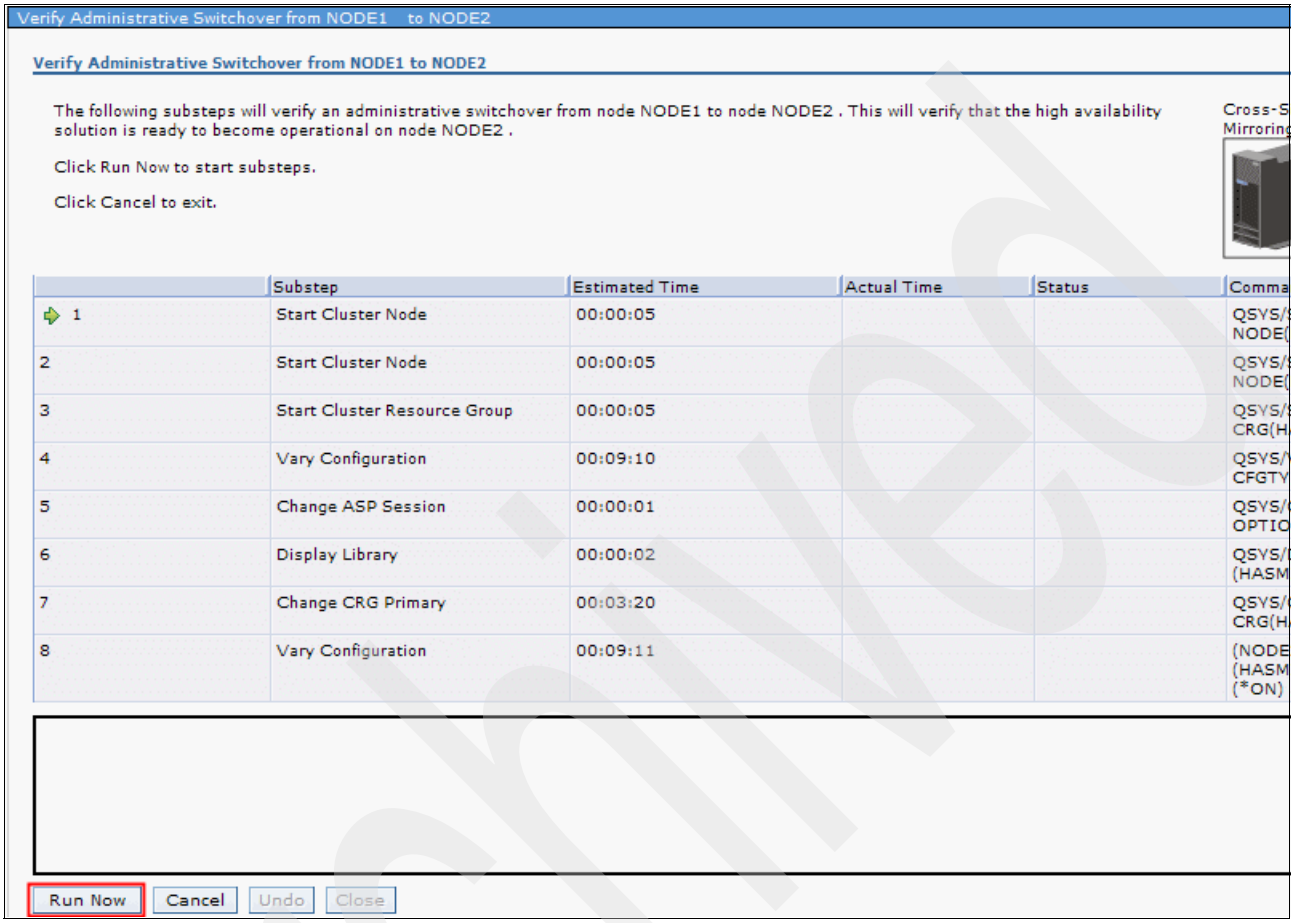


Figure 6-62 Switchover window

CHGCRGPRI is running as shown in Figure 6-63.

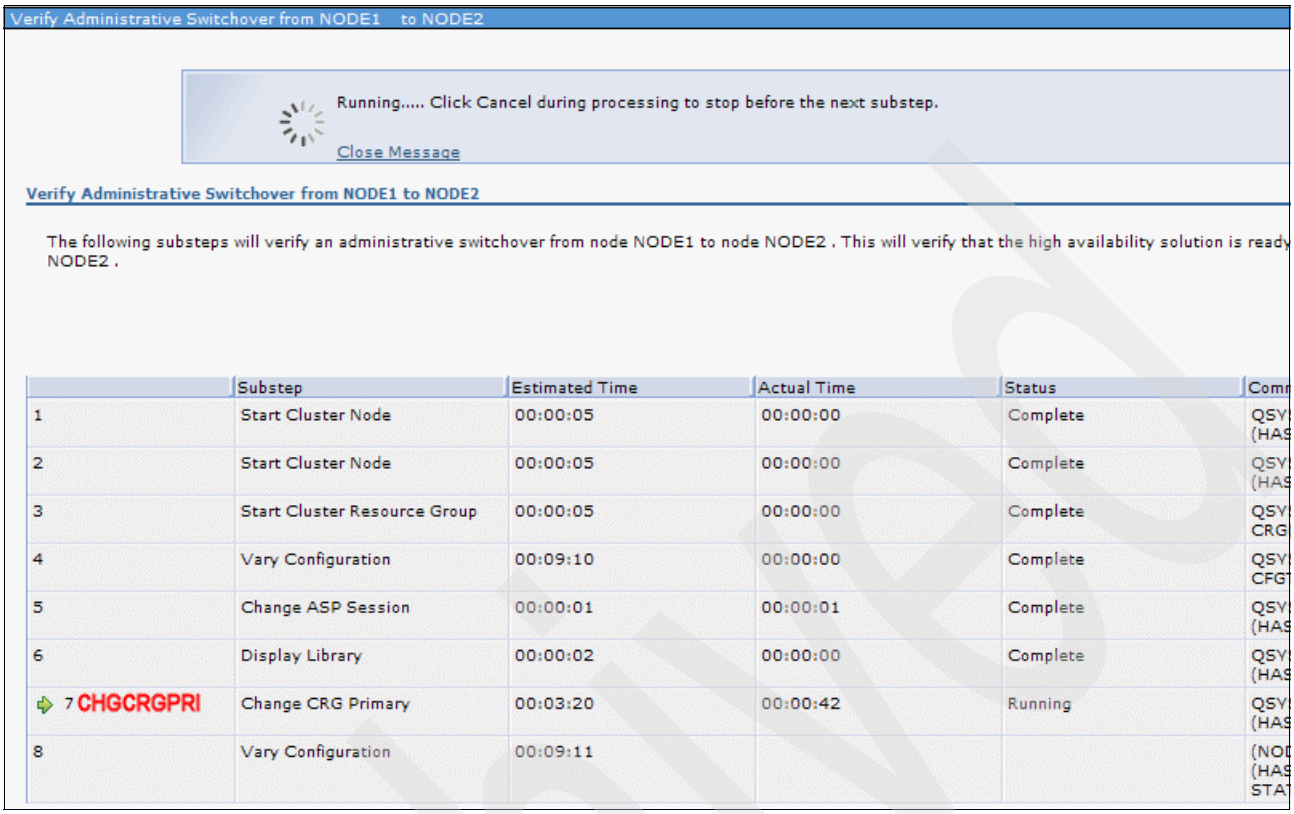


Figure 6-63 CHGCRGPRI command

3. Click **Close** (Figure 6-64).

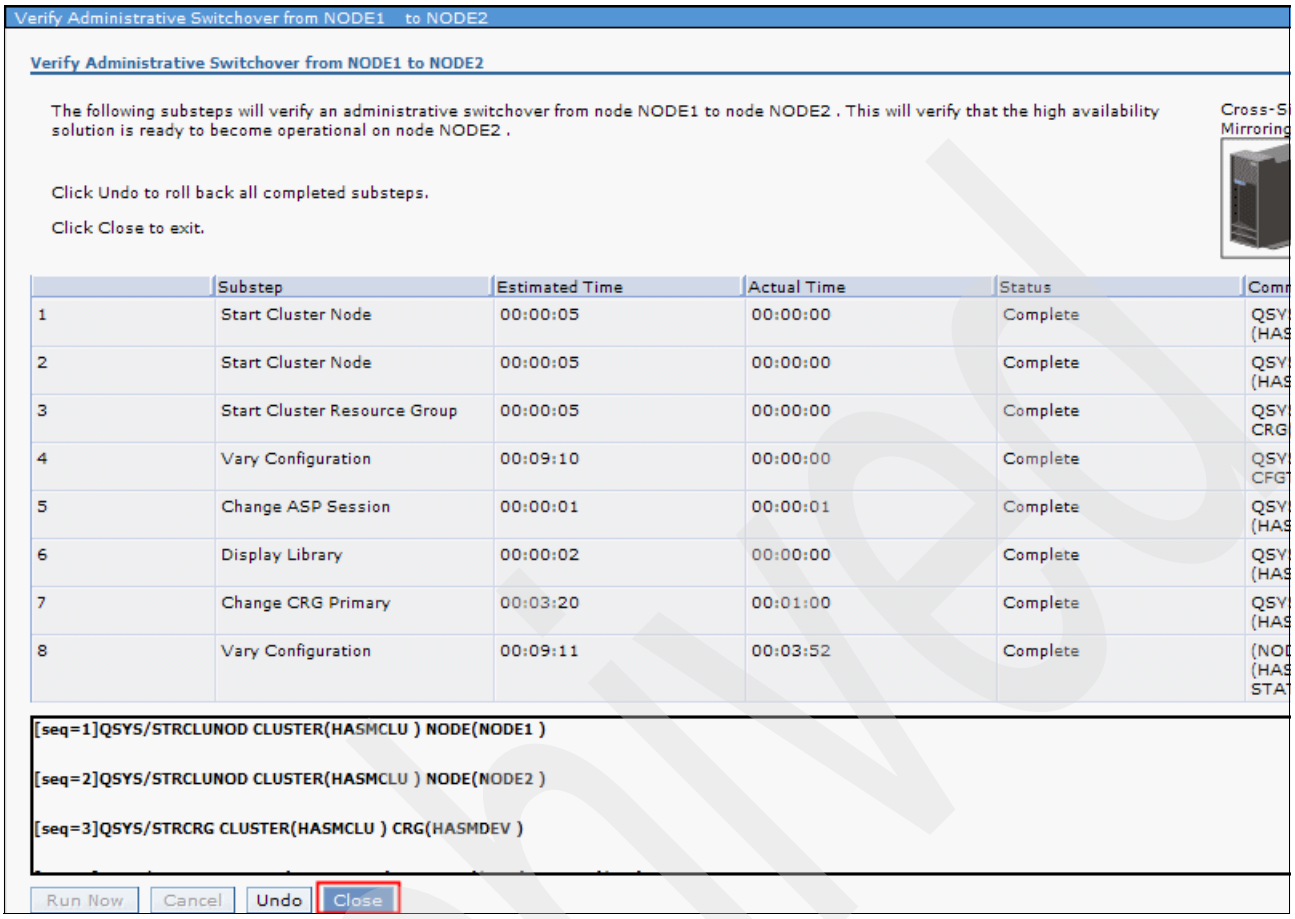


Figure 6-64 Switchover complete

4. Click **Close** to return to the Set up high availability solution main panel, as shown in Figure 6-65.

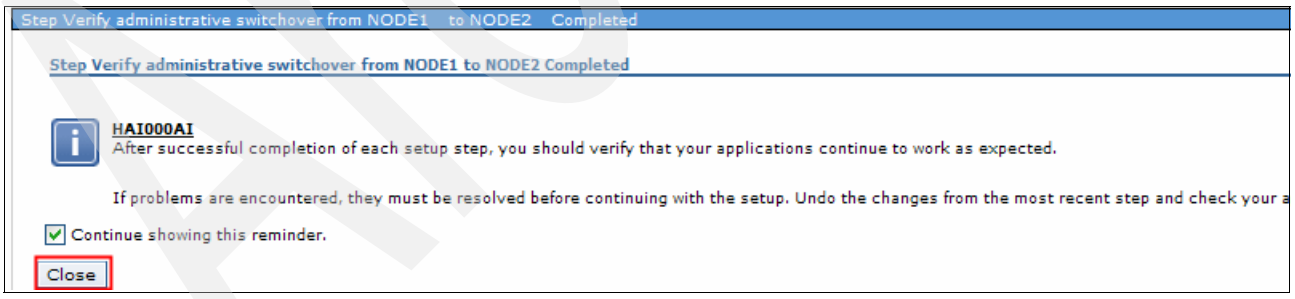


Figure 6-65 Switchover completed

Before running a switchover back to NODE1 test that all your applications are working as expected.

5. Perform the verify administrative switchover from NODE2 (new primary) to NODE1 (backup) (Figure 6-66).

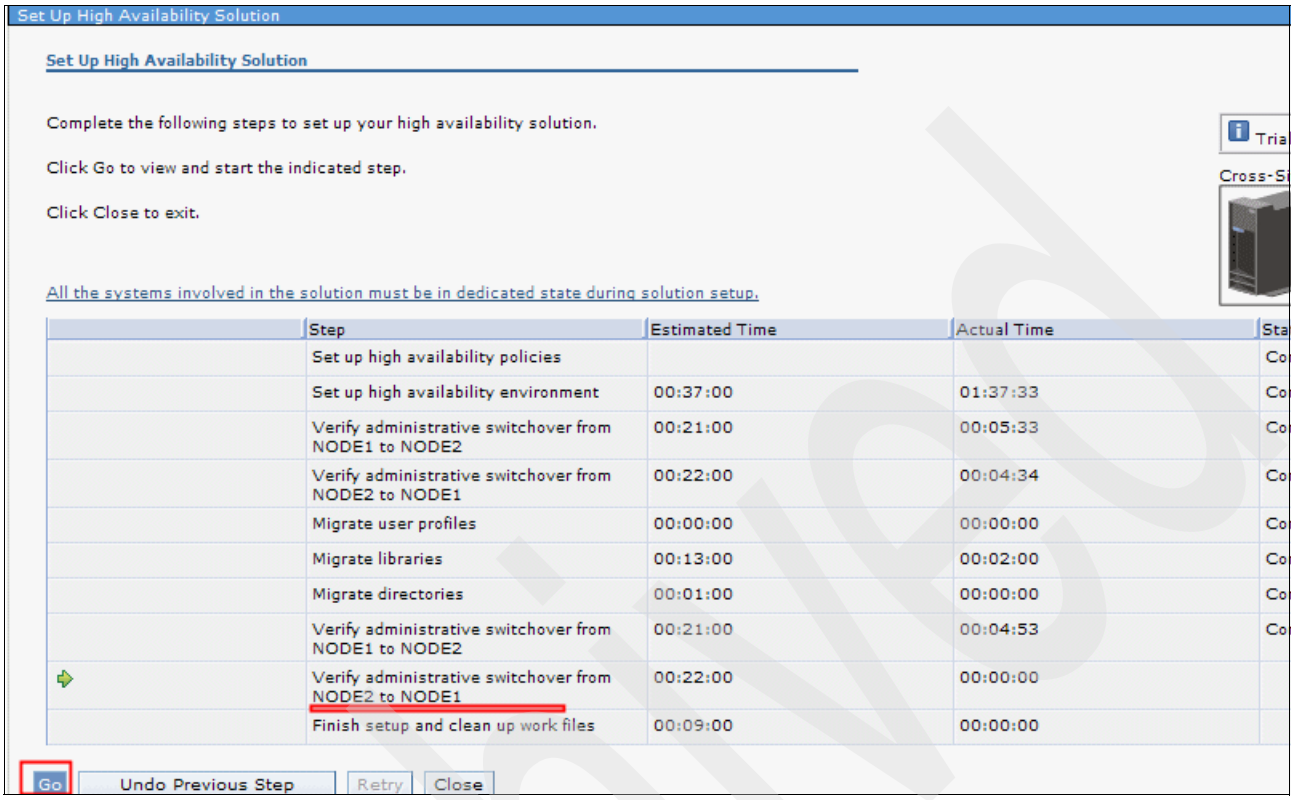


Figure 6-66 Verify administrative switchover

You will see in the next windows the same steps done to switch from NODE1 to NODE2. Click **Run now** to perform each substep, as shown in Figure 6-62 on page 133.

6. Click **GO** to perform the Finish setup and clean up work files task, as shown in Figure 6-67.

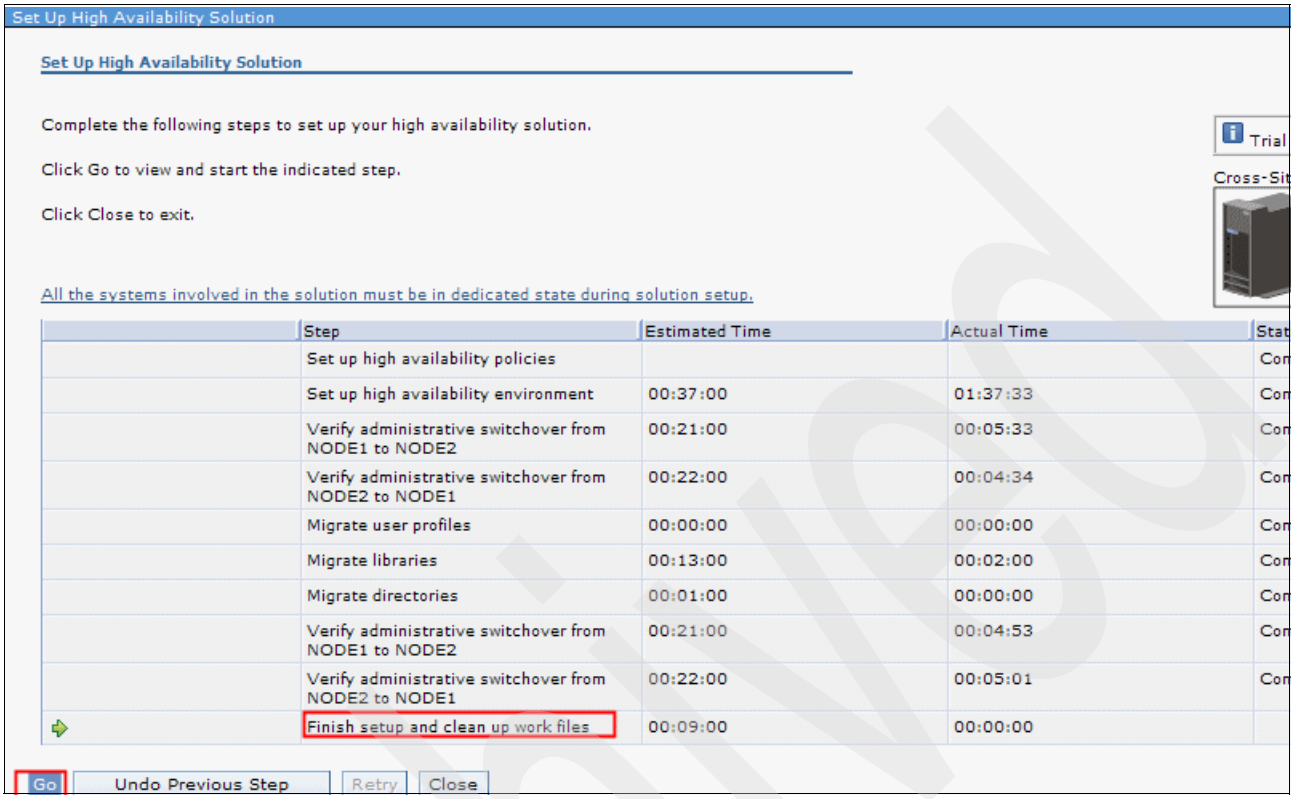


Figure 6-67 Finish set up and clean up work file task

7. Click **Run Now** on the next window, as shown in Figure 6-68.

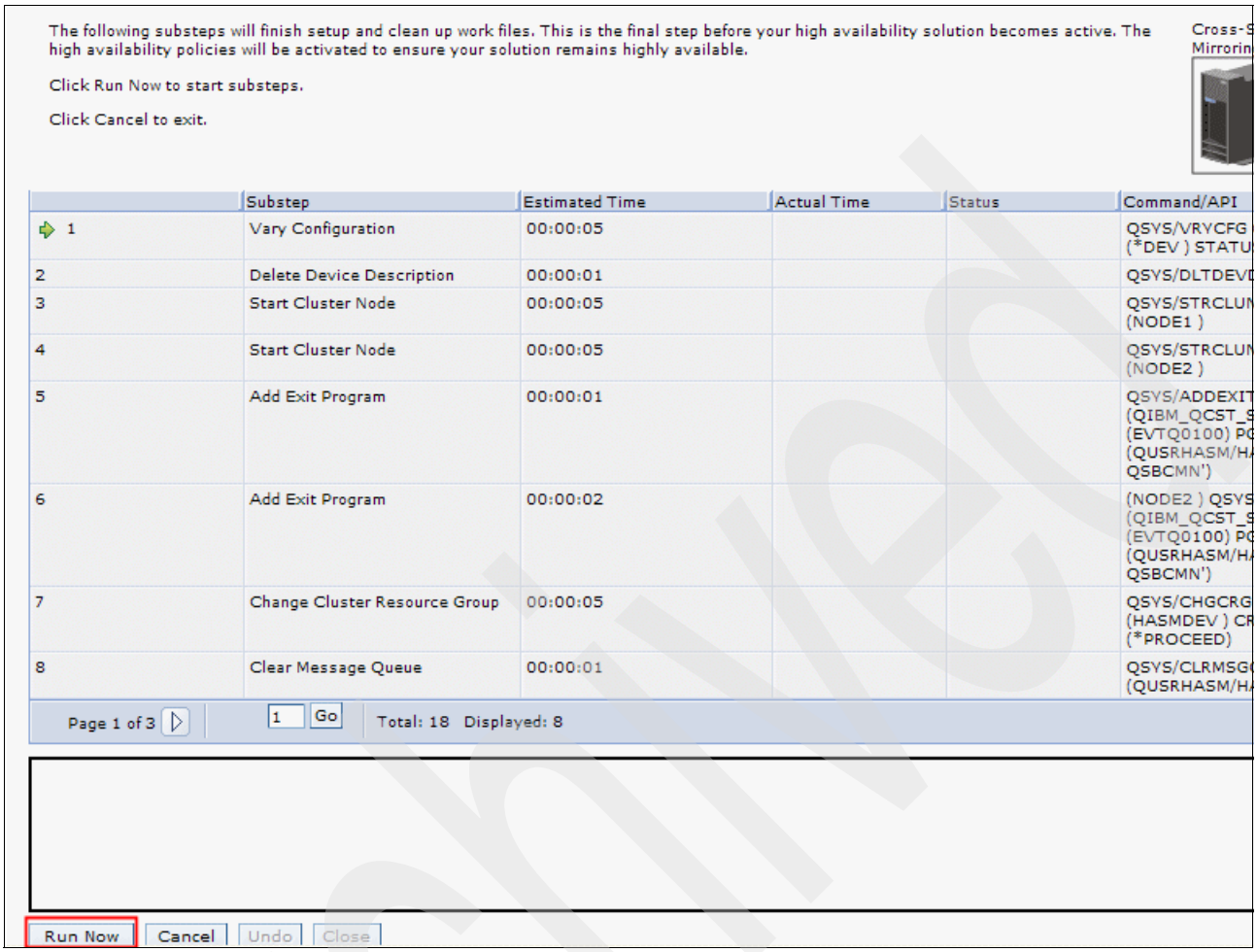


Figure 6-68 Finish and clean up work files window

8. A warning message appears next stating:

   ```
   You have selected to finish the setup of your high availability solution and
   clean up work files. Once started, you cannot undo previous setup steps because
   all setup steps will be permanently completed.
   ```

   If you have done all the tests to verify that your high availability solution is working click **Continue**.

9. Click **Close** to exit (Figure 6-69).

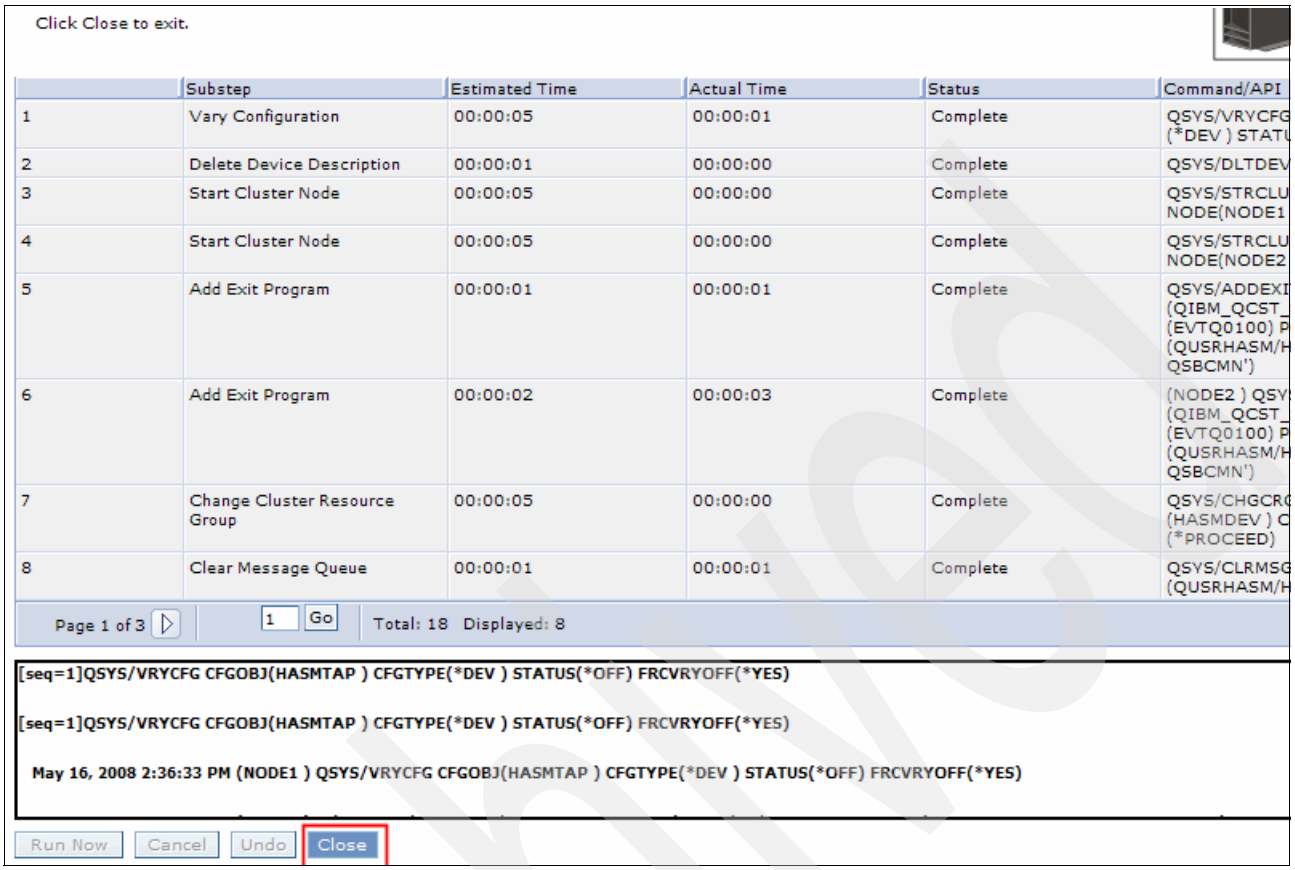


Figure 6-69 Finish set up step window

10. On next panel click **Close** to exit.

    All the setup substeps have been successfully completed (Status column shows Complete), as shown in Figure 6-70

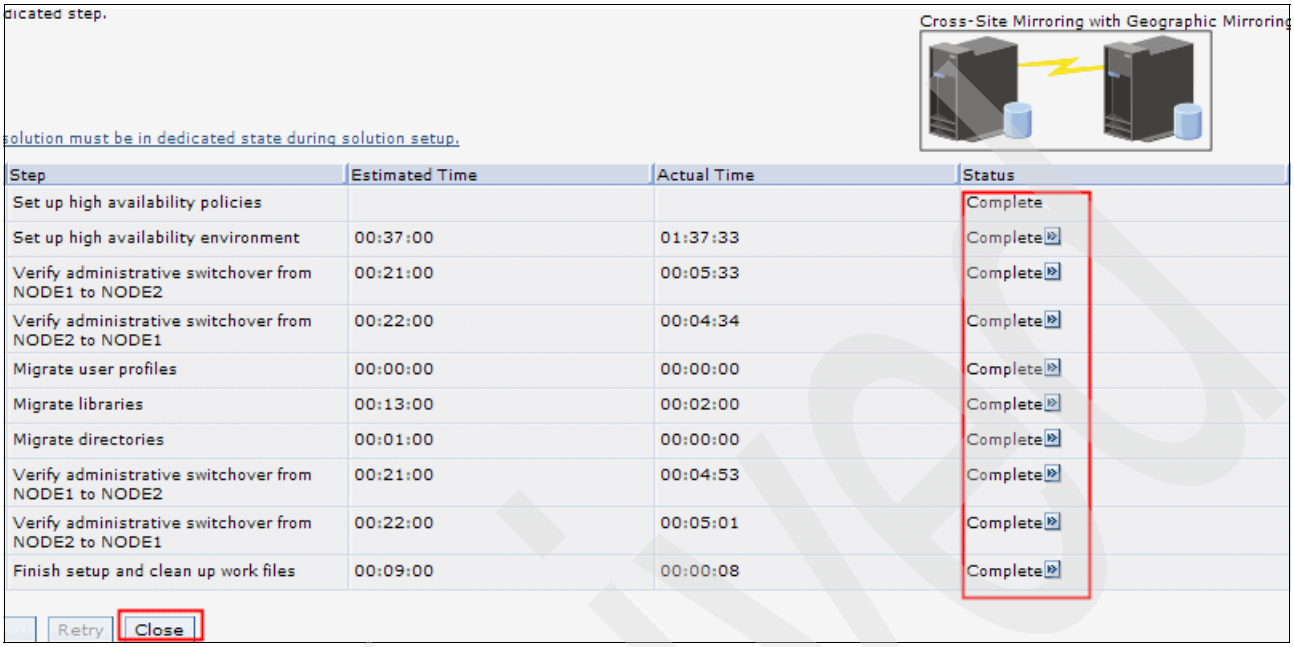


Figure 6-70 Set up substeps completed

On the next panel you see the Display Log option on each substep (Figure 6-71).

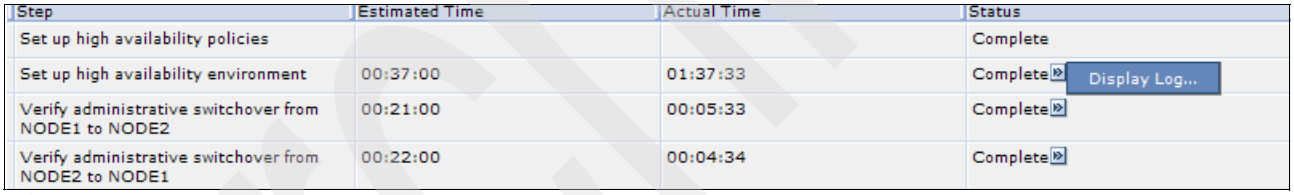


Figure 6-71 Display Log

11. Select **Display Log** and a window appears, as shown in Figure 6-72.

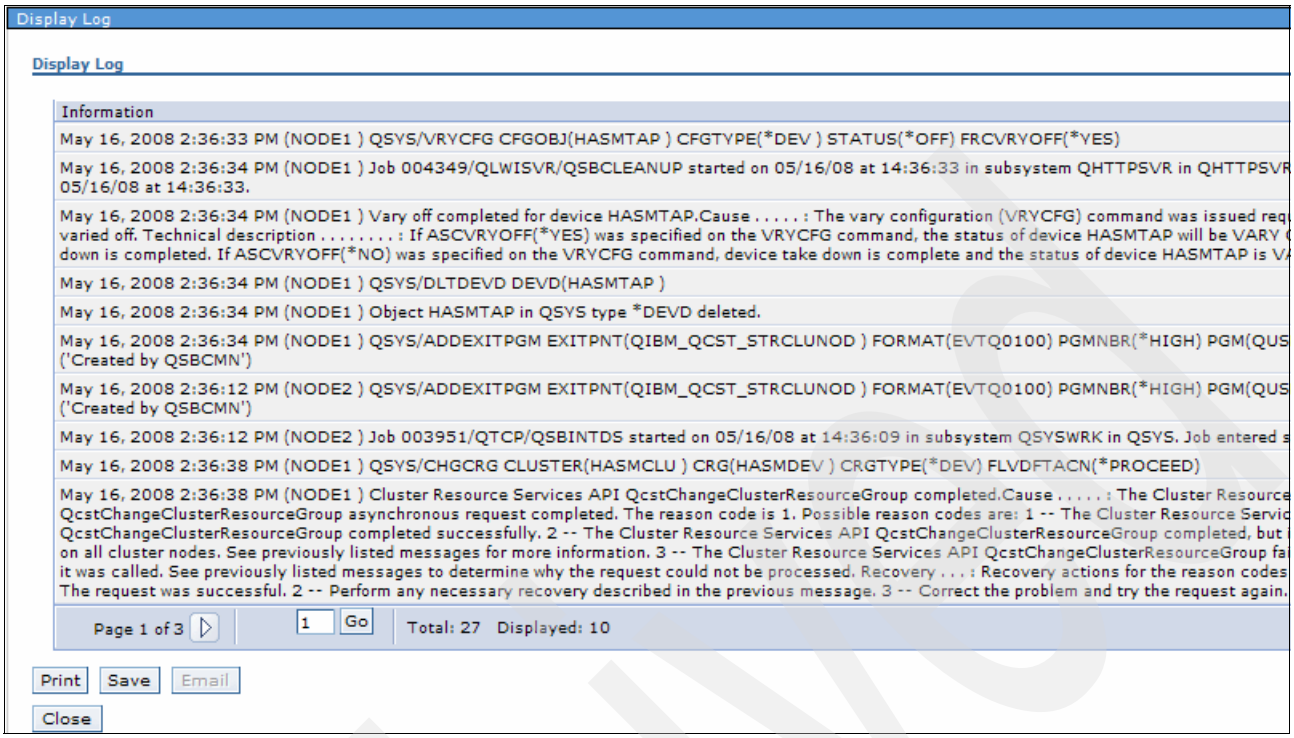


Figure 6-72 Display Log example

12. On the next window click **Close** to exit from the set up (Figure 6-73).

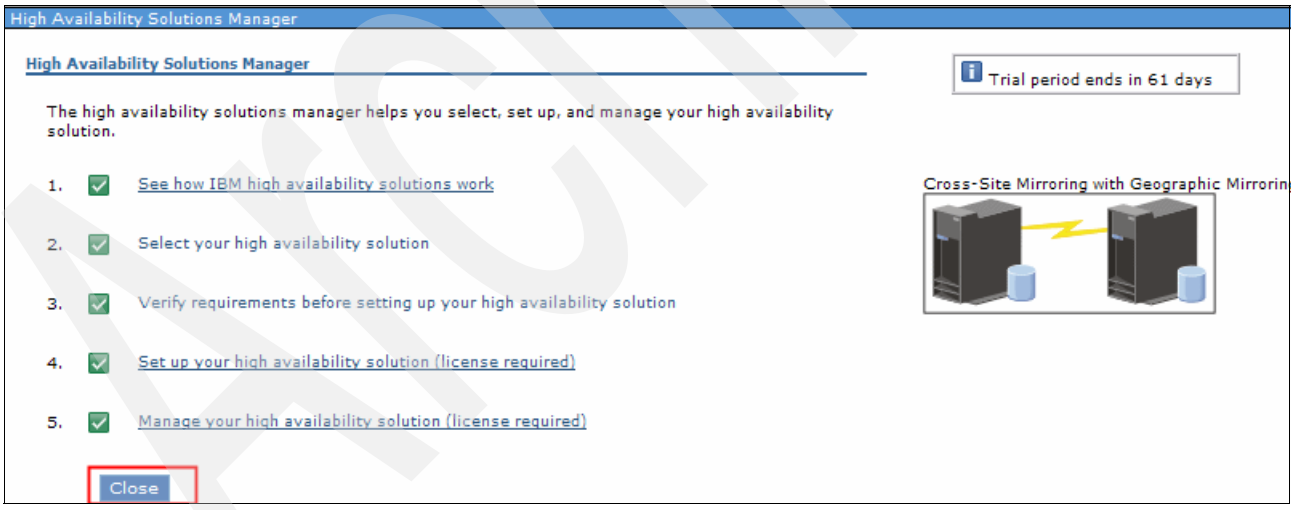


Figure 6-73 Set up completed

6.7 Managing your high availability solution

The High Availability Solutions Manager graphical interface allows you to manage your high availability solution through solution-level tasks that are generated dynamically based on the current status of your solution. In addition to these tasks, you can also manage high availability resources that comprise your solution and view event messages.

Start on the High Availability Solution Manager panel. (We showed how to get here in 6.2, “HASM GUI” on page 90.) Select **Manage your high availability solution**, as shown in Figure 6-74.

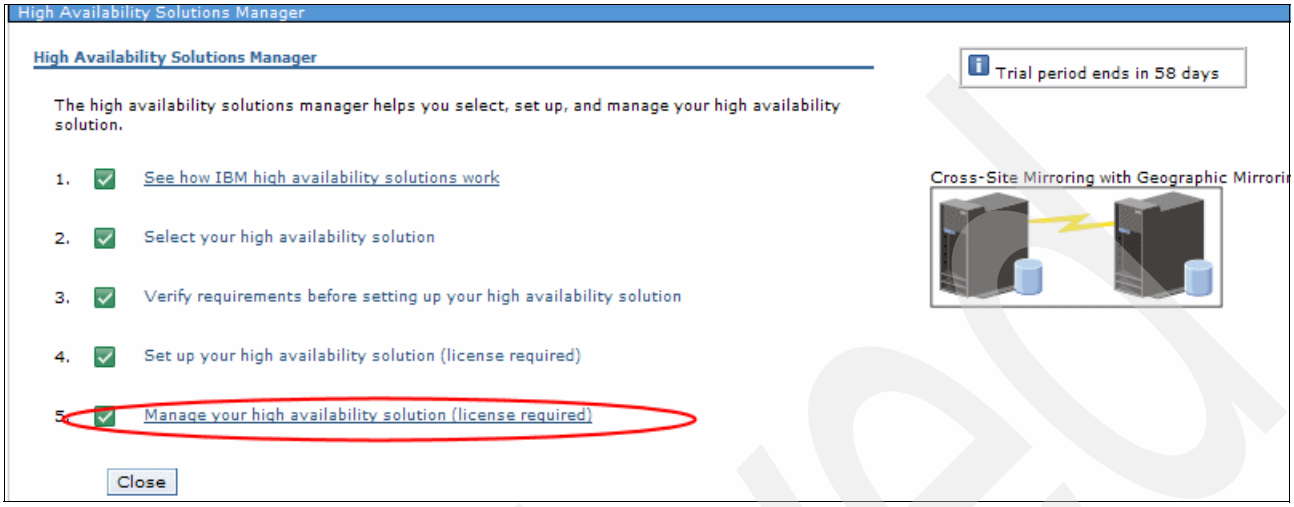


Figure 6-74 Manage your high availability solution

There are three sections on the Manage Your High Availability Solution page:

- The Manage Your High Availability Solution section provides an at-glance view of the status of the high availability solution and quick access to solution-level actions.
- The High Availability Solution Resources section provides a tabbed list of all of the high availability solution resources. Each tab gives a detailed view of each resource, along with possible actions to perform on the resource.
- The Event Log section presents the list of events that have occurred in the high-availability solution.

Once you select **Manage your high availability solution** a window appears, as shown in Figure 6-75.

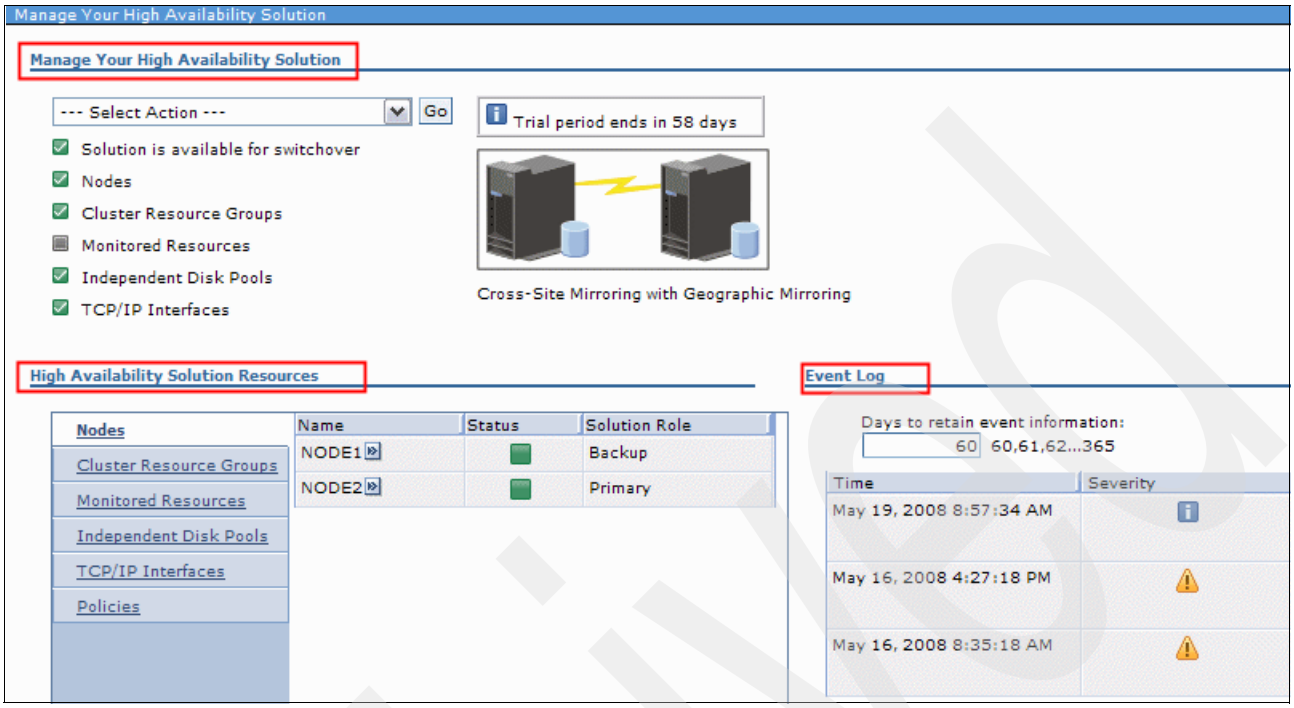


Figure 6-75 Manage Your High Availability Solution

Manage your high availability

Select the action that you want to perform from the Manage Your High Availability Solution window drop-down menu (see the red rectangle), then click **GO** to perform it. See Figure 6-76.

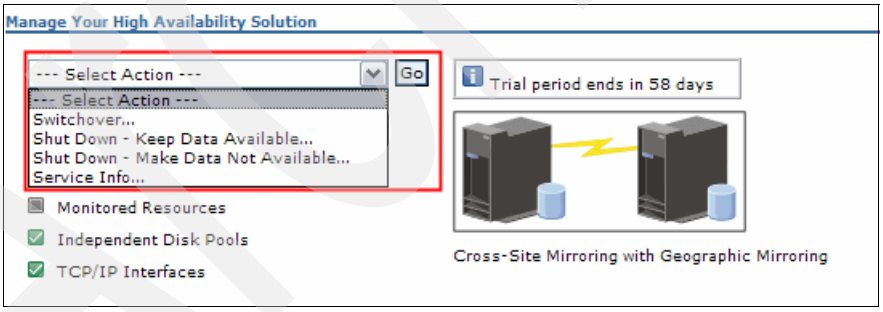


Figure 6-76 Select action window

The actions that can be done are listed in the drop-down menu.

Switchover

To do this:

1. Select the action **Switchover** and then click **Go** (Figure 6-77).

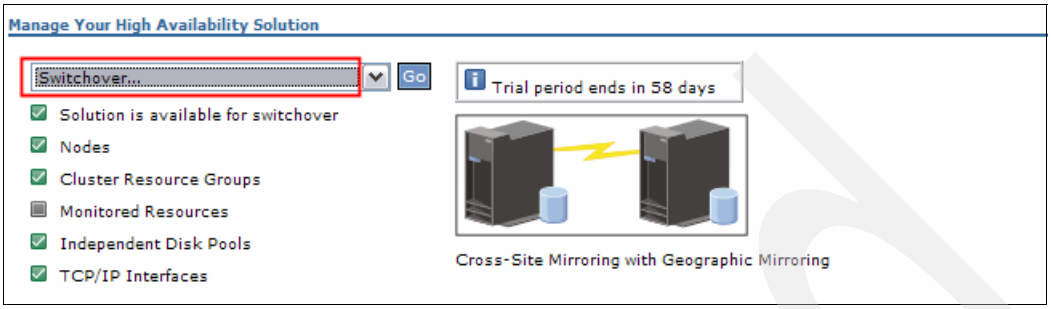


Figure 6-77 Switchover action

2. Click **Run Now**, as shown in Figure 6-78.

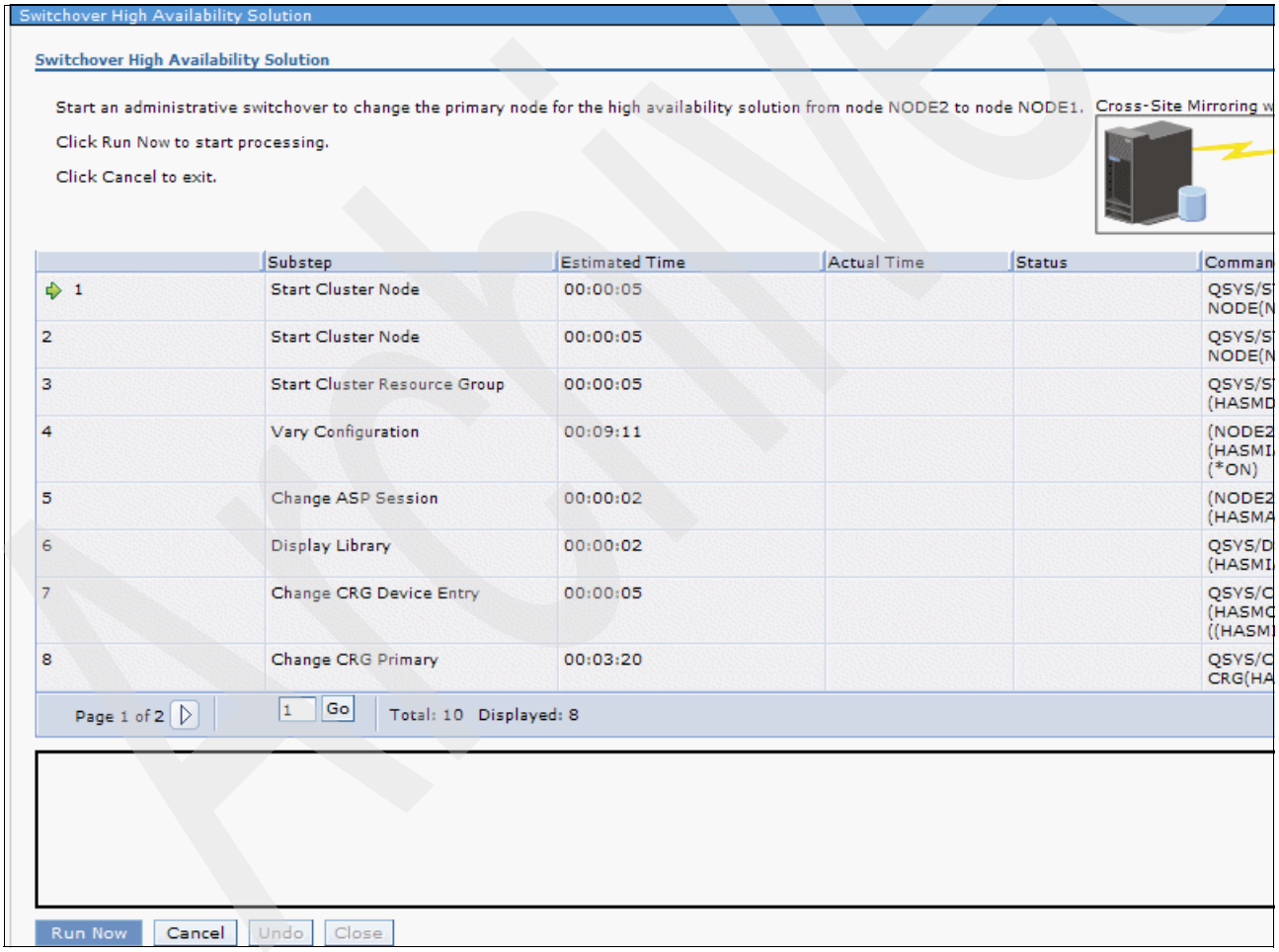


Figure 6-78 Perform switchover

Shut Down - Keep Data Available and Shut Down - Make data not available

Take the following steps:

1. Select the action **Shut Down-Keep data available/or Shut Down Make data not available** from the drop-down menu and click **Go**, as shown in Figure 6-79.

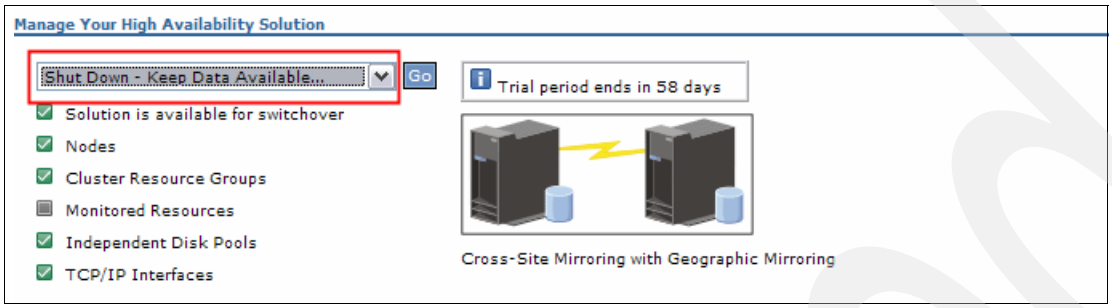


Figure 6-79 Shut Down - Keep Data Available action

2. Click **Run Now** to perform the action shown in Figure 6-80 (Shut down keep data available).

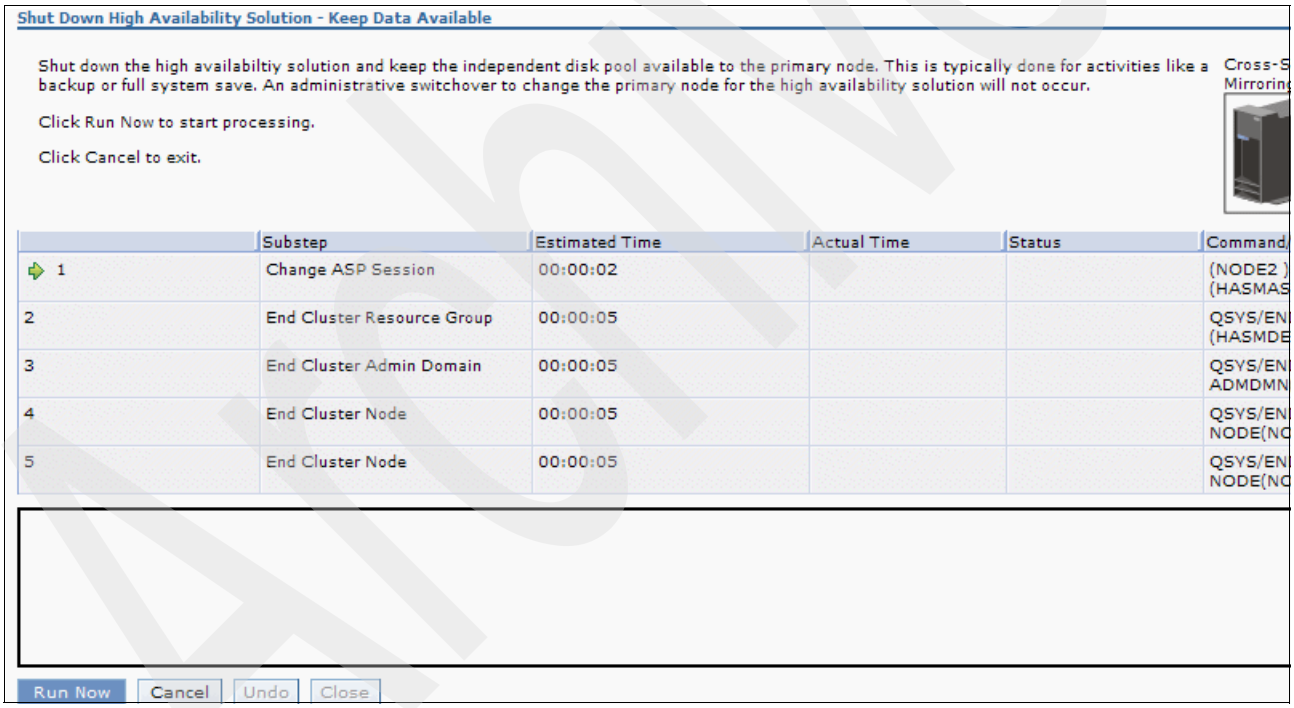


Figure 6-80 Perform shut down - Keep Data Available

3. Click **Run Now** on the Shut Down - Make data not available window, as shown in Figure 6-81.

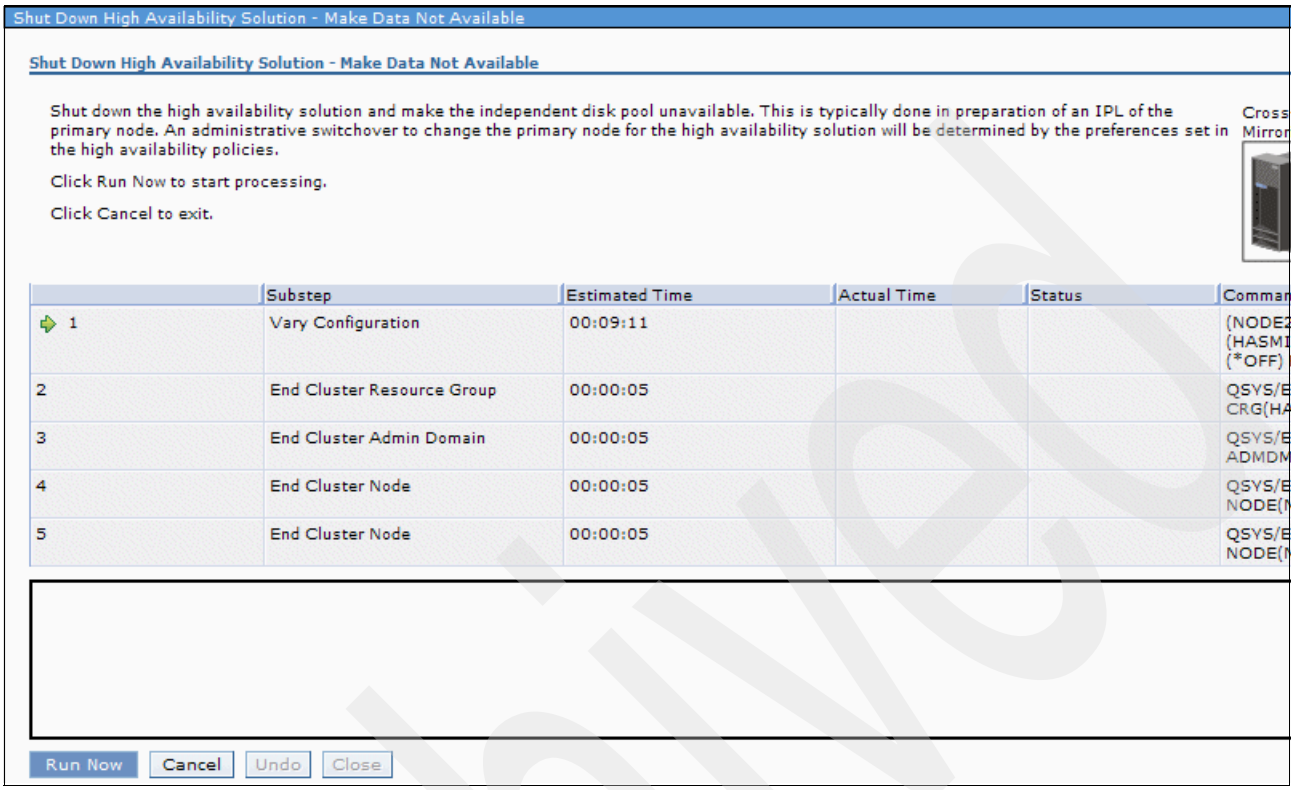


Figure 6-81 Perform shut down and make data not available

Service Info

Select the action **Service Info** from the drop-down menu and then click **Go**, as shown in Figure 6-82.

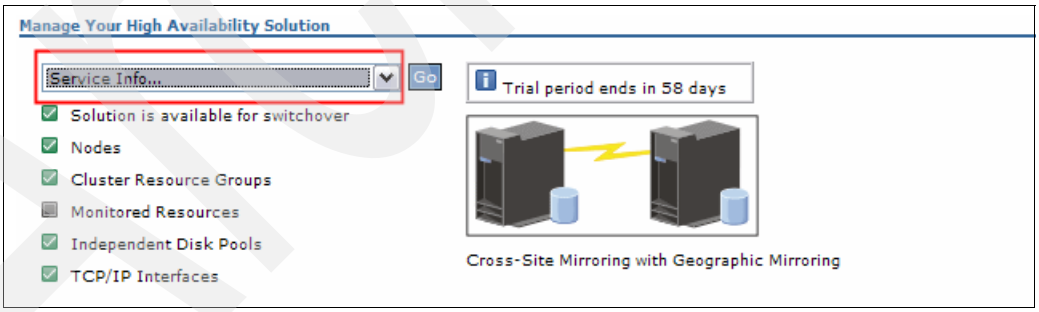


Figure 6-82 Service Info action

The next panel shows the running state of the Service Info action (Figure 6-83).

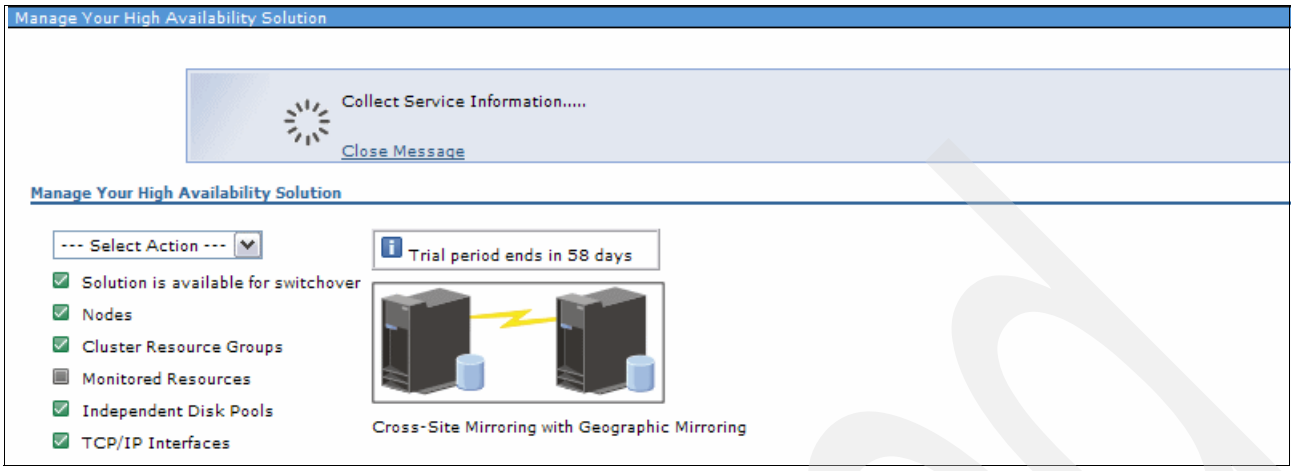


Figure 6-83 Service Info action

Recover partition

To do this:

1. Select the action **Recover Partition** and click **Go** (Figure 6-86 on page 148).

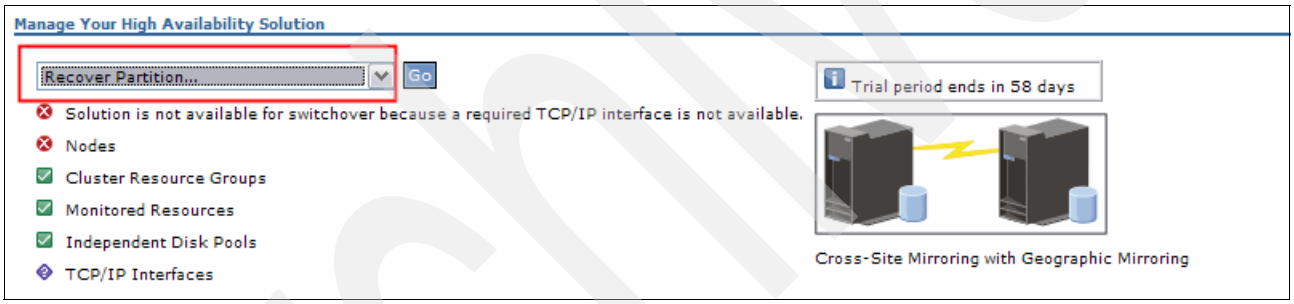


Figure 6-84 Recover Partition action

2. Click **Run Now** to perform the action, as shown in Figure 6-84.

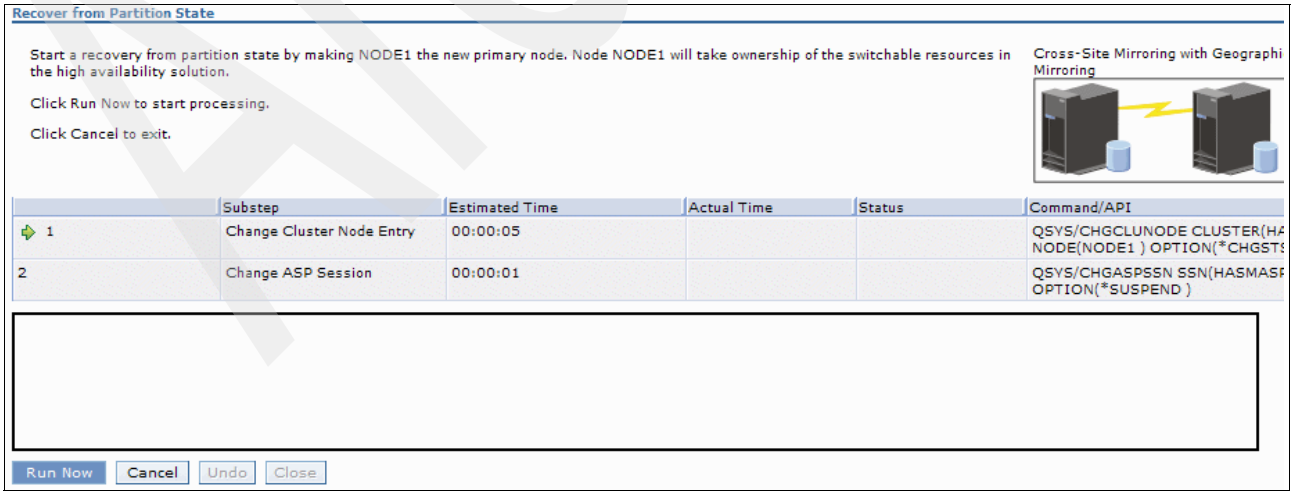


Figure 6-85 Perform recover partition

From the command line run WRKCLU and select option 6 (Work with cluster nodes). The partition status is shown in Figure 6-86.

```
                          Work with Cluster Nodes
Local node . . . . . . . . . . . . . . . . :   NODE2
Consistent information in cluster  . . . . :   Yes

Type options, press Enter.
  1=Add   2=Change   4=Remove   8=Start   9=End   20=Dump trace

                             Potential
                             Node  Mod
Opt   Node       Status       Vers Level   -------Interface Addresses-------

      NODE1      Partition       6     0   10.0.3.11         10.0.4.11
      NODE2      Active          6     0   10.0.3.12         10.0.4.12
```

Figure 6-86 WRKCLU command

3. The next panel explains the meaning of partition state and how to recover from it. Click **Continue** after you have reviewed the explanation (Figure 6-87).

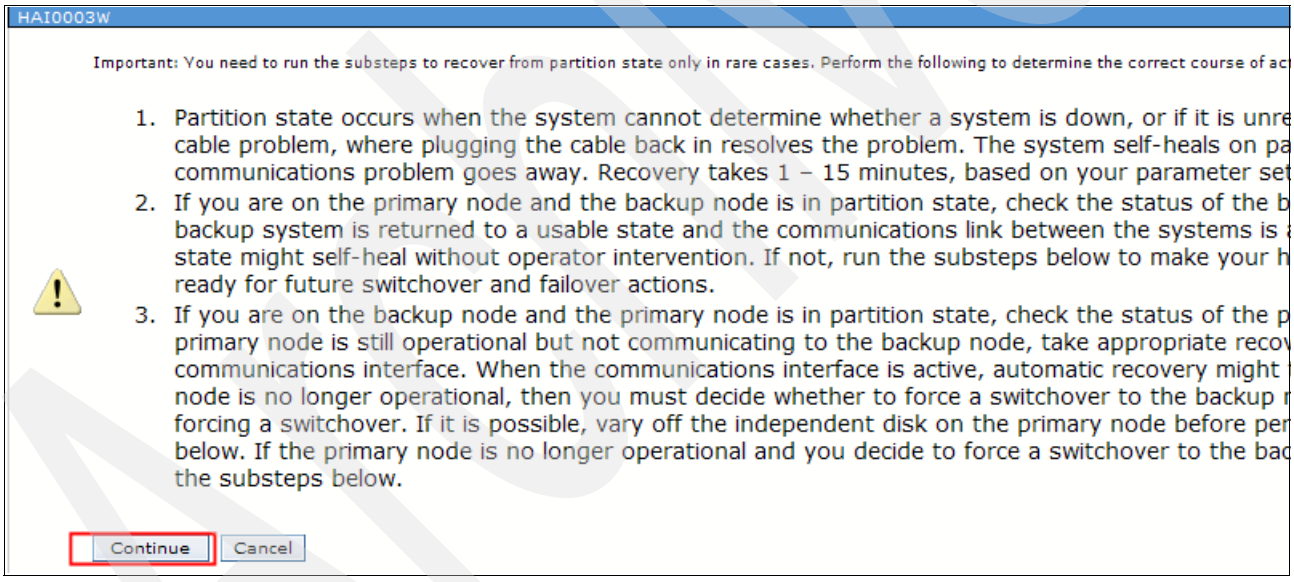


Figure 6-87 Warning message about recover partition action

4. Click **Run Now**, as shown in Figure 6-88.

   The command QSYS/CHGCLUNODE CLUSTER(HASMCLU) NODE(NODE1) OPTION(*CHGSTS) is performed on NODE2. This action changes the status of NODE1 from partition to failed, and it makes NODE2 the new primary node, as shown in Figure 6-89. NODE2 takes ownership of the switchblade resources in the high availability solution (Figure 6-88).

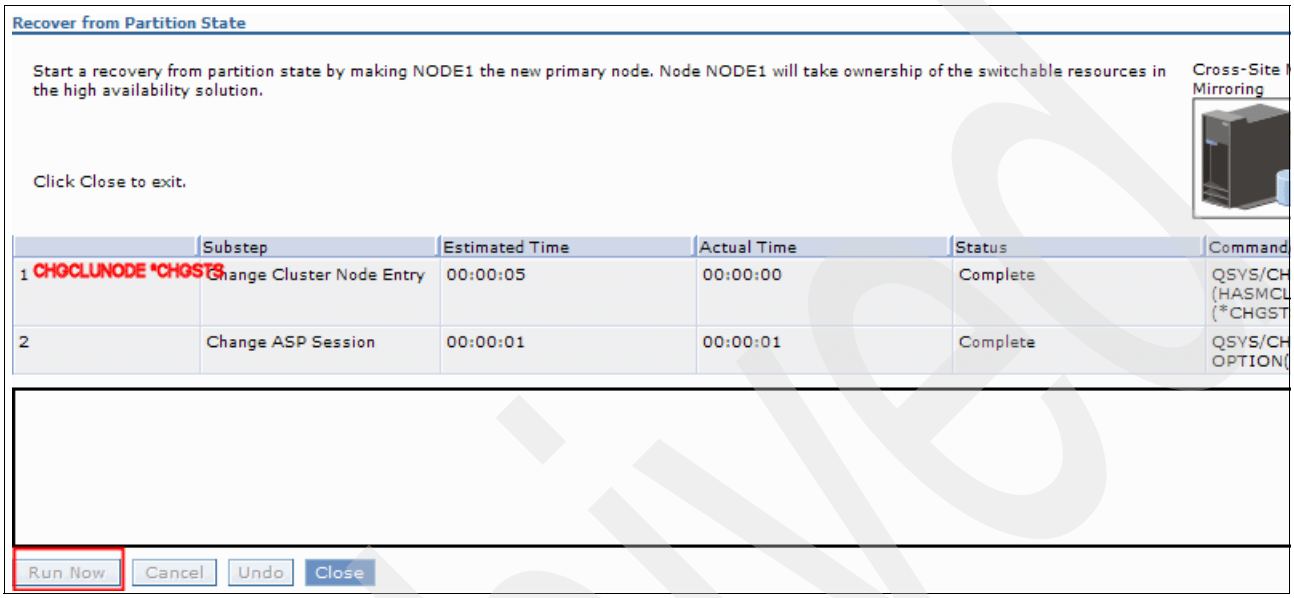


Figure 6-88 Perform recover from partition state

If we run WRKCLU and select option 6 (Work with Cluster Nodes) we will see the panel shown in Figure 6-89.

```
                          Work with Cluster Nodes

Local node . . . . . . . . . . . . . . . . :   NODE2
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
1=Add   2=Change   4=Remove   8=Start   9=End   20=Dump trace

                           Potential
                           Node  Mod
Opt   Node       Status      Vers Level   -------Interface Addresses-------

      NODE1      Failed         6     0   10.0.3.11         10.0.4.11
      NODE2      Active         6     0   10.0.3.12         10.0.4.12
```

Figure 6-89 WRKCLU

5. Click **Close** to return to the previous panel (Figure 6-90).

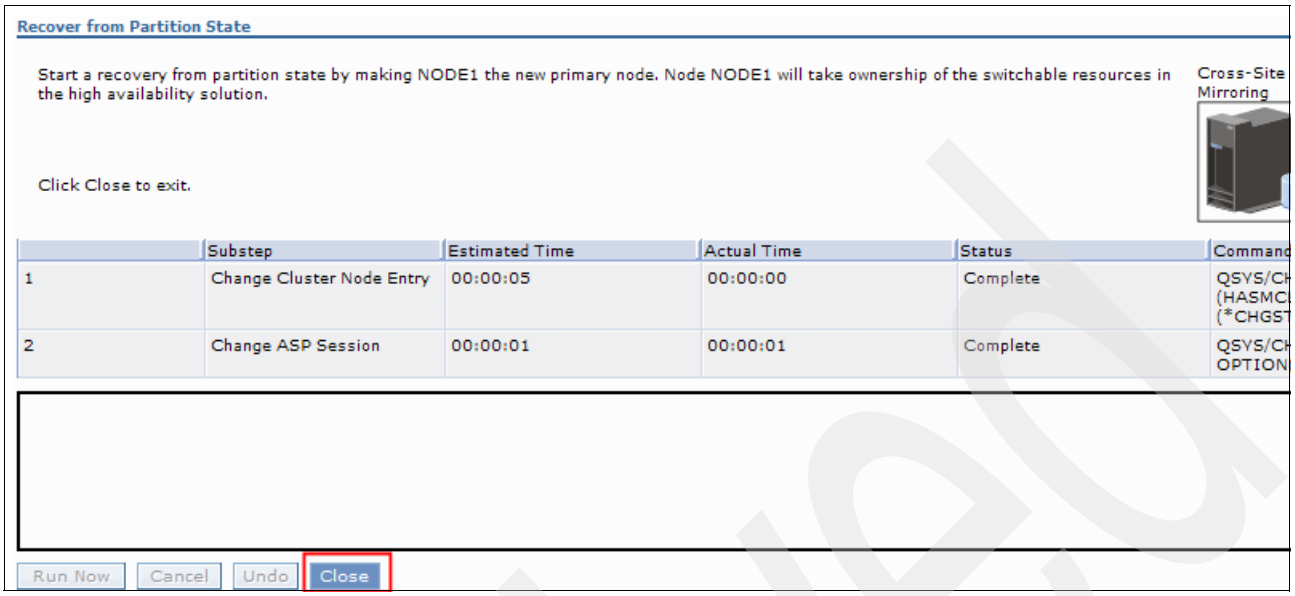


Figure 6-90 Perform recover partition action

Resume

Under the HASM GUI you can manage your high availability configuration. Be aware that here the action *resume* means that if your high availability solution is shut down this action resumes the high availability solution and makes data available. (*Resume* in a cross-site mirroring configuration has a different meaning.)

1. On the Manage yOur High Availability Solution panel select **Resume** and click **Go** to continue, as shown in Figure 6-91.

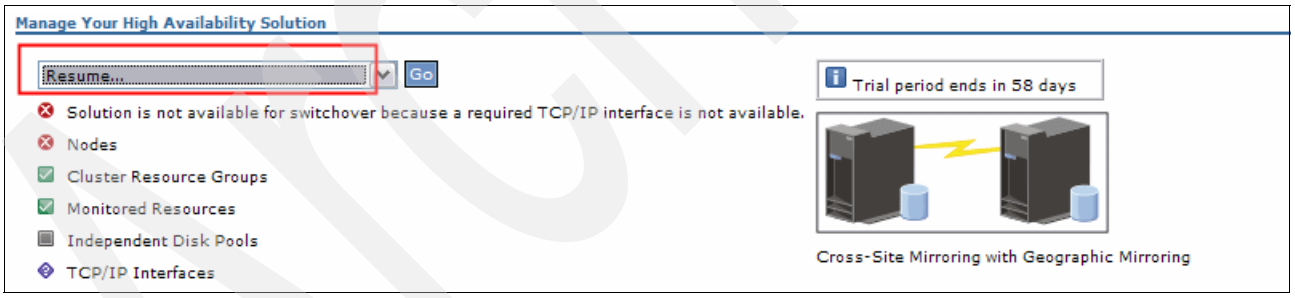


Figure 6-91 Resume action

2. Click **Run Now** to continue, as shown in Figure 6-92.

   If the high availability solution was shut down this action resumes the high availability solution and makes data available.

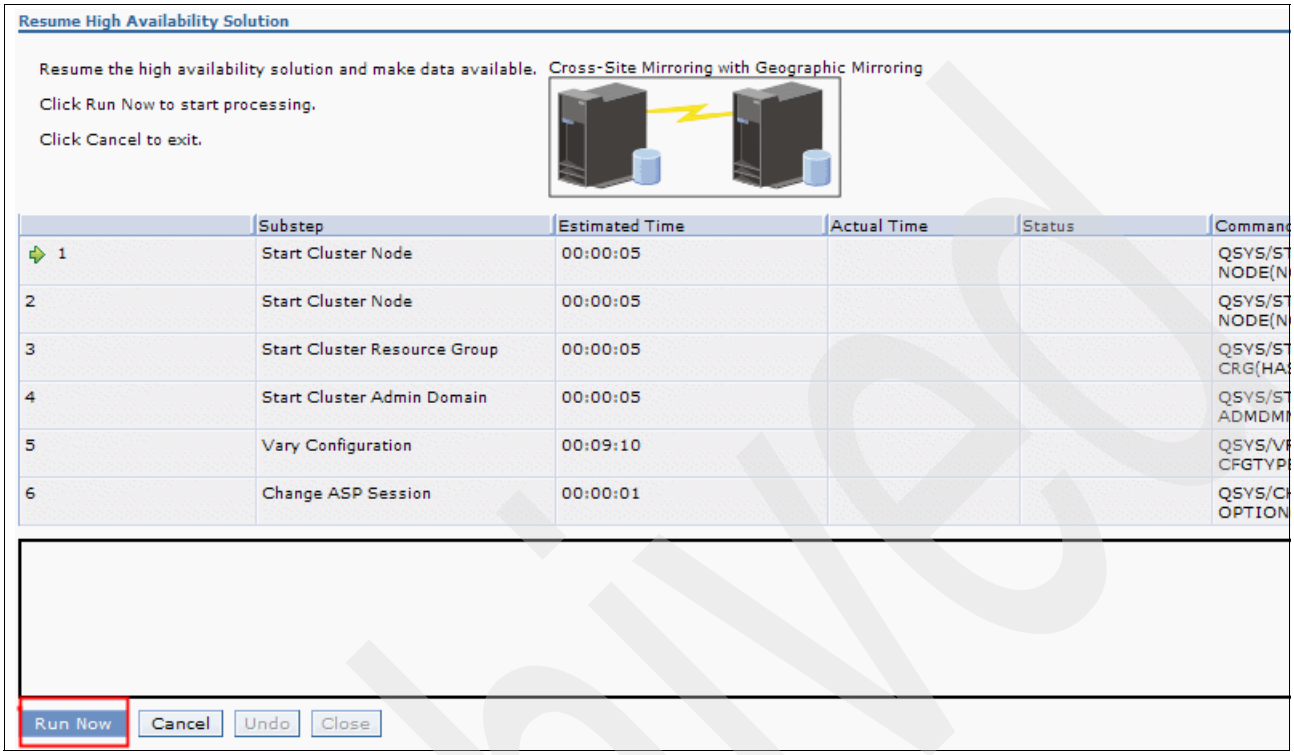


Figure 6-92 Perform resume action

If one of the substeps listed above fails the resume action stops. The resume will continue after you have solved the problem. See the text message area to find the error message, as shown in Figure 6-93.

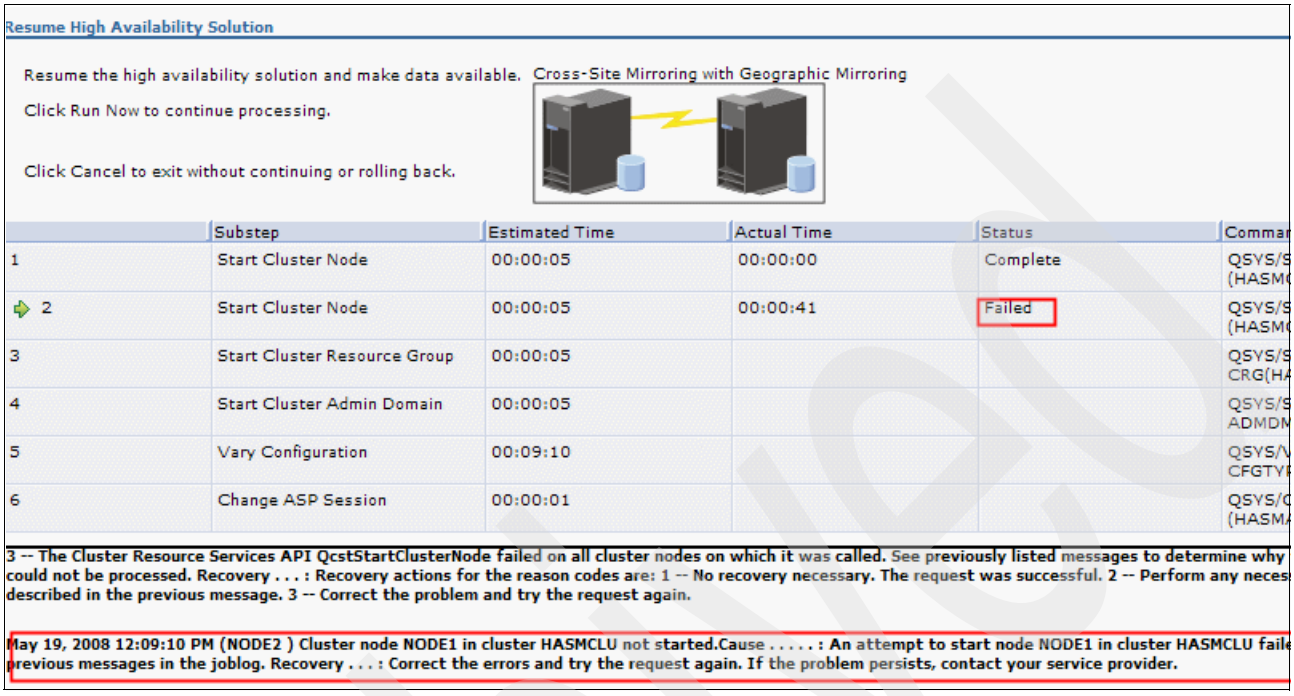


Figure 6-93 Resume failed example

3. Click **Close** to return to previous panel, as shown in Figure 6-94.

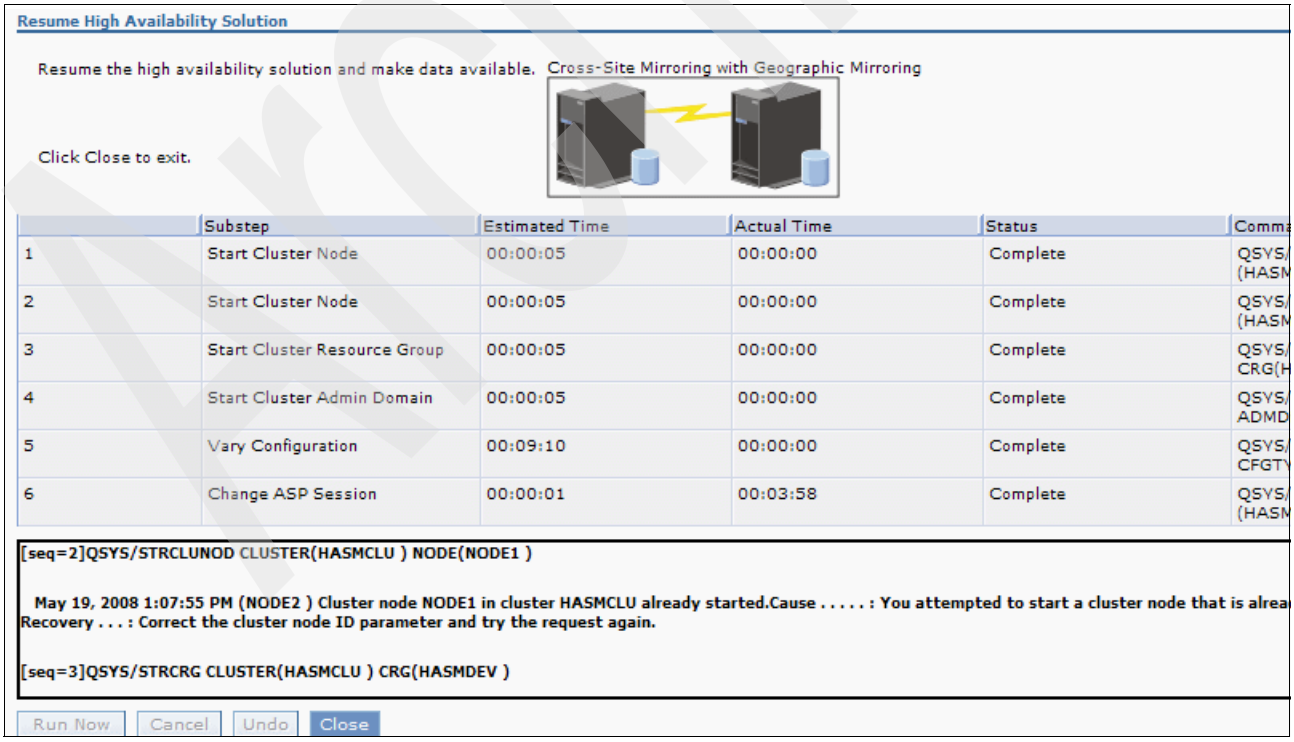


Figure 6-94 Resume completed

During the resuming action the Change ASP Session command is running, as shown in step 6 in Figure 6-95.

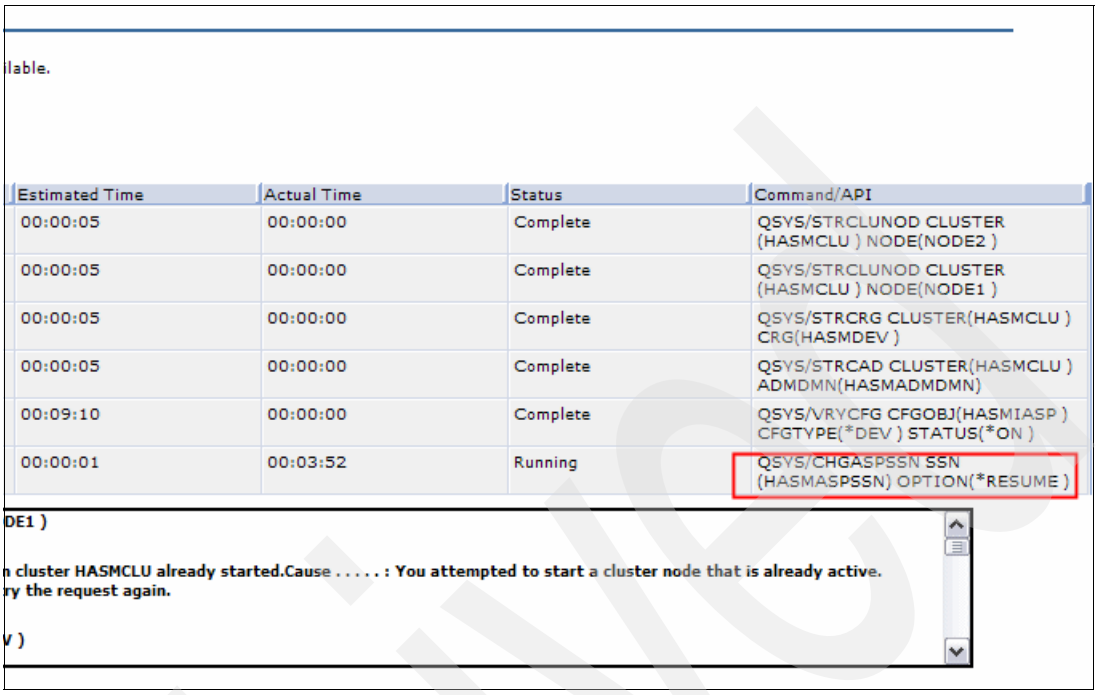


Figure 6-95 Resuming action completed

From the command line you can check the status of the resuming action running the DSPASPSSN command, as shown in Figure 6-96. The MIRROR copy of the iASP changes from RESUMING to ACTIVE.

```
                          Display ASP Session
Display ASP Session                          NODE2
                                                        05/19/08  13:22:20
   Session . . . . . . . . . . . . . . . . . :   HASMASPSSN
   Type  . . . . . . . . . . . . . . . . . . :     *GEOMIR
   Mode  . . . . . . . . . . . . . . . . . . :     SYNC
   Suspend timeout . . . . . . . . . . . . . :     120
   Synchronization priority  . . . . . . . . :     *MEDIUM
   Track space . . . . . . . . . . . . . . . :     0

                                                                    Bottom
                          Copy Descriptions

 ASP          ASP                                            Data
 Device       Copy               Role         State         State    Node
 HASMIASP     HASMRMTCPY     PRODUCTION     AVAILABLE       USABLE    NODE2
 HASMIASP     HASMLCLCPY        MIRROR        ACTIVE        USABLE    NODE1
```

Figure 6-96 DSPASPSSSN command

High availability solution resources

Under this section you can manage the following resources:

- Nodes
- Cluster resource groups
- Monitored resources
- iASP
- TCP/IP Interfaces
- Policies

Managing nodes

To manage nodes:

1. Starting from the Manage High Availability Solution Resources panel, refer to the step on page 142 to see how to get there, then select **Nodes**, and from the drop-down menu (red circle in Figure 6-97) select **Work with all nodes**.

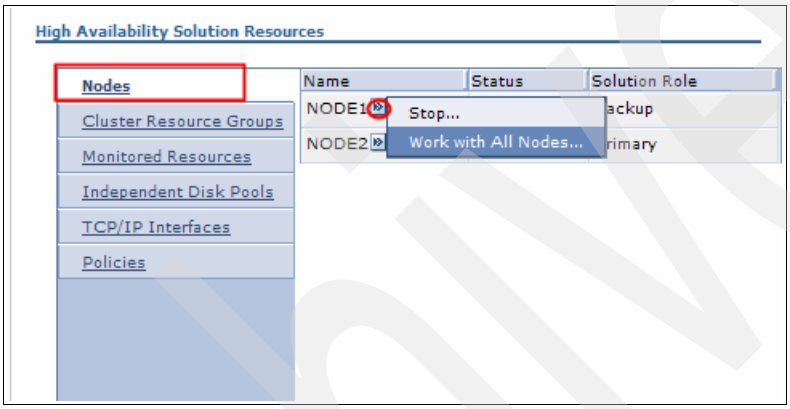


Figure 6-97 Manage node window

2. Drop down the context menu (see red circle in Figure 6-98) and click the action desired.

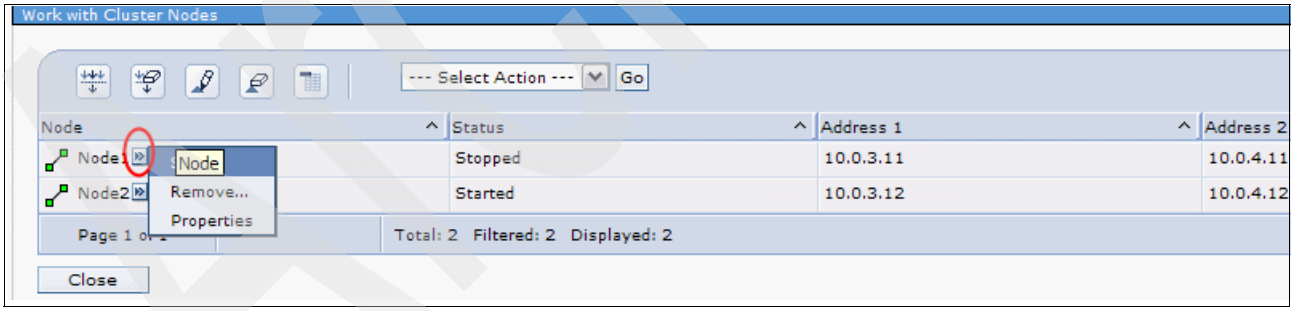


Figure 6-98 Work with node window

You can perform any of the following functions on the nodes in the high availability solution:

- Monitor the status of nodes.
- Display or edit node properties.
- Start a node.
- Stop a node.
- Work with all nodes.

Managing cluster resource group

To manage a cluster resource group:

1. Select the **Cluster Resource Groups** tab, then click the context menu and from the drop-down menu choose **Work with CRGs** (Figure 6-99).

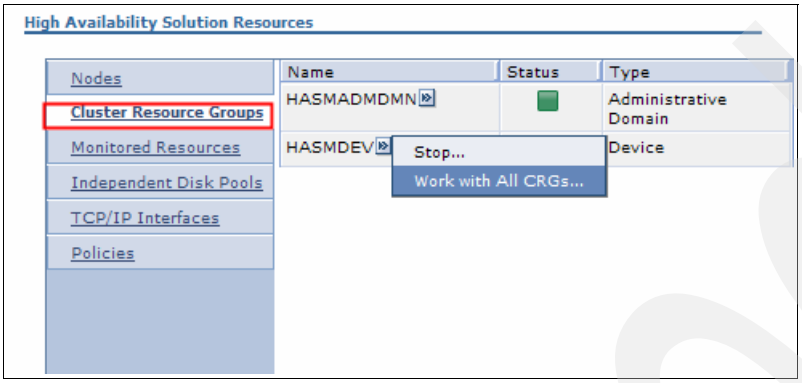


Figure 6-99 Work with cluster resource group window

2. Select the action to perform from the drop-down menu, as shown in Figure 6-100.

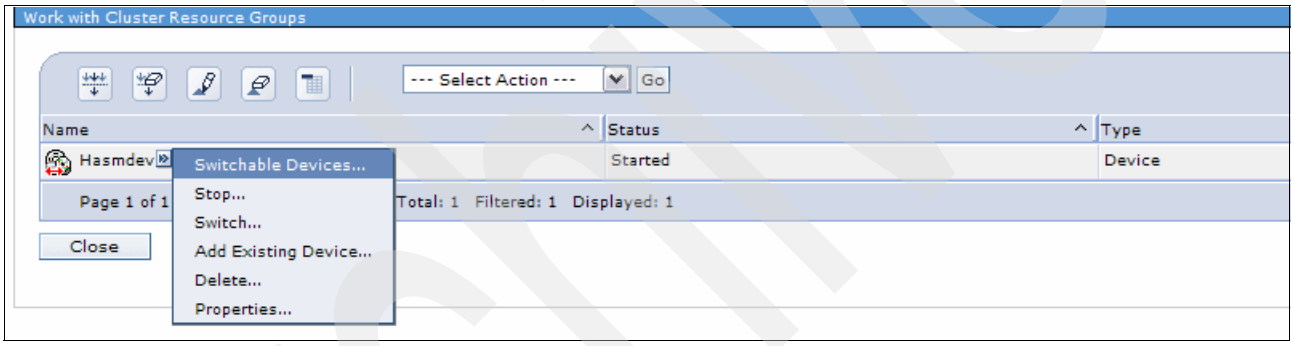


Figure 6-100 Cluster resource group action

The following functions can be performed on the CRGs:

- Monitor the status of CRGs.
- Start a CRG.
- Stop a CRG.
- Delete a CRG.
- Work with all CRGs.
- Display or edit CRG properties.

Managing monitored resources

You can manage monitored resources in your high availability solution by using the High Availability Solutions Manager graphical interface. These resources are monitored, and when they are changed on one node, those changes are propagated to other nodes in the high availability solution.

Select **Monitored Resources**, as shown in Figure 6-101.

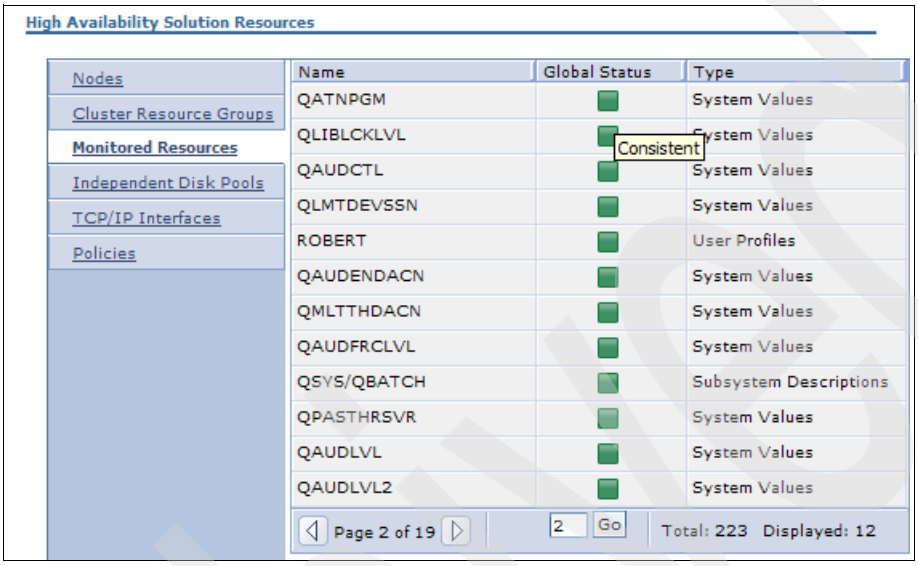


Figure 6-101 Monitored resources window

The global status values are listed in Figure 6-102.

Icon	Status	Description
	Consistent	The values for all the resource's attributes monitored by the system are the same on within the cluster administrative domain.
	Inconsistent	The values for all the resource's attributes monitored by the system are not the same within the cluster administrative domain.
	Pending	The values of the monitored attributes are in the process of being synchronized acros administrative domain.
	Added	The monitored resource entry has been added to the monitored resource directory in administrative domain but has not yet been synchronized.
	Ended	The monitored resource is in an unknown state because the cluster administrative dor ended, and changes to the resource are no longer being processed.
	Failed	The resource is no longer being monitored by the cluster administrative domain and th removed. Certain resource actions are not recommended when a resource is being syr cluster administrative domain. If the resource represented by an MRE is a system obje deleted, renamed, or moved to a different library without removing the MRE first. If a renamed or moved to a different library, the global status for the MRE is Failed and an the resource on any node after that are not propagated to any node in the cluster ad

Figure 6-102 Global status value

Managing Independent disk pools

To manage independent disk pools:

1. Select **Independent Disk Pools** and then select **Work with All Independent Disk Pools** from the context menu, as shown in Figure 6-103.

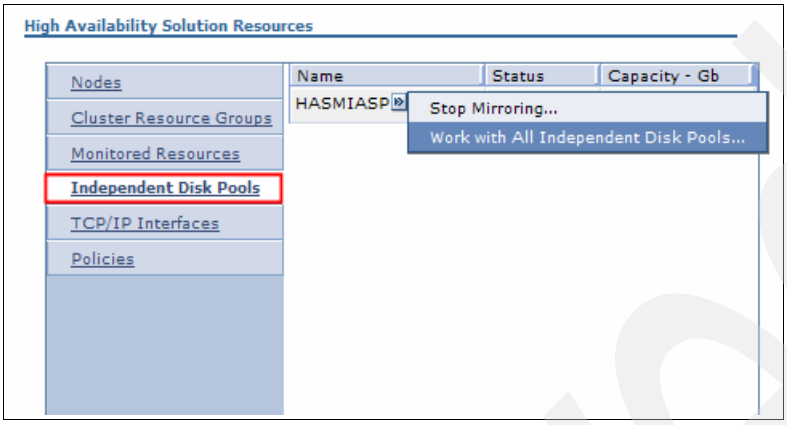


Figure 6-103 Manage independent disk pool window

2. Select the context menu to select the action to perform, as shown in Figure 6-104.

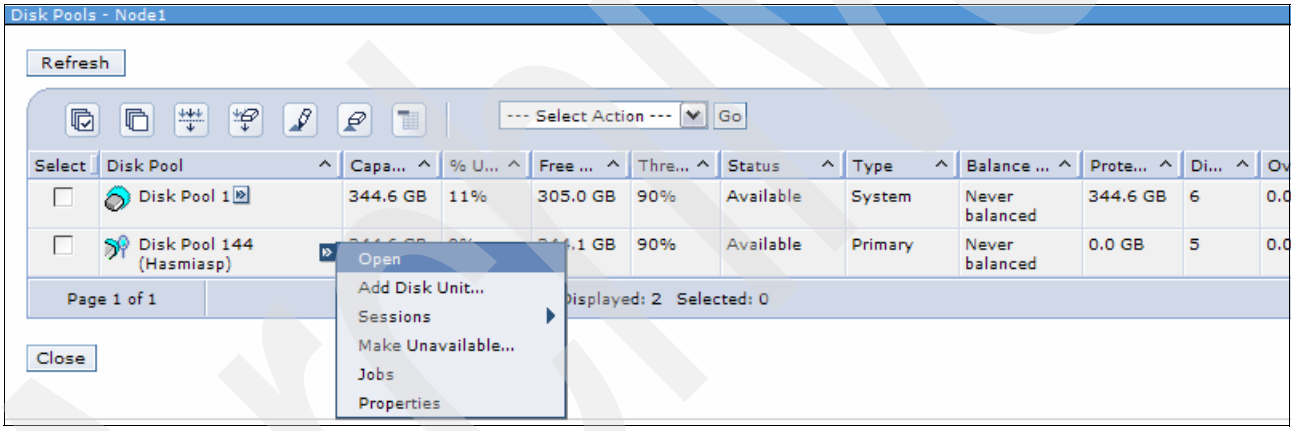


Figure 6-104 Independent disk pools action window

The following functions can be done on the Independent disk pools:

- Monitor the status of independent disk pools.
- Start mirroring.
- Stop mirroring.
- Work with all independent disk pools.
- Display or edit properties.

Managing TCP/IP interfaces

To manage TCP/IP interfaces:

1. On next window select **TCP/IP Interfaces**, then select **Work with All TCP/IP Interfaces** (Figure 6-105).

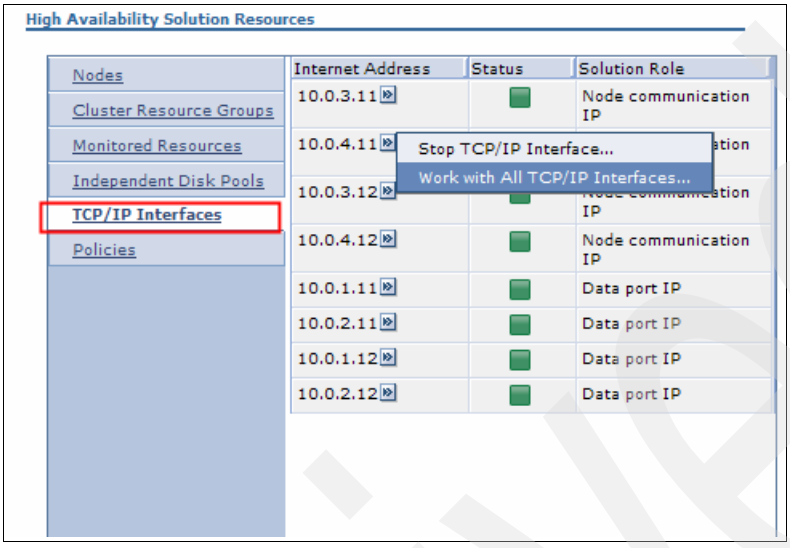


Figure 6-105 TCP/IP Interfaces window

2. Select the action to perform from the drop-down menu, as shown in Figure 6-106.

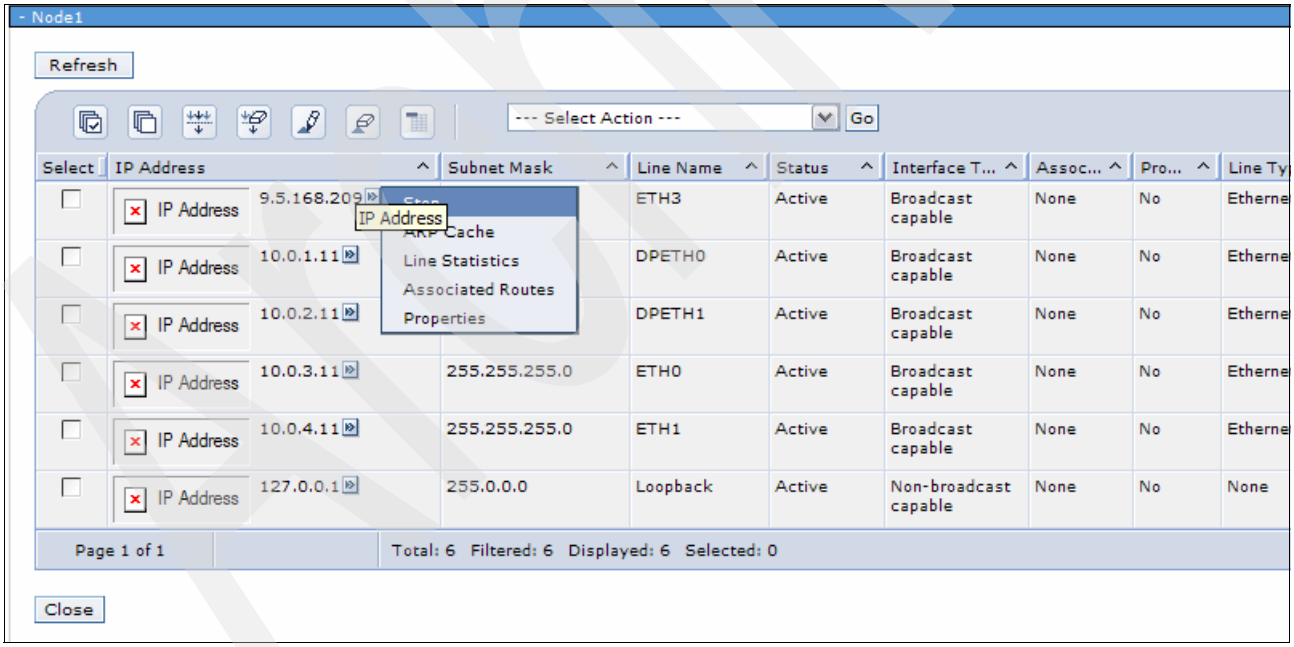


Figure 6-106 Work with TCP/IP Interfaces window

You can perform the following actions on the TCP/IP interfaces:

- Monitor the status of TCP/IP interfaces.
- Start TCP/IP interfaces.
- Stop TCP/IP interfaces.
- Work with all TCP/IP interfaces.

Managing policies

Policies define the actions that might occur within your high-availability environment. These policies were established when you set up your high availability solution. Refer to 6.6.1, “Getting started with the setup of your high availability solution” on page 105.

1. Select **Policies** on the next panel, as shown in Figure 6-107.

Figure 6-107 Manages policies window

2. Click **Edit** to change the policies associated with your high availability solution, as shown in Figure 6-108.

Figure 6-108 Edit policies window

The policies that you can manage in the high availability solution are:

- Action when a user profile is deleted
- Action to be done before the primary node enters in restricted state
- Action to be done before the primary node performs a power down

Managing event log

The event log provides an informational view of warnings and error messages logged for your high availability solution.

Each entry in the event has a date and time stamp, severity level, and description, as shown in Figure 6-109.

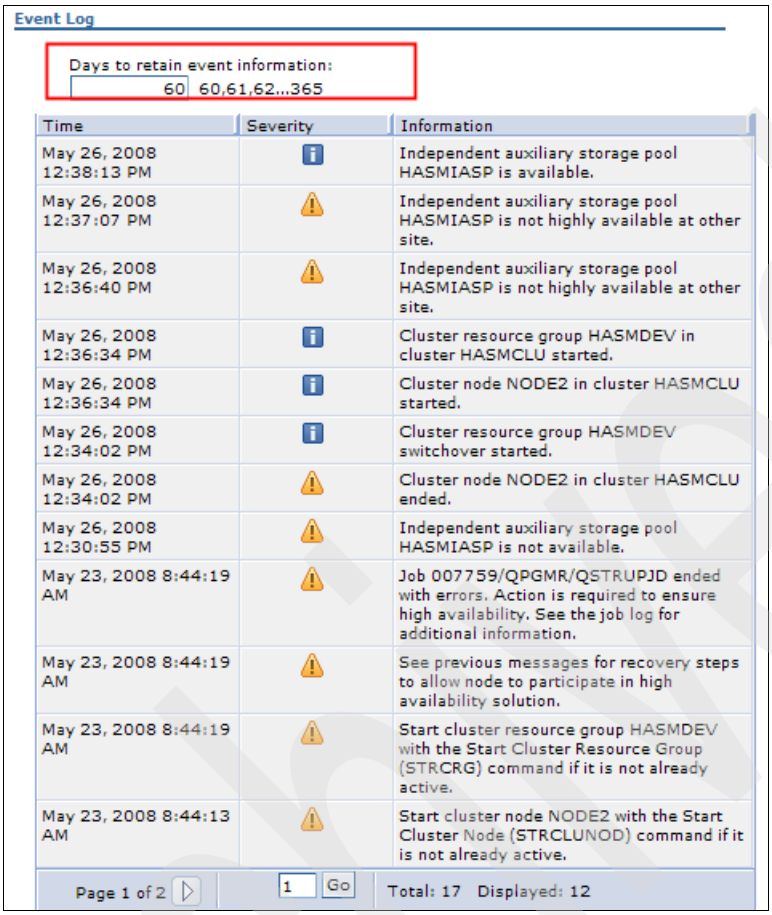


Figure 6-109 Event log window

The next window shows the event log severity level description (Figure 6-110).

Severity	Description
	The log contains informational messages.
	The log contains warning messages and should be examined.
	The log contains error messages and should be examined.

Figure 6-110 Severity level description window

Programs for starting and ending the application

Data areas exist in the QUSRHASM library that should be used and modified by the customer to specify the program that will start up and shut down their application. The format of the 20-character data areas is a 10-character program name followed by a 10-character library name. For example:

```
'STARTUP   MYLIB '
```

QSTARTAPP should contain the program that will start the application.

QSHUTDOWN should contain the program that will shut down the application.

7

Cluster Resource Services graphical user interface

In this chapter we discuss the use of the new Cluster Resource Services graphical user interface (GUI) provided by IBM System Director Navigator. This GUI provides you with a task-based approach for setting up and maintaining a high availability environment that uses PowerHA for i.

We guide you through a complete setup of an environment using metro mirror with independent auxiliary storage pools (iASPs) as an example on how this GUI works. We also provide you with information about how to set up a FlashCopy environment with the task-based GUI.

In addition to these step-by-step instructions we point you to other tasks that can be achieved using this new GUI.

7.1 Cluster GUI history

IBM System i (iSeries) cluster infrastructure came out with OS/400 at V4R4M0. At first we had no GUI access, but we could use supplied application programming interfaces (APIs).

In OS/400 V5R1M0 we had big changes with the cluster GUI being created. Since then we have made enhancements to GUI, but there have been no major changes.

However, with IBM i 6.1 we see the introduction of a brand new GUI for IBM i clustering management and disk management provided by IBM System Director Navigator.

7.2 Setting up an environment using metro mirror

In the following sections we walk you through a complete setup for an environment using metro mirror together with an iASP in a cluster environment. For more information about metro mirror refer to 4.5, “Metro mirror” on page 58. We assume that DSCLI is already installed on both nodes. Information about how to do this can be found in “Installing DSCLI” on page 87.

7.2.1 Create an iASP on the production node

Create an iASP on the Storage system attached to the production partition by using IBM Systems Director Navigator for i5/OS:

1. Once in the System Director Navigator, select **Configuration and Services** in the Navigation tree, as shown on Figure 7-1.

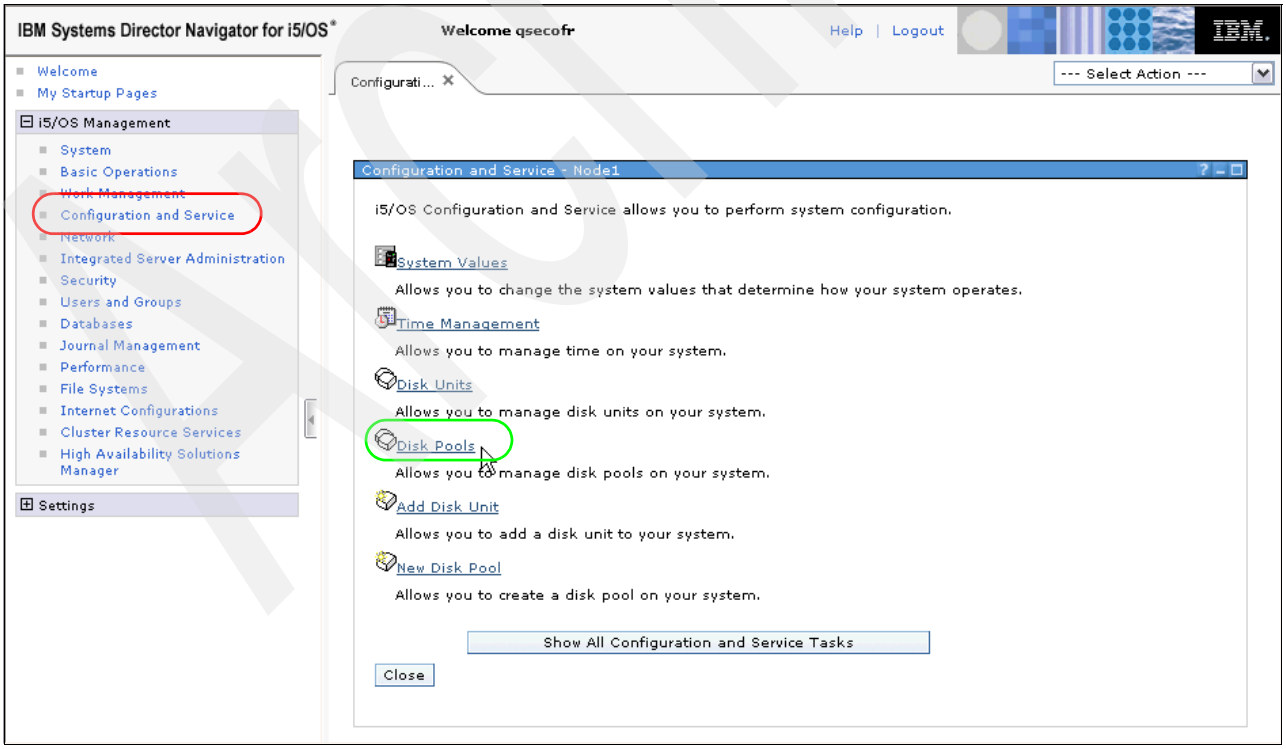


Figure 7-1 Systems Director Navigator: Launch configuration and services

2. On the Configuration panel, which is shown next, select **Disk Pools**. On the Disk Pools panel select **New Disk Pool** from the pull-down, as shown in Figure 7-2. Click **Go**. This brings up the New Disk Pool wizard.

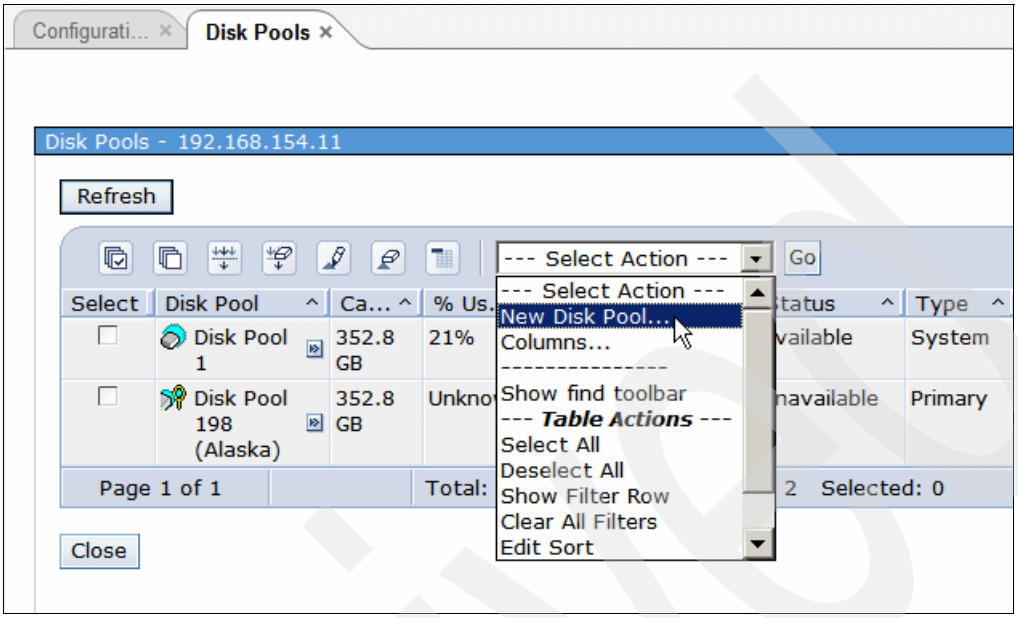


Figure 7-2 New Disk Pool

3. On the first wizard panel click **Next**. On the next wizard panel choose **Primary** from the Type of disk pool pull-down, as shown in Figure 7-3. Click **Go**.

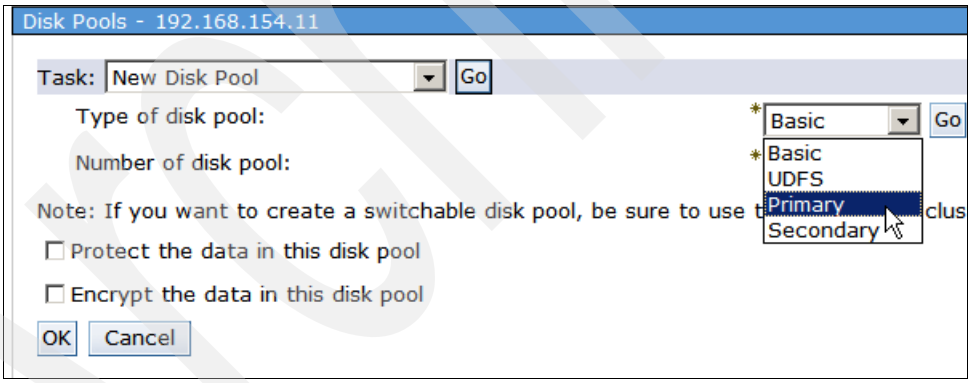


Figure 7-3 Primary disk pool

4. On the next wizard panel, specify the name of the new disk pool and click **OK**. On the next panel select the disk pool to the disk units that will be added. Select the disk pool that you are creating and click **Next**. The wizard now asks you to select which type of protection to use for the new disk pool, mirrored or parity-protected. Select the relevant option.

5. The list of available disk units is now shown to you. Since the Storage system is configured and attached to production node you see the available LUNs to add to the iASPs. Select the LUNs that you want to include in the iASP and click **Add**. In our example we add two protected LUNs of the size 35 GB, as shown in Figure 7-4.

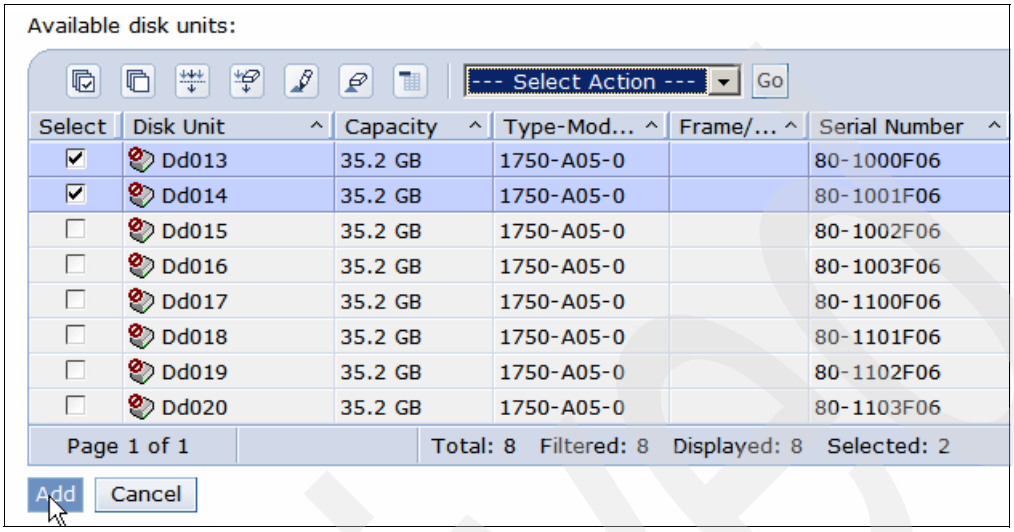


Figure 7-4 Add disk units to iASP

Note: The model of a LUN denotes whether the LUN is defined in the Storage system as parity protected or unprotected. For instance, model A05 means that the volume is defined as parity protected.

6. On the next panel you see the summary of the new disk pool. Check the disk units and characteristics of the pool and click **Finish**. The Summary panel is shown in Figure 7-5.

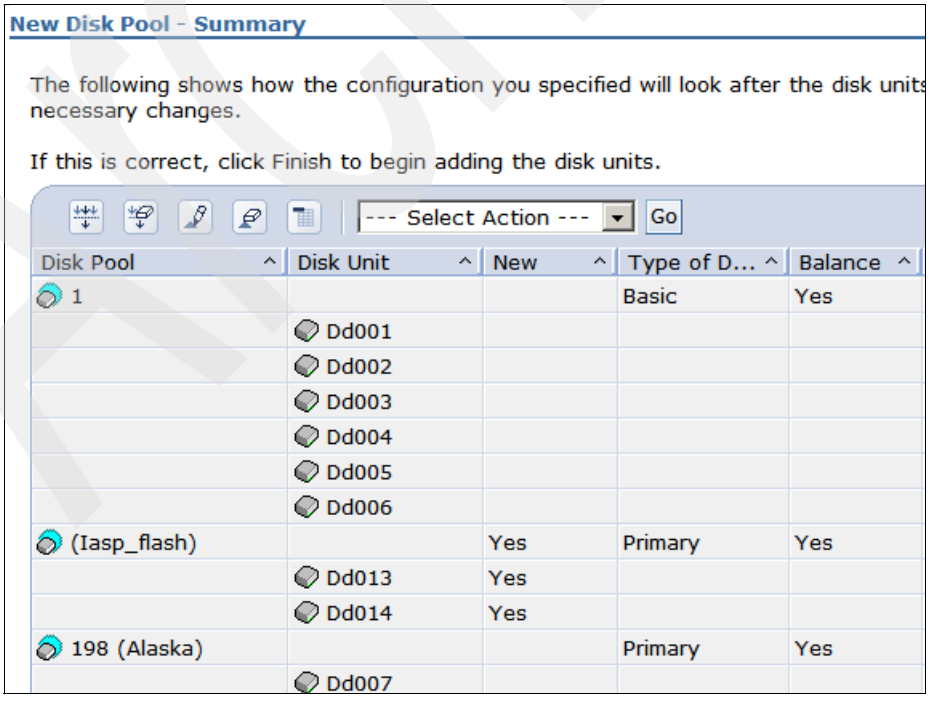


Figure 7-5 Summary of new disk pool

You can now observe the progress of creating the iASP (Figure 7-6).

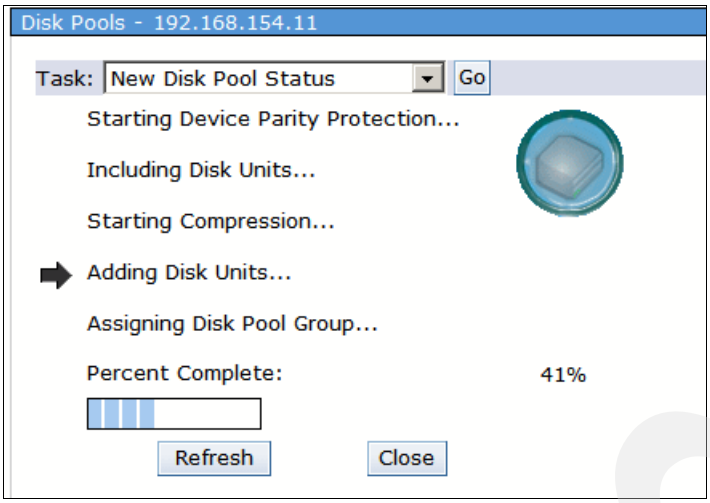


Figure 7-6 Creating a disk pool

7.2.2 Set up metro mirror on the external Storage system

Once your iASP is created you now have to set up the metro mirror configuration on your IBM Storage system. Further information about how to do this can be found in IBM *System Storage Copy Services and IBM i - A Guide to Planning and Implementation,* SG24-7103. Make sure that you also configure the reverse direction of the metro mirror connection. This is needed after a failover.

7.2.3 Preparing the scenario

Perform the steps described in this section to prepare for using metro mirror on an iASP. Once these steps are complete you will not need to do them again.

Create cluster

Create a cluster containing the production and backup partitions. Perform the following steps in IBM Systems Director Navigator for i5/OS:

1. Select **Cluster Resource Services** in the navigation tree, as shown in Figure 7-7.

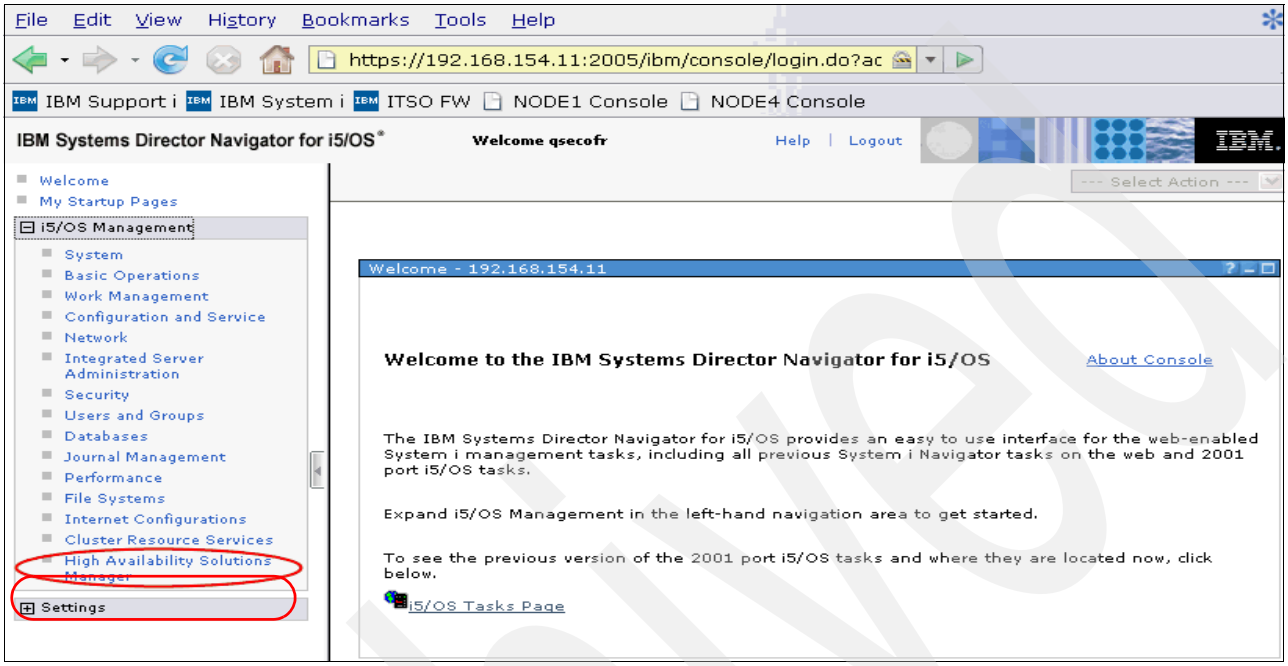


Figure 7-7 Systems Director: Cluster resource services

2. On the Cluster Resource Services panel select **New Cluster** (Figure 7-8). This starts the New Cluster wizard. The first wizard panel presents you with initial information and requirements for the cluster. After checking them, click **Next**. On the next panel insert the name of the cluster, as shown in Figure 7-8, and click **Next**.

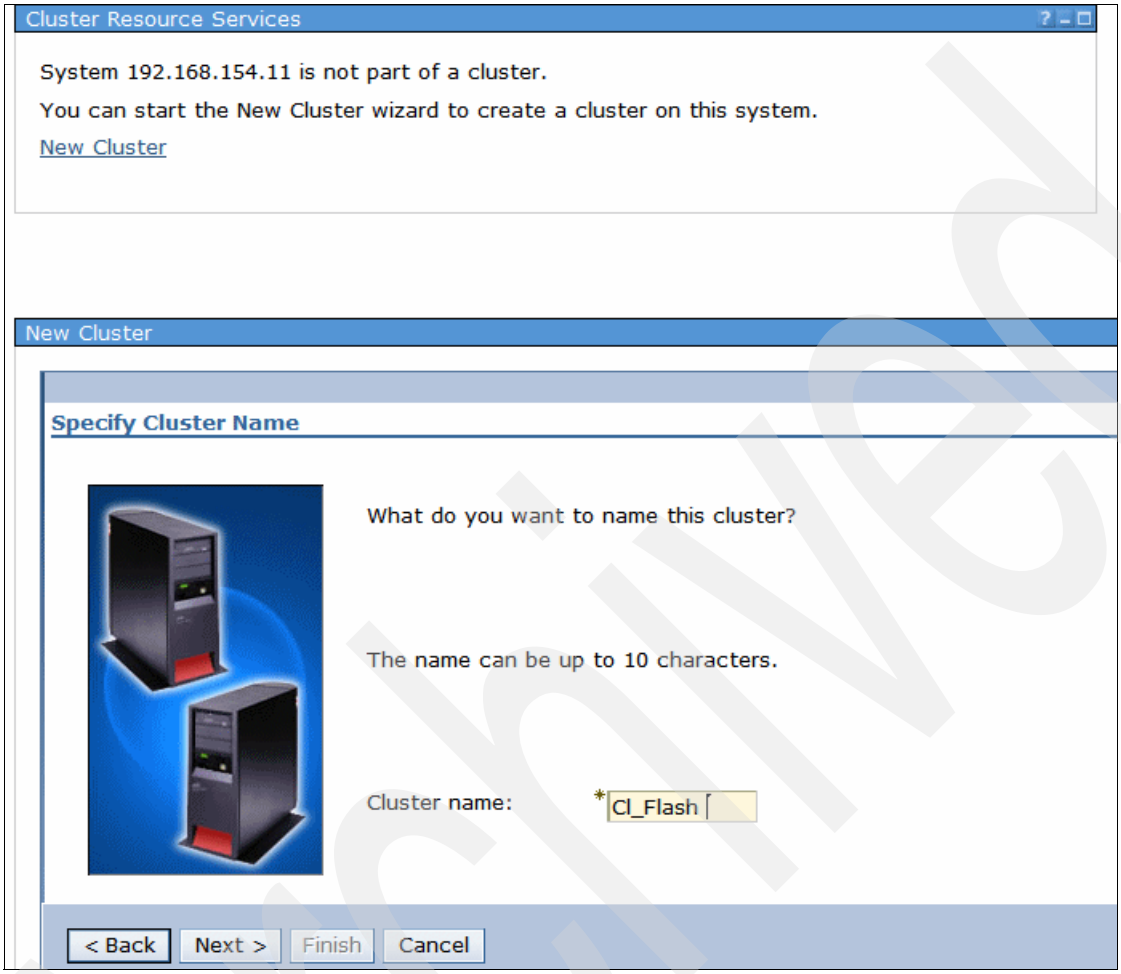


Figure 7-8 Insert the name of the cluster

3. You are now asked to insert the IP address and name of the first node in the cluster. The IP address that you enter here is used for cluster heartbeating. Note that the IP address and system name of the partition on which you are creating the cluster are already filled in. You may change them or add the second IP interface, or just click **Next** (Figure 7-9).

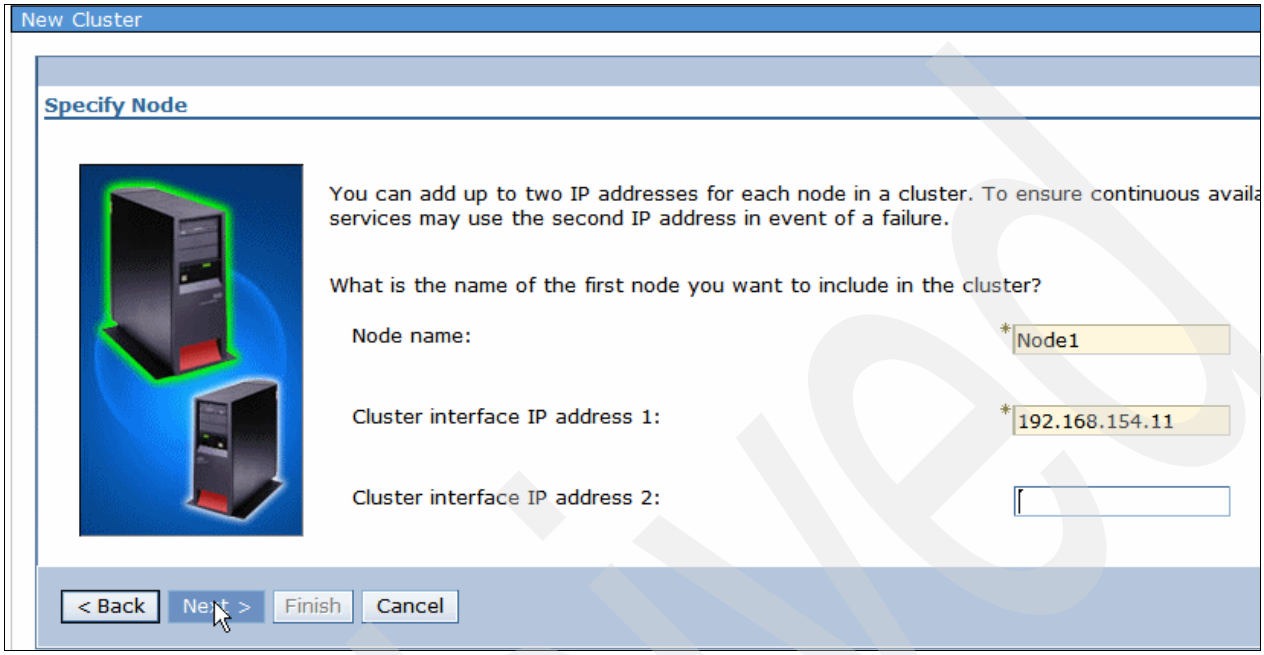


Figure 7-9 Insert IP address of first node

4. On the wizard panel that follows, insert the name and IP address of the second node (Backup partition), as shown in Figure 7-10. Click **Next**.

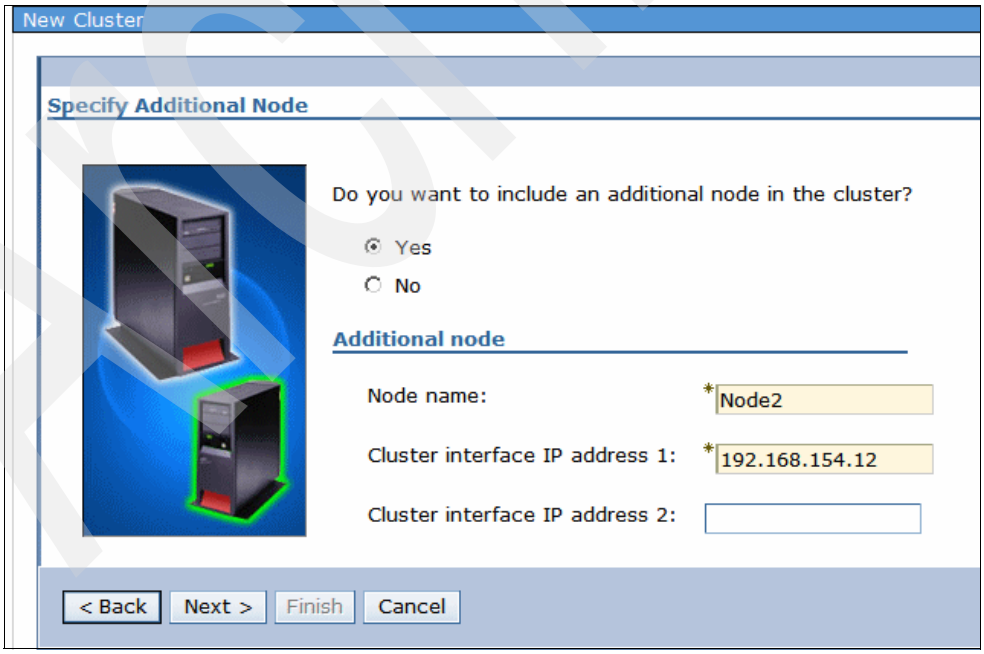


Figure 7-10 Insert second node

5. On the next wizard panels insert the version of the cluster, and, if you want, the name of the message queue to receive clustering messages. Then you see an option to select

switchable software that is automatically started with the cluster. You may select one, or just click **Next**.

6. The wizard now shows you the summary of the new cluster, as shown in Figure 7-11. After checking the summary click **Finish**.

> **Important:** In a high availability perspective, we recommend defining two cluster interface IP addresses. IBM i clustering functionality uses them for communication between cluster nodes. If one fails, the cluster can still work properly. With only one IP address, the cluster will become partitioned if there is a failure.

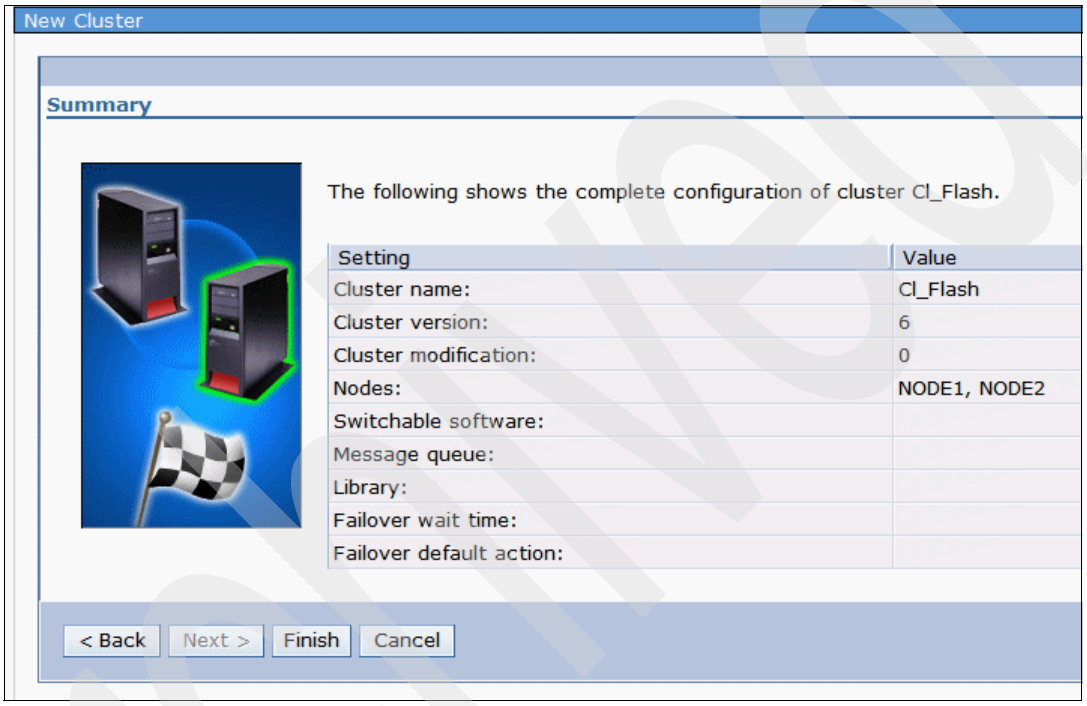


Figure 7-11 Cluster summary

7. Now the cluster is created and both nodes are started. To observe this, open the Cluster resources tab and select **Work with cluster nodes**. On the Nodes tab you can see the status and IP address of each node in the cluster. To stop the node, remove the node, or see node properties, select the appropriate action from the pull-down at the node. See Figure 7-12.

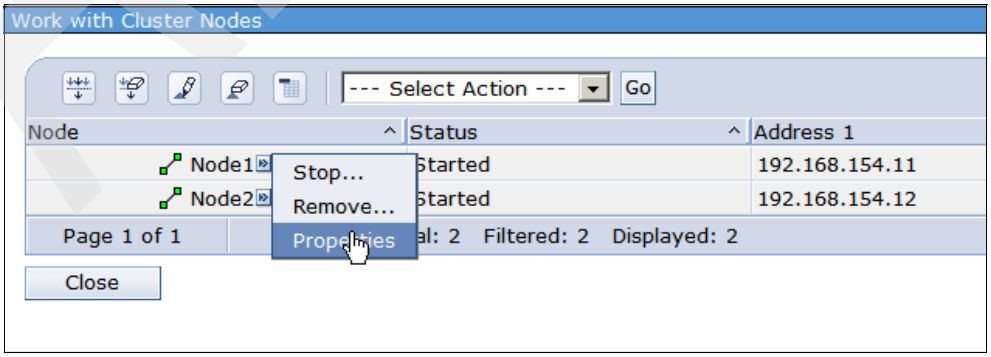


Figure 7-12 Work with cluster nodes

Creating a device domain

Create a device domain and include the cluster nodes by doing the following steps:

1. Still in the System Director tab of cluster resources, select **Work with cluster nodes**. On the Work with Cluster Nodes panel expand the pop-up window at the first node and select **Properties**, as shown in Figure 7-13.

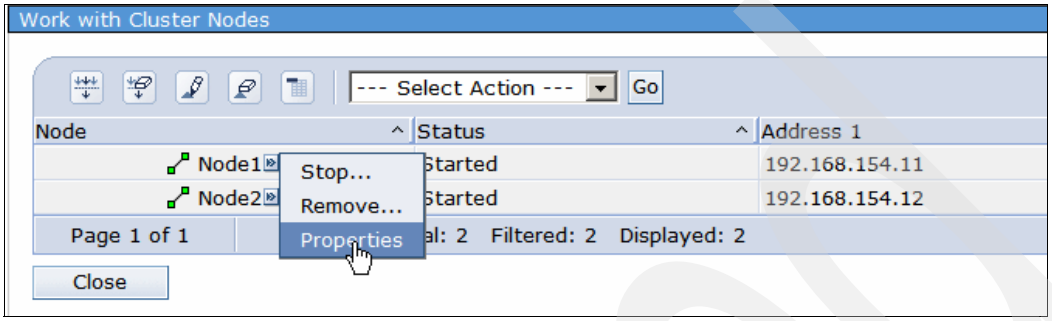


Figure 7-13 Node properties

2. On the Properties → name → Cluster Nodes panel, which is shown next, select **Clustering**. This brings up the window shown in Figure 7-14. Insert the name of device domain and click **OK**.

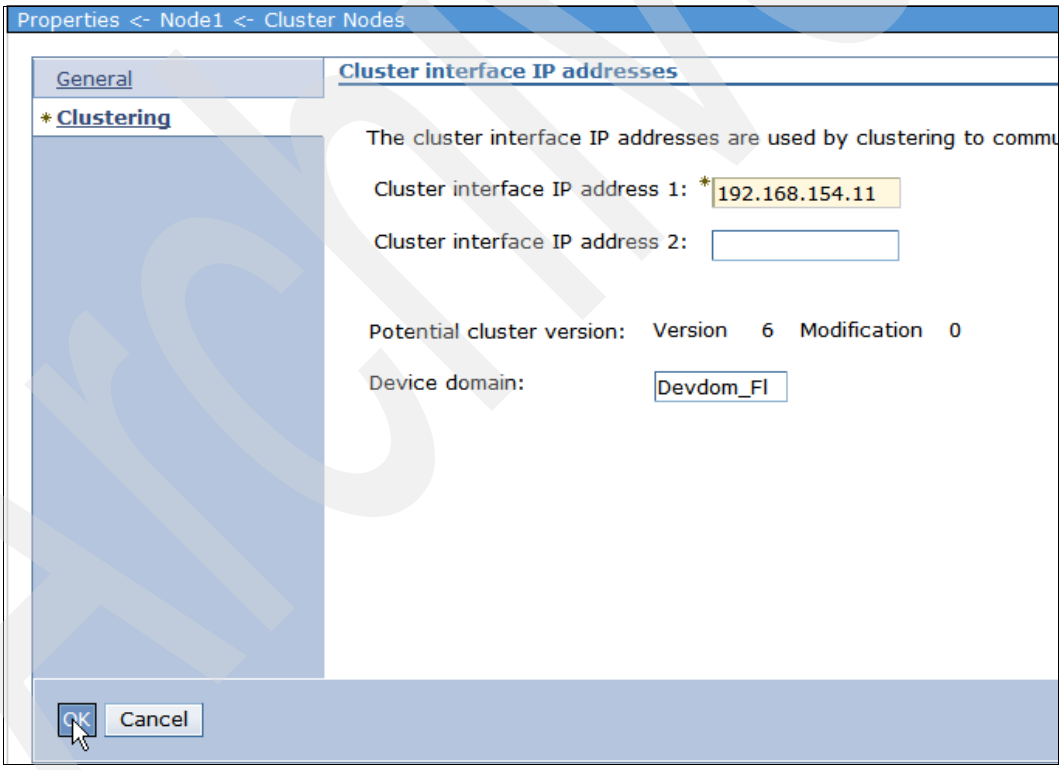


Figure 7-14 Device domain name

3. Now the device domain is created and node 1 is added to the new device domain. To add node 2 to the same device domain, perform the same steps for node 2 that you used to add node 1. That is, expand the pop-up at node 2 in the Work with Cluster Nodes window and select **Properties**. On the Properties → name → Cluster Nodes panel select **Clustering** and insert the name of the device domain that you created. Click **OK**.

4. You can check that both nodes are added to the device domain by selecting **Display device domains** on the Cluster resources tab and selecting the device domain that you created and add nodes to it. Both nodes should be shown at the device domain, as shown in Figure 7-15.

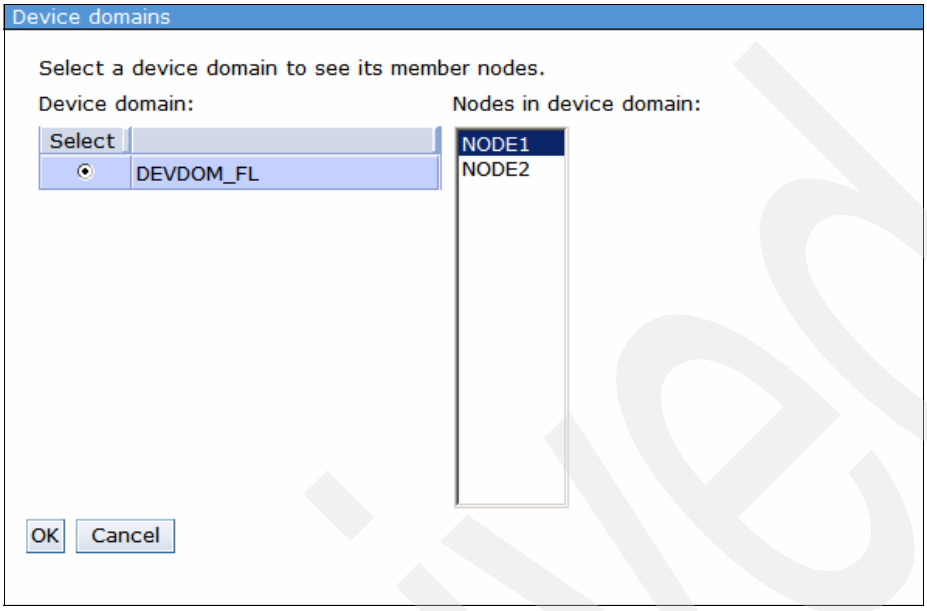


Figure 7-15 Device domain with both nodes

Create device description of iASP on backup node

On the backup node you must manually create the device description of the iASP with the same device description name and resource name as are used for iASP on the production node. To do this:

1. Open the IBM i session on the backup node
2. Type in the command `CRTDEVASP`.
3. Press F4.
4. Insert the device description name and resource name, and press Enter. If you want, type in the description of iASP, and press Enter again.

An example of creating a device description is shown in Figure 7-66 on page 197.

```
                    Create Device Desc (ASP) (CRTDEVASP)

Type choices, press Enter.

Device description . . . . . > IASPMETRO     Name
Resource name  . . . . . > IASPMETRO     Name
Relational database  . . . . . .   *GEN
Message queue  . . . . . .   *SYSOPR       Name
  Library  . . . . . .                     Name, *LIBL, *CURLIB
Text 'description' . . . . . .   Decsription of IASP for Metro Mirror

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

Figure 7-16 Create device description of an iASP

Note that at this time only the device description of the iASP is needed on the backup node, while the disk pool itself does not need to be created.

Create Device CRG

In the next step you must create a device CRG that controls switchover and failover for your metro mirror iASP. To do so:

1. In IBM System Director Navigator expand **i5/OS Management**, select **Cluster Resource Services**, and then on the right-hand pane select **Work with Cluster Resource Groups**. In the next window, under Select Action choose **New Device CRG**, as shown in Figure 7-17, and press **Go**.

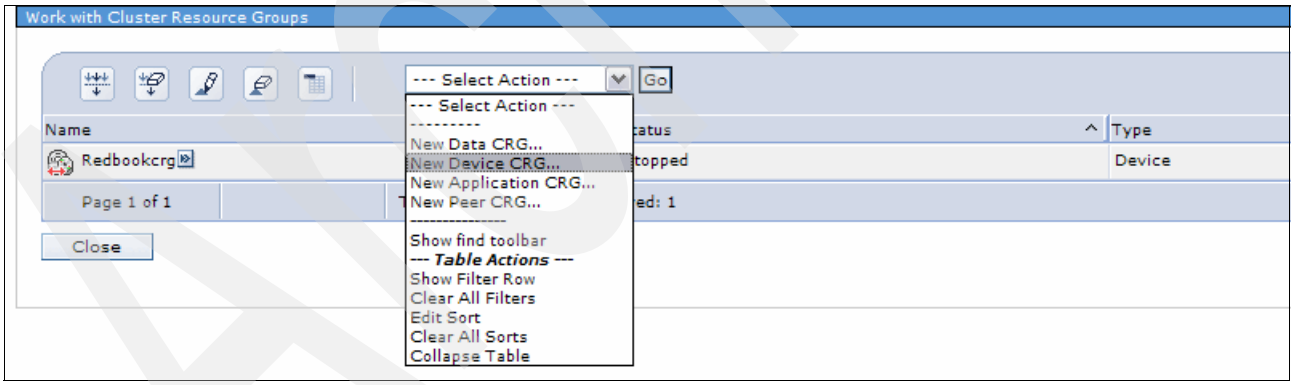


Figure 7-17 Create new Device CRG

2. You are presented with a welcome panel. After pressing **Next** on that welcome panel specify the name of the new device CRG, as shown in Figure 7-18. You can also add a description. If you want to make use of a CRG exit program, check the according box.

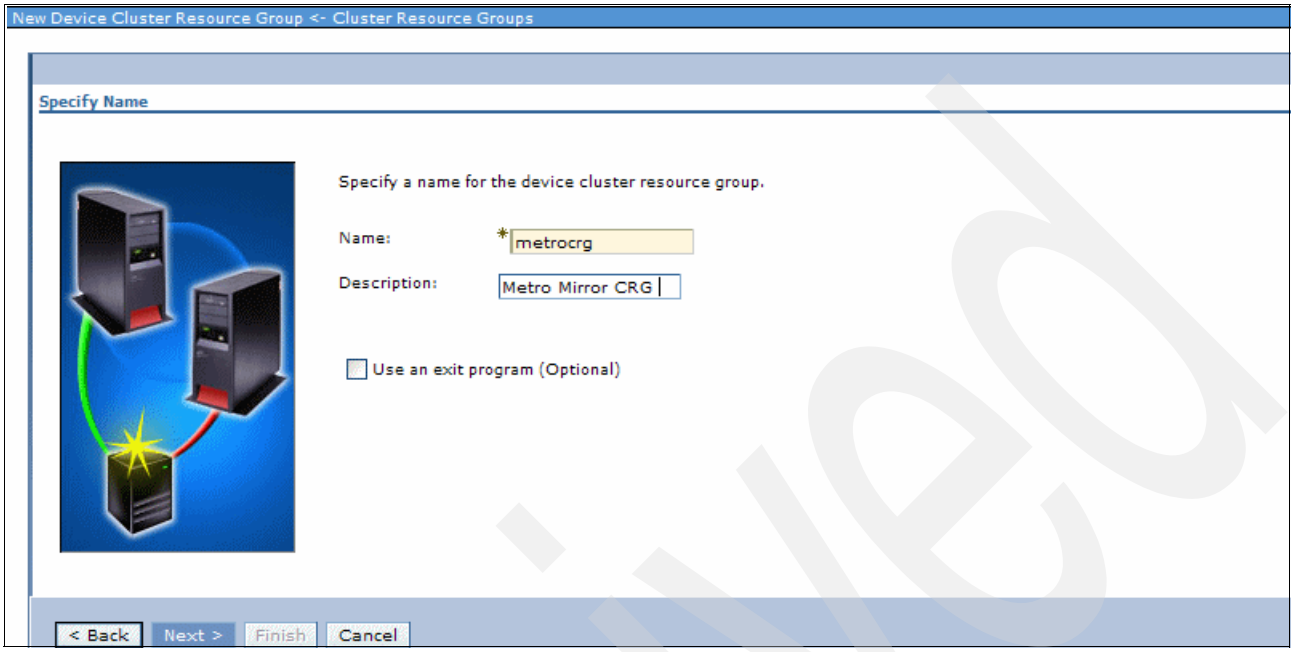


Figure 7-18 Name for new Device CRG

3. On the next panel choose your primary node. In our setup, this is node3, as shown in Figure 7-19. Make sure to also check the box for **Specify site name**, as this is required for a node that is part of a metro mirror environment.

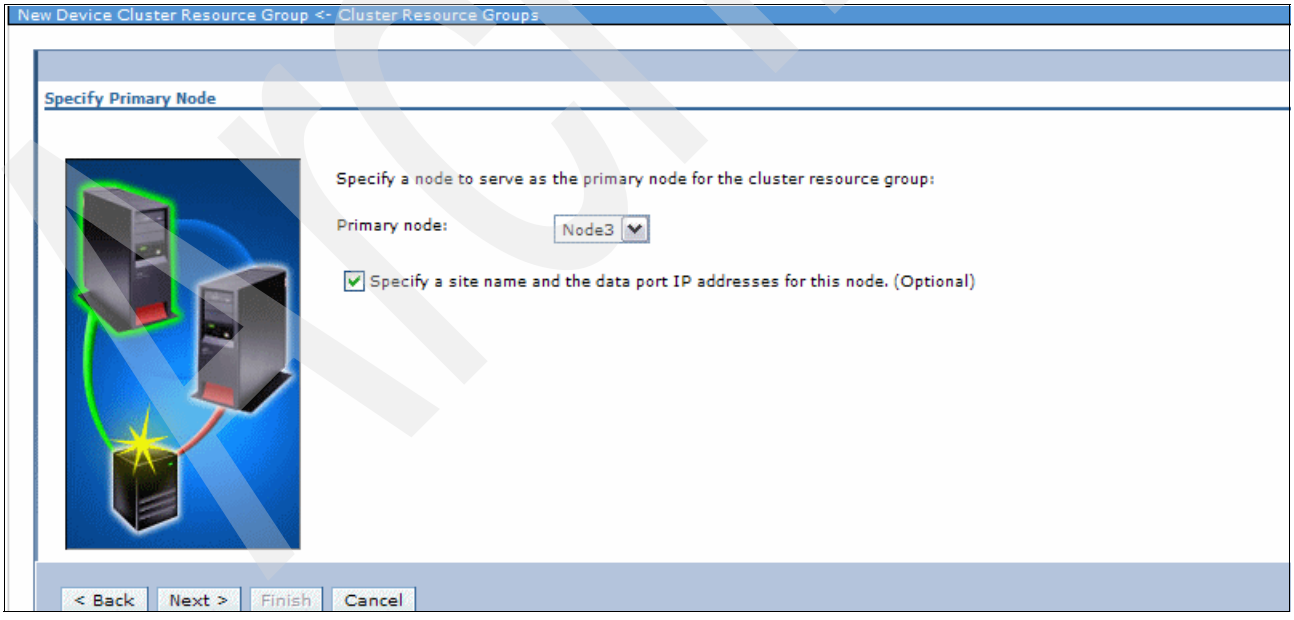


Figure 7-19 Specify Primary Node

4. Specify the site name as shown in Figure 7-20. Do not specify a data port IP address as you would do in a geographic mirroring environment. Metro mirror works between the Storage systems, so data ports are not used here.

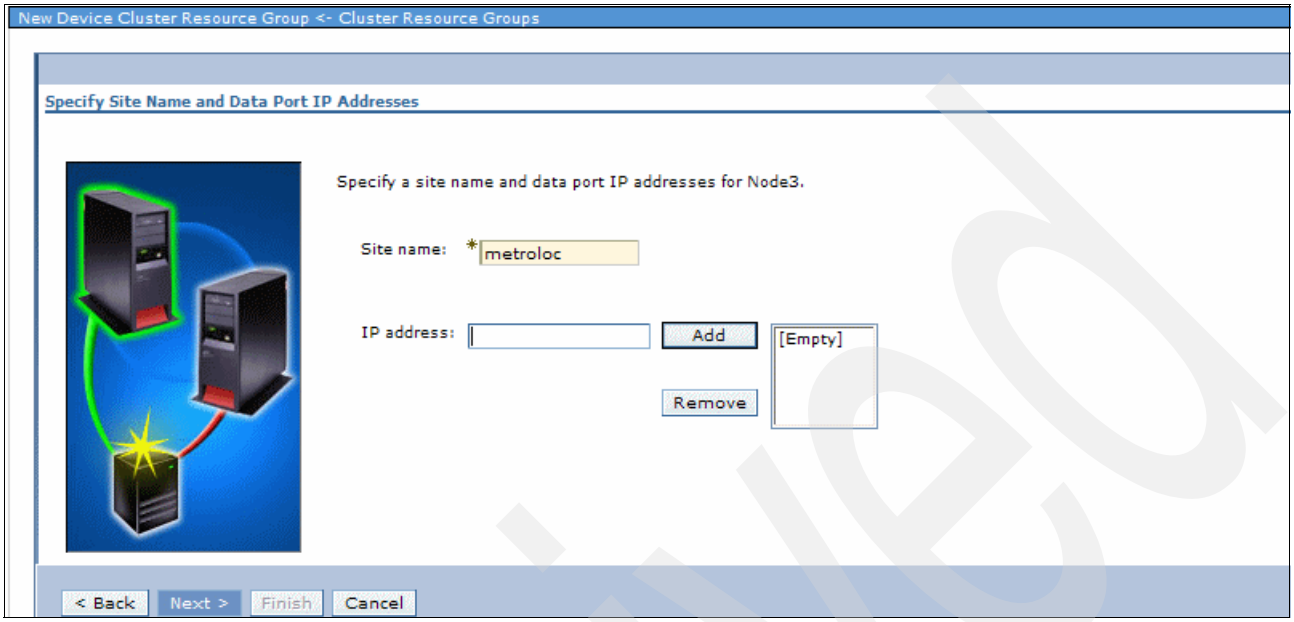


Figure 7-20 Add site name for primary node

5. Clicking **Next** brings you to a window where you can specify whether you want to add another node to this device CRG. You can also specify which node that should be, as shown in Figure 7-21. Again, make sure that you check the option to specify a site name.

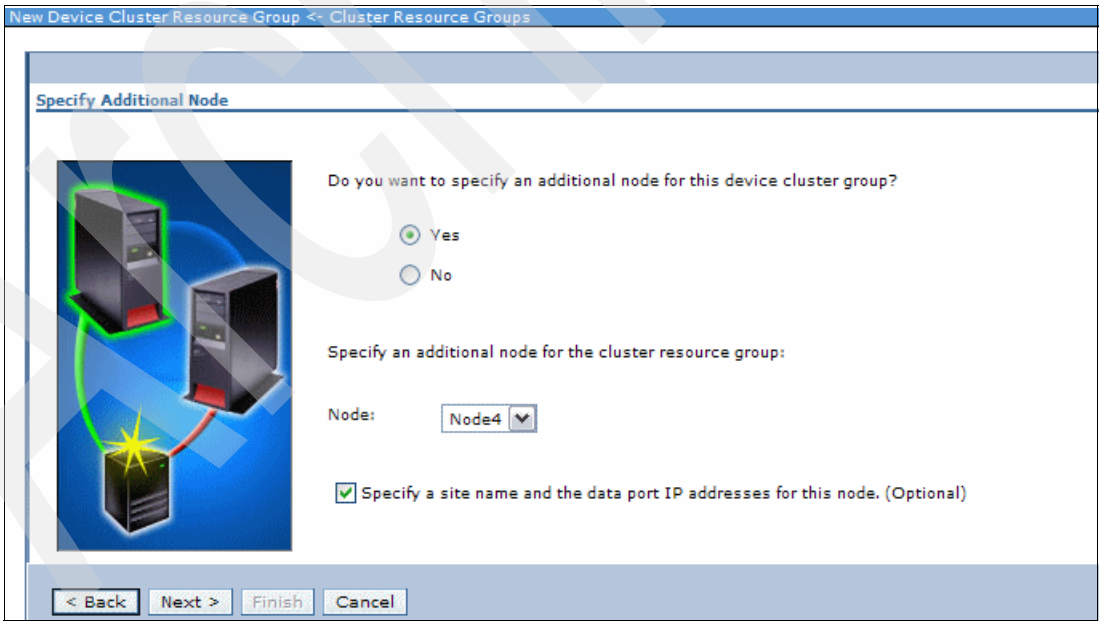


Figure 7-21 Specify backup node

6. After you specified the site name (and no data port addresses) in the same way that you did for the primary node you then get the opportunity to define a failover message queue specific to this CRG, as shown in Figure 7-22. As we have already specified a cluster-wide failover message queue, we do not choose this option here. Remember that if there is a cluster-wide failover message queue the CRG failover message queue is ignored.

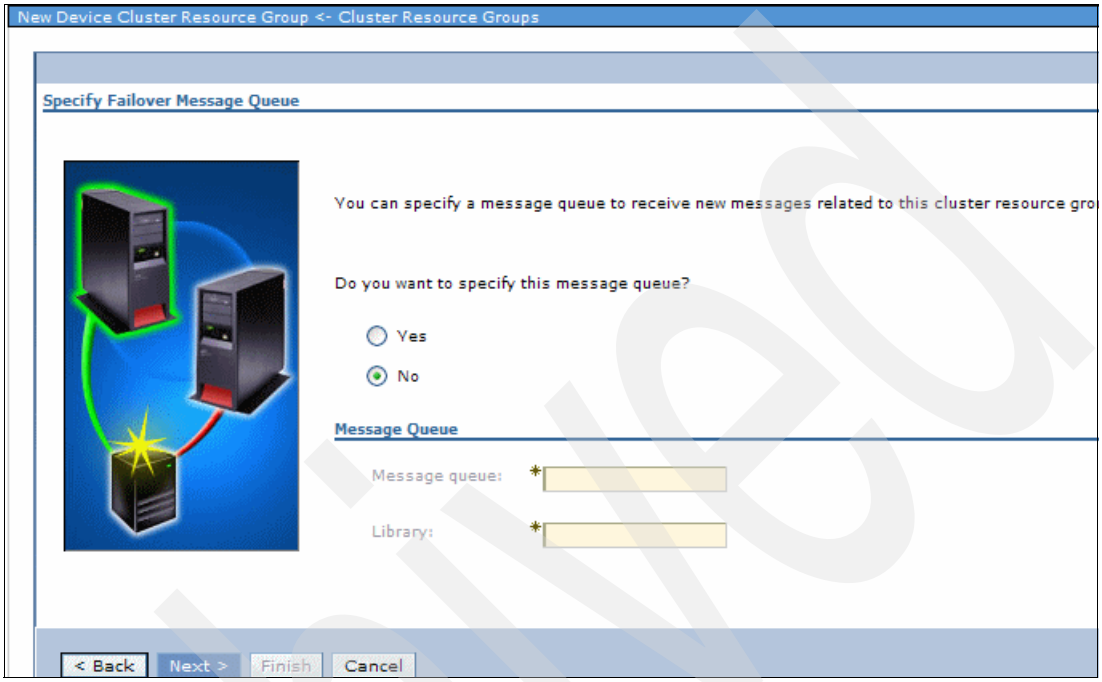


Figure 7-22 Specify Failover Message Queue

7. After clicking **Next** specify what kind of hardware is controlled by the CRG that you are creating. As shown in Figure 7-23, choose **Auxiliary storage pool** and click **Next**.

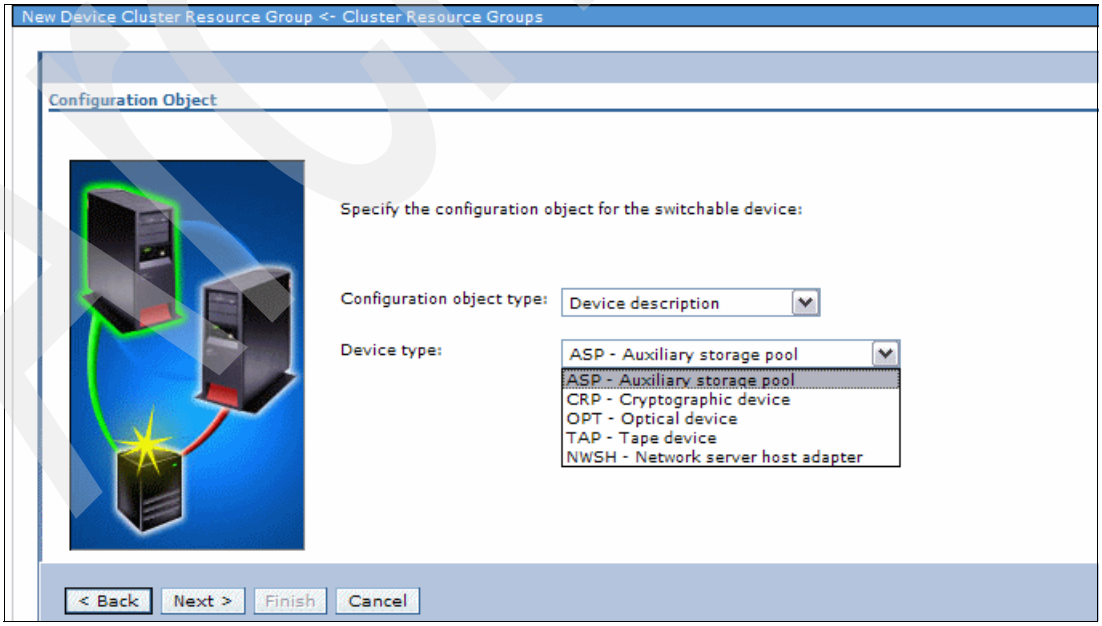


Figure 7-23 Specify configuration object

8. As the iASP already exists, we choose the according option to add an existing iASP to the CRG and also specify the name of that iASP, as shown in Figure 7-24.

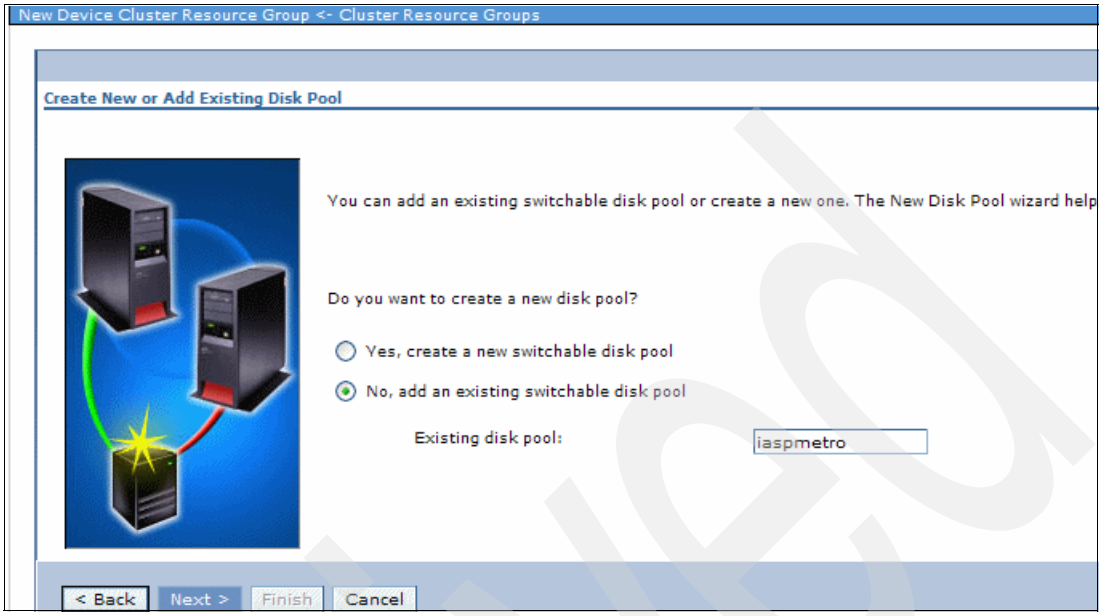


Figure 7-24 Add existing iASP to CRG

9. After clicking **Next** you are presented with on overview of the configuration data that you have just entered, as shown in Figure 7-25. Make sure that everything is correct and click **Finish**.

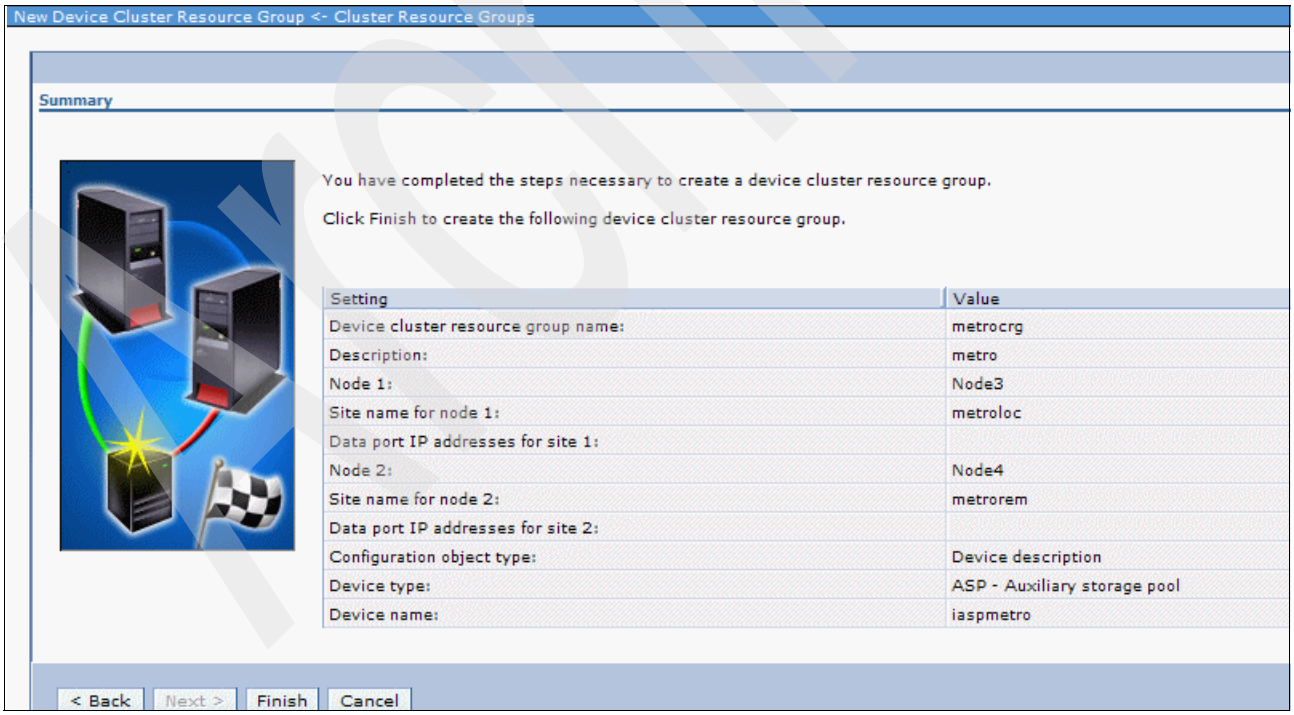


Figure 7-25 Create CRG: Summary

10. Once creation is done you can see your new CRG in the CRG list, as shown in Figure 7-26. By selecting the double arrow beside the CRG name you can now start that CRG.

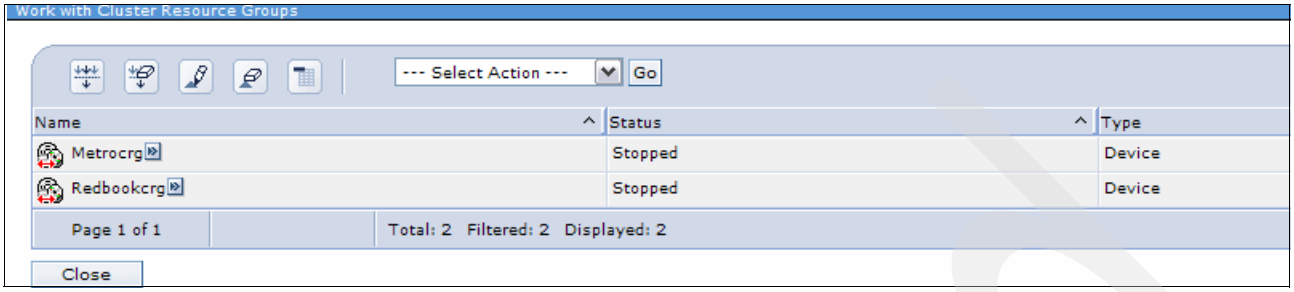


Figure 7-26 CRG creation completed

Create ASP copy description and ASP session description

In the next steps we create the ASP copy descriptions and the ASP session description. These objects are used by PowerHA for i to manage the communication between clustering and the Storage system.

1. On the disk pool window of IBM System Director Navigator select the double arrow beside the iASP. Choose **Session** → **New** → **Metro Mirror**, as shown in Figure 7-27.

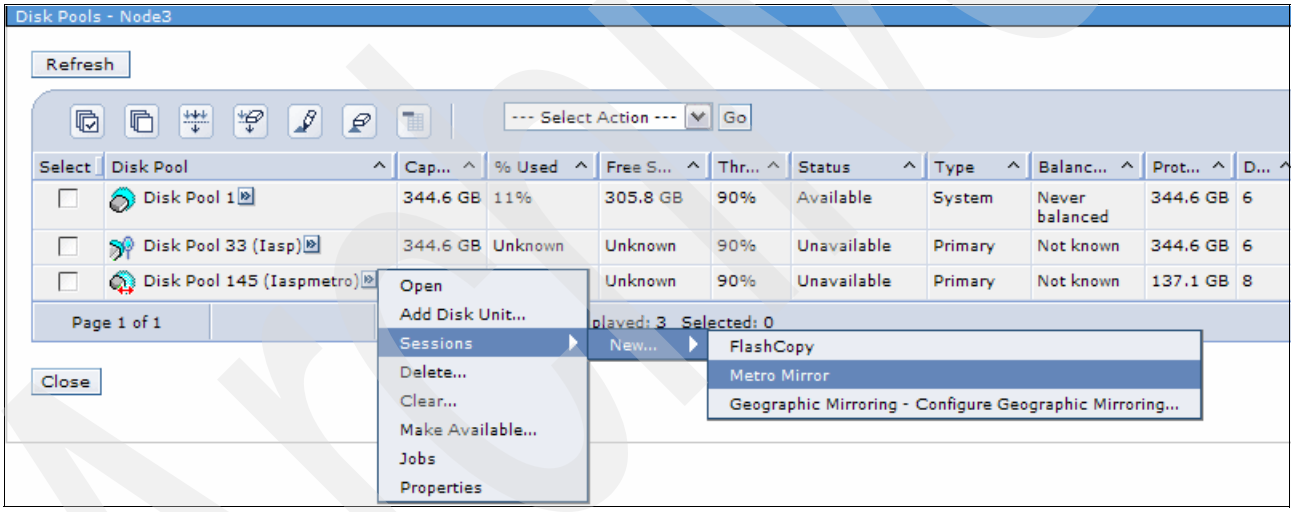


Figure 7-27 Create new session for metro mirror

2. The next window shows you the welcome wizard for the session created. Click **Next**, as shown in Figure 7-28.

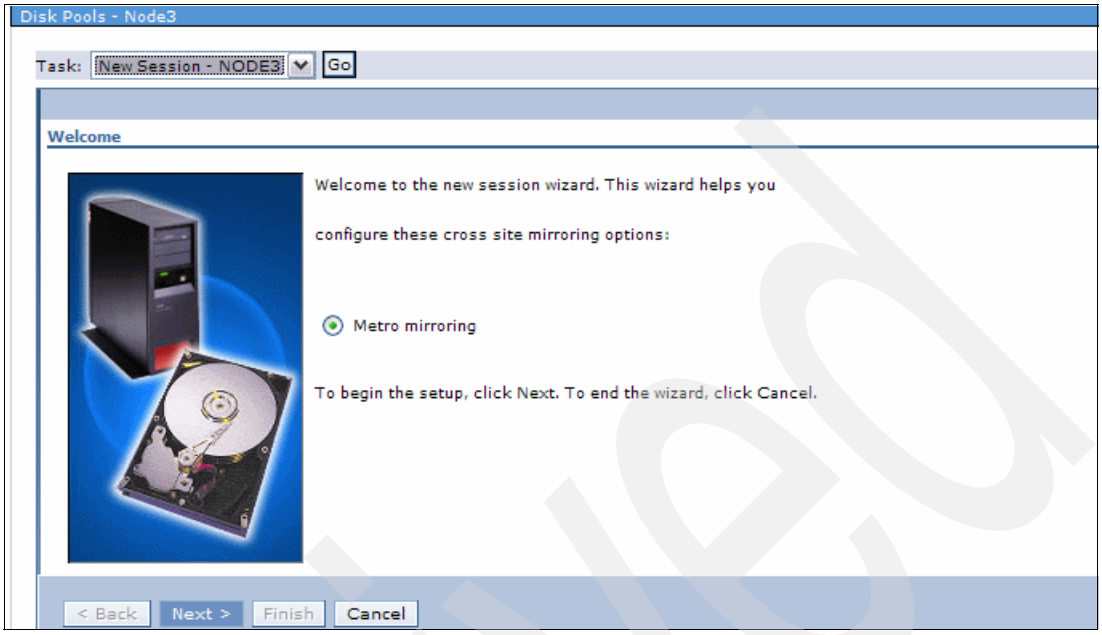


Figure 7-28 New session metro mirror wizard

3. Create an ASP copy description for the local copy of your iASP. As shown in Figure 7-29, the list of copy descriptions is empty at the moment. Click **Add Copy Description** to create a new one.

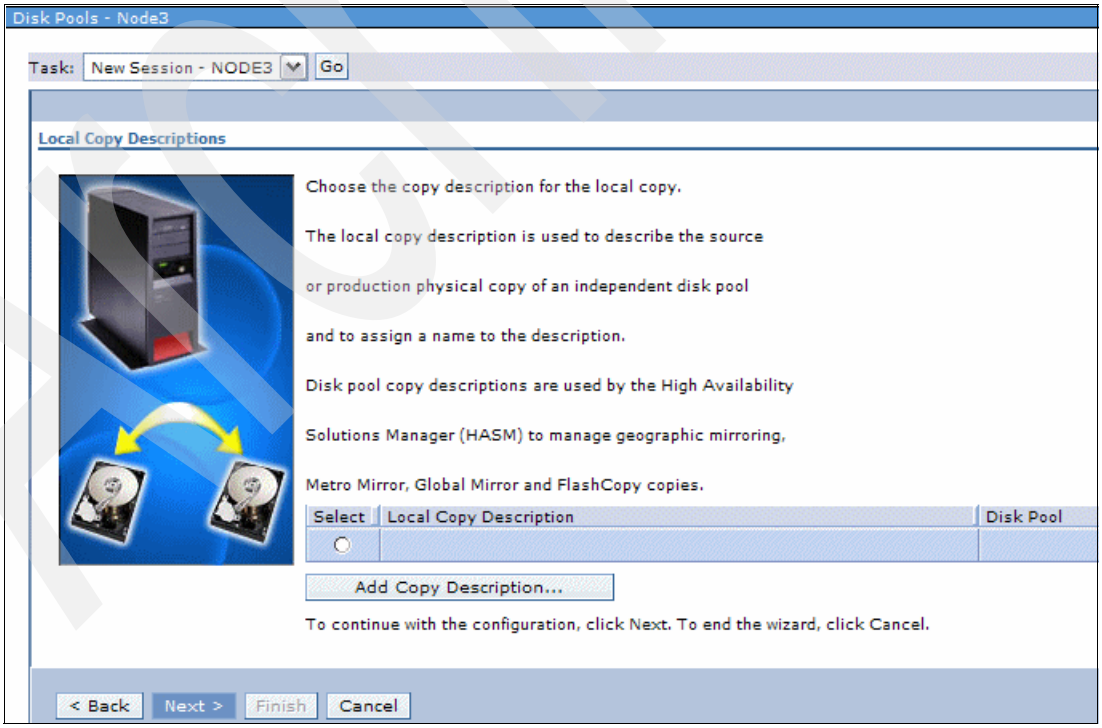


Figure 7-29 Add copy description for local copy

4. You first have to provide a name for your copy description (metroloc in our example), as well as the CRG that is associated with your iASP and the site name of the local node, as shown in Figure 7-30. As this is a copy description for a metro mirror session, you also must provide some information about your Storage system. Select the **Add** button on the right-hand side of the storage hosts table to do so.

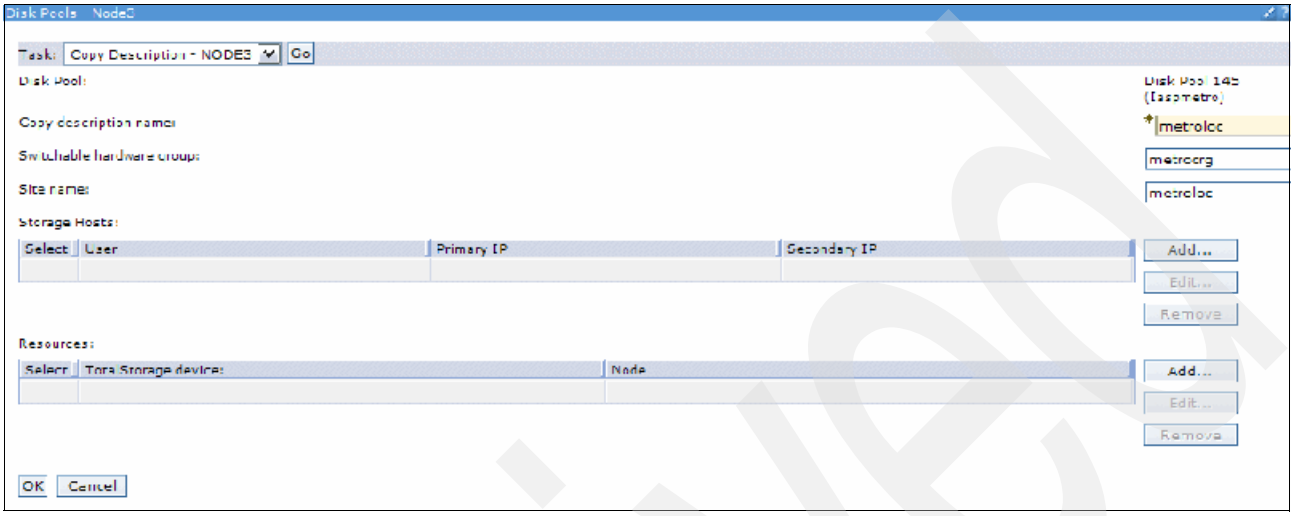


Figure 7-30 Copy description: Basic info

5. In the next panel enter the user ID that is used to access the storage management console, its password, and the IP address of the storage management console, as shown in Figure 7-31. Clicking **OK** brings you back to the panel shown in Figure 7-30. On that panel you then select the **Add** button on the right-hand side of the resources table.

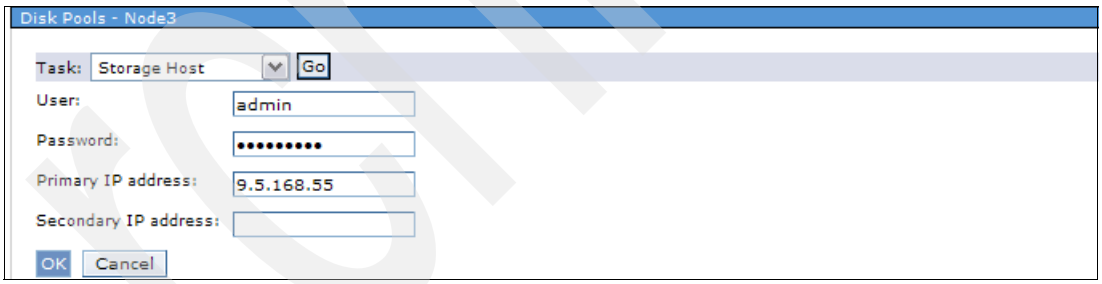


Figure 7-31 Copy description: Sign on to storage server information

6. The last step opens up the window shown in Figure 7-32. This table is empty. By clicking the **Add** button you can enter the LUN ranges that your iASP occupies in your Storage system. Once you have entered all the necessary data, click **OK**.

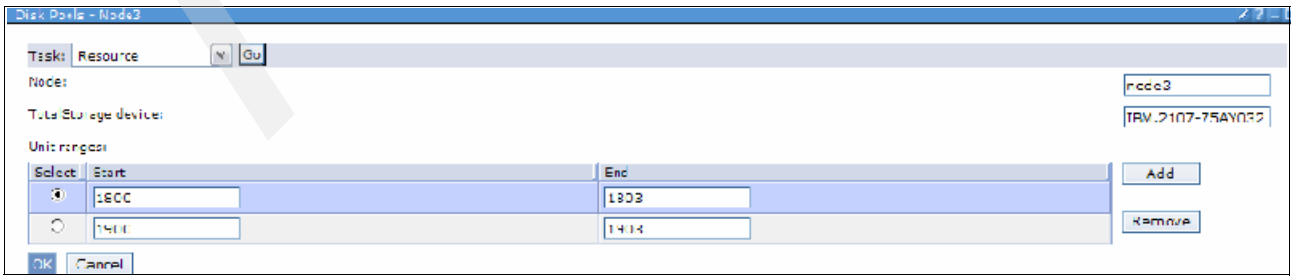


Figure 7-32 Copy description: LUN configuration

7. The wizard presents you with an overview of the data that you just entered, as shown in Figure 7-33. If all the data is correct, click **OK**.

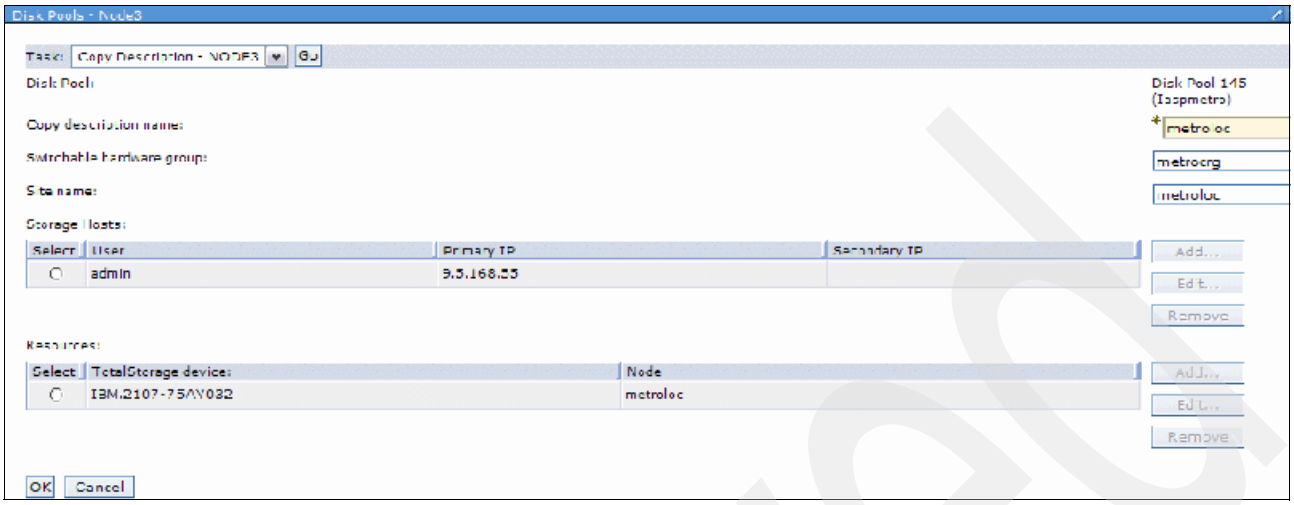


Figure 7-33 Copy description: Overview

8. You are brought back to the panel that shows you the available copy descriptions. Notice that now the table contains the copy description that you just created. Make sure to check the radio button next to the copy description, as shown in Figure 7-34, and click **Next**.

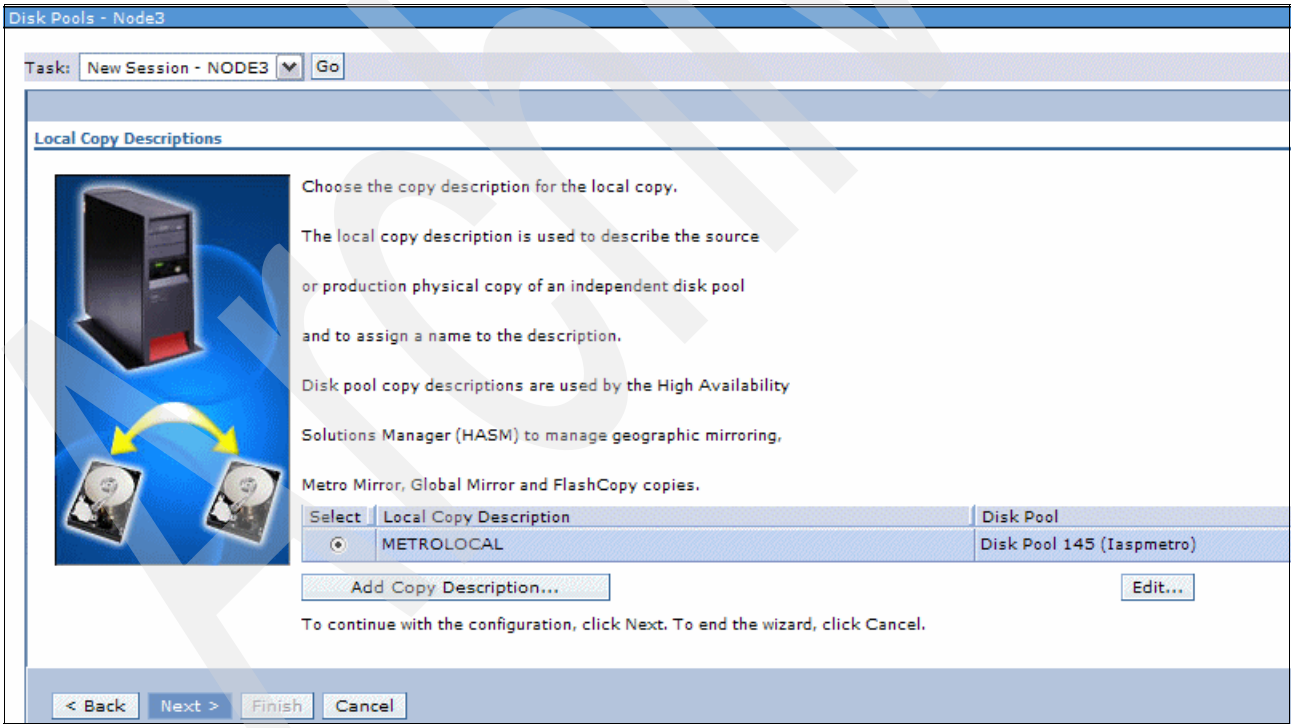


Figure 7-34 Choose copy description

9. You now must create a copy description for the remote copy of your iASP, as shown in Figure 7-35. The steps required to do this are exactly the same as for the local copy. Make sure that you enter the correct information in the panels for the storage server, site name, and LUN ranges.

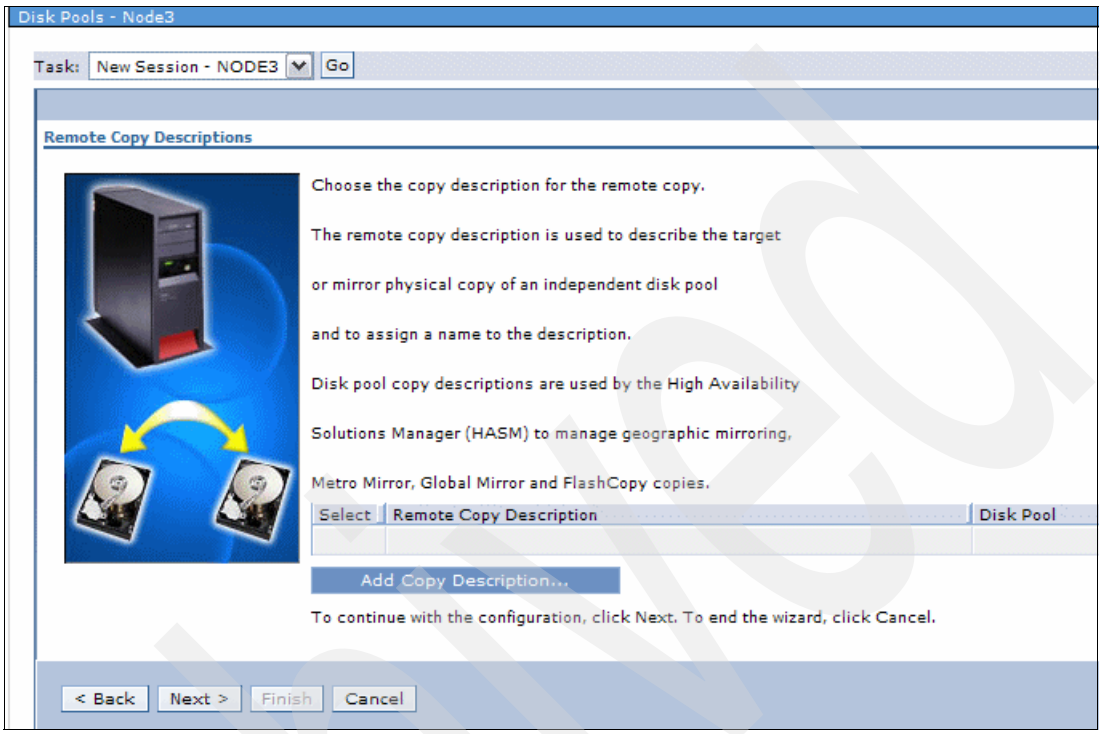


Figure 7-35 Add copy description for remote copy

10. Once you have successfully added the remote copy description and clicked **Next** you can set up the session itself. You first must provide a session name, as shown in Figure 7-36.

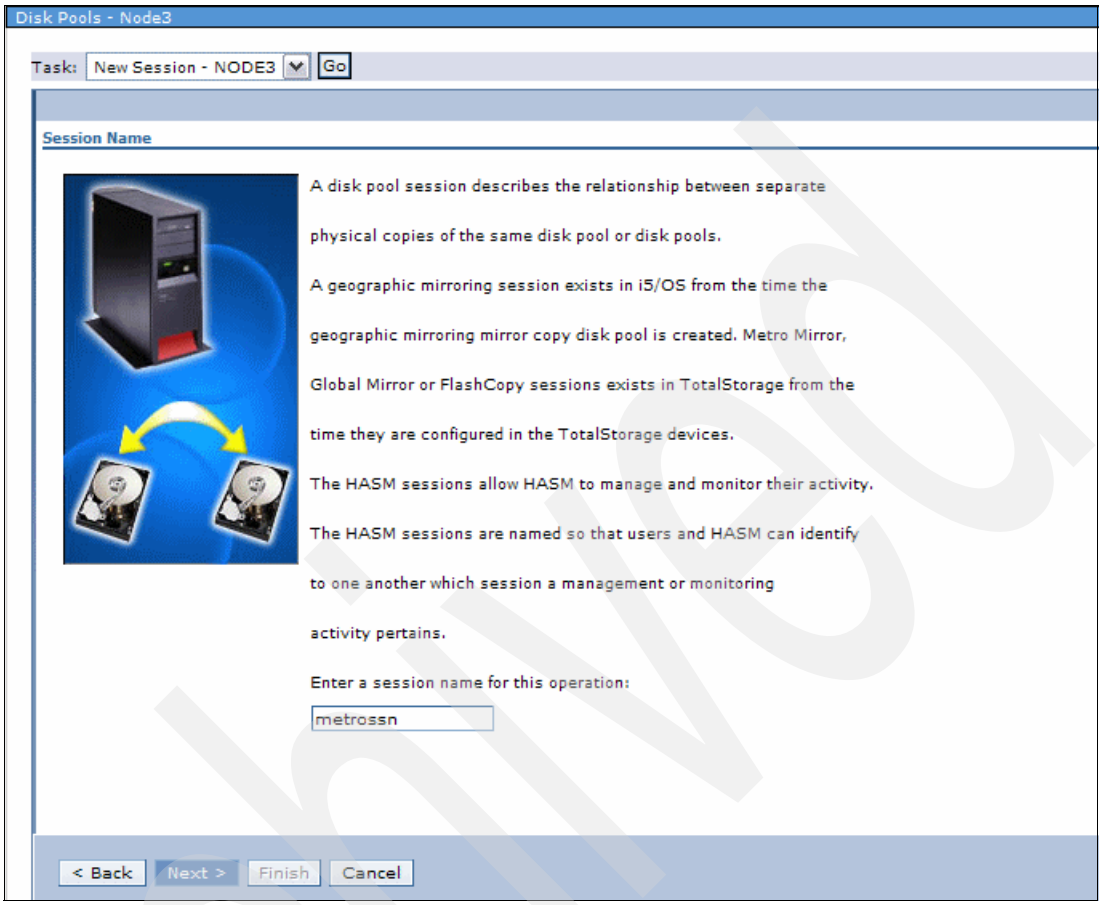


Figure 7-36 Create session description: Name

11. Clicking **Next** provides you with an overview of the session that you are about to create. If all the information is correct, click **Finish**, as shown in Figure 7-37.

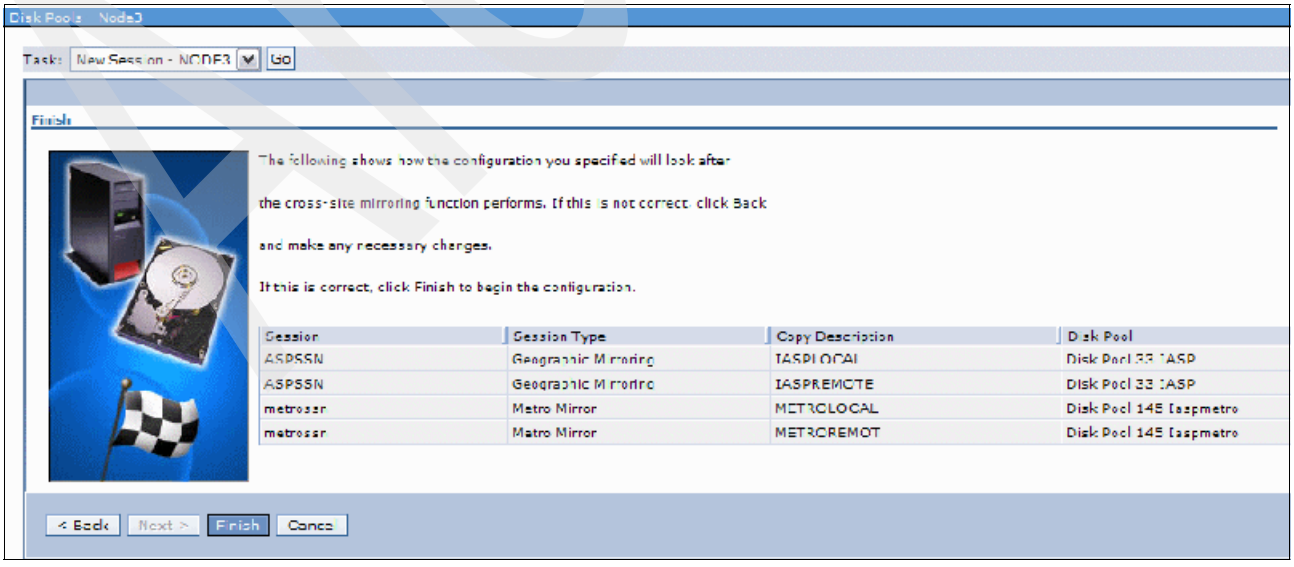


Figure 7-37 Metro mirror session overview

Using the GUI, this session is automatically started when it is created. You are now ready to use your iASP metro mirror environment.

If you receive the error shown in Figure 7-38 then this is an indication that you have not done the configuration of metro mirror correctly on your Storage system. Remember that you must define a metro mirror path from your primary iASP to the backup iASP (used during normal production process), as well as a metro mirror path from the backup iASP to the primary iASP (used after a switchover or failover occurred).

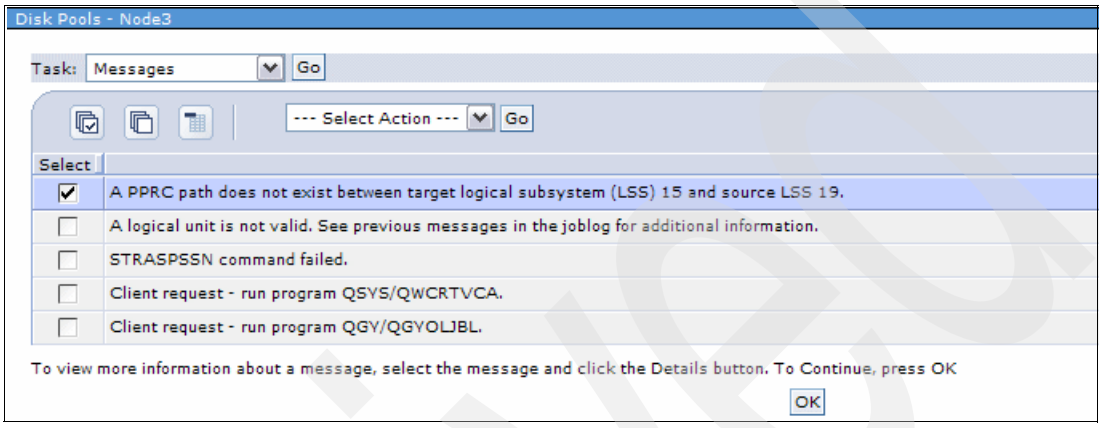


Figure 7-38 Error message

7.3 Working with the metro mirror environment

In this section we describe the various steps that you can take while working with your metro mirror environment.

7.3.1 Suspending

The suspend operation can be done from either the source or the target node of the metro mirror session and does not require that your iASP be varied off. Suspend only interrupts the sending of data from the source Storage system to the target Storage system. In a metro mirror environment a suspend is always done with tracking. Access to the iASP on the backup system is *not* possible in a suspended status.

To suspend a metro mirror session start in the Disk pools window. (We showed how to get there in 7.6.5, “Disk GUI” on page 226.)

1. Select the double arrow of the metro mirror iASP and from the drop-down menu and select **Sessions** → **Open**, as illustrated in Figure 7-39.

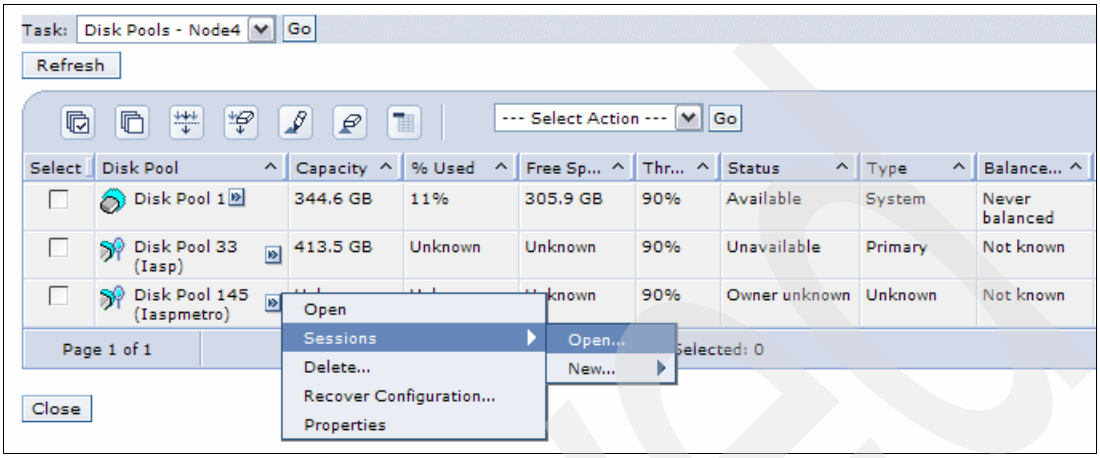


Figure 7-39 Working with Sessions

2. Select the **Metro Mirror** session, then from the drop-down menu choose **Suspend** and click **Go**, as shown in Figure 7-40.

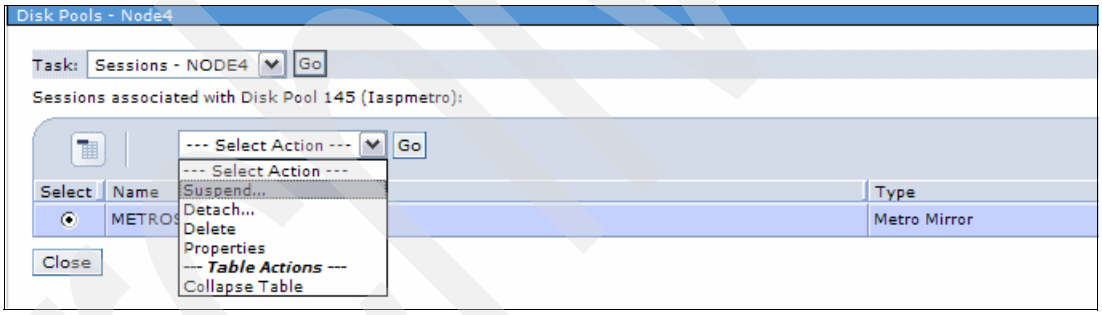


Figure 7-40 Choose the action suspend

3. Click **Suspend** to confirm the action, as shown in Figure 7-41.

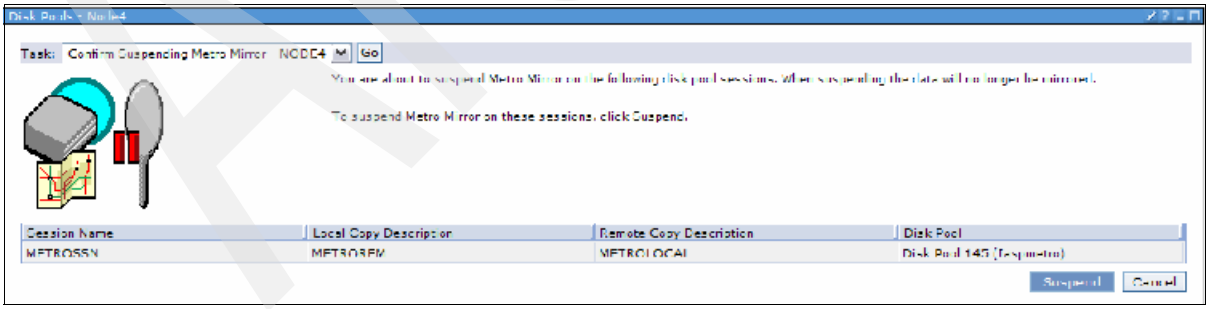


Figure 7-41 Confirm Suspend

4. Click **OK** to exit, as shown in Figure 7-42.

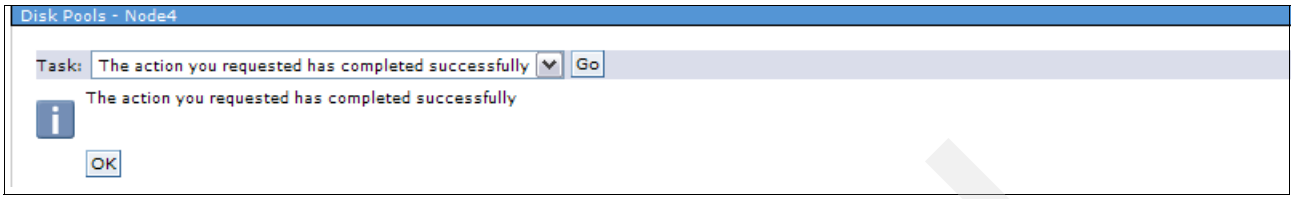


Figure 7-42 Suspend completed

7.3.2 Resuming

The resume operation re-establishes a suspended session. Metro mirror between the source and the target Storage system is started and tracked changes are replicated to the target. It can be run from either the source or the target node of the metro mirror session.

1. To resume a metro mirror session start on the Disk pools panel. Select the double arrow against the metro mirror iASP and from the drop-down menu then select **Sessions** → **Open**, as illustrated in Figure 7-43.

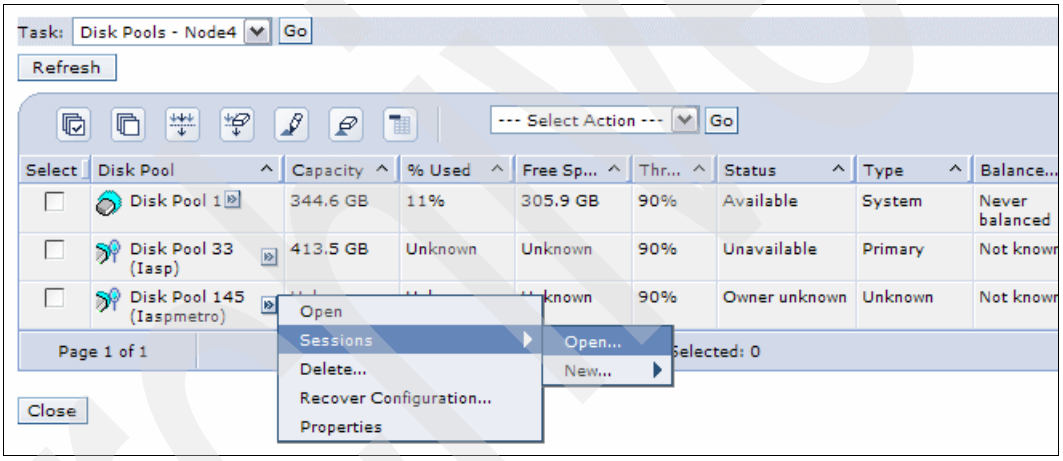


Figure 7-43 Working with sessions

2. Select the **Metro Mirror** session, then from the drop-down menu choose **Resume** and click **Go**, as shown in Figure 7-44.

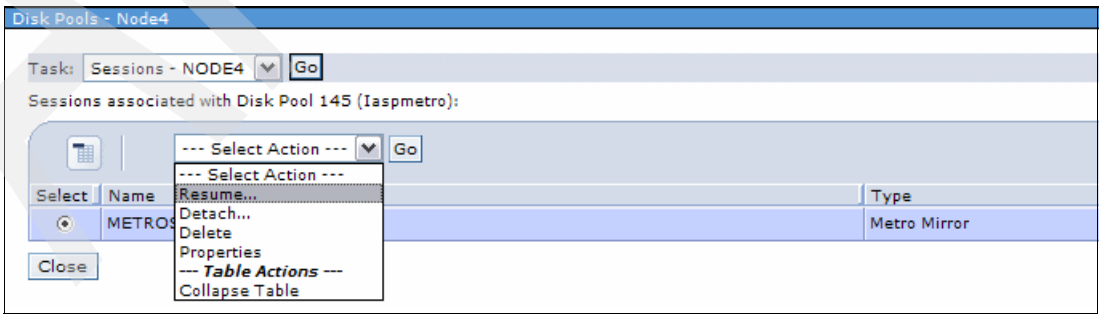


Figure 7-44 Resume action

3. Click **Resume** to confirm, as shown in Figure 7-45

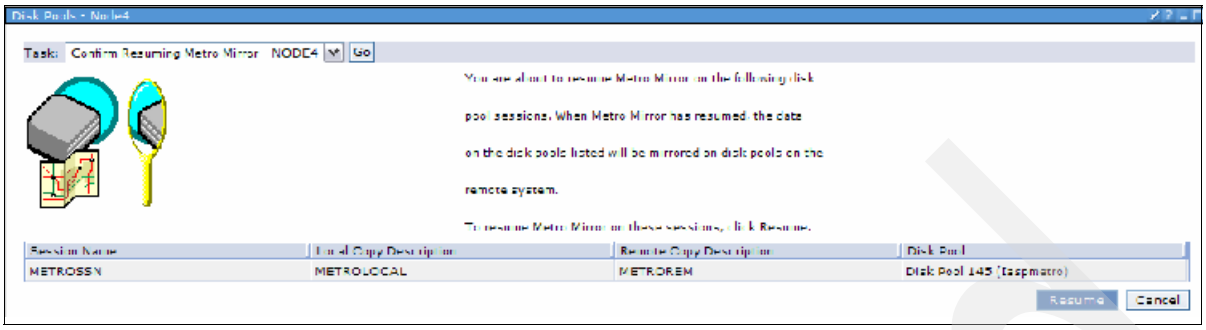


Figure 7-45 Confirm resume action

4. Click **OK** to exit, as shown in Figure 7-46

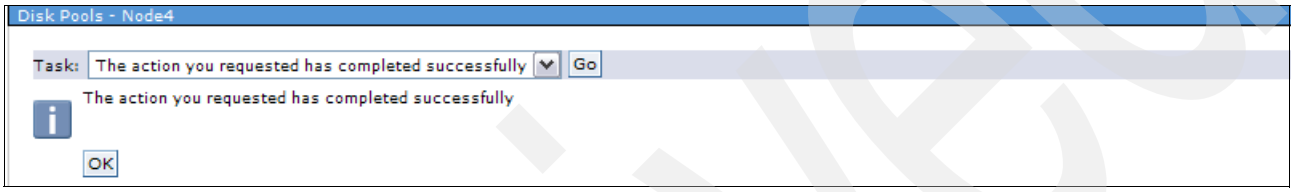


Figure 7-46 Resume completed

7.3.3 Detaching

The detach operation allows us to access the mirror copy to perform save operations or data mining. You must detach the mirror copy from the production copy.

The detach must be done from the current source node of the metro mirror with the iASP in VARY-OFF.

1. From command line run the DSPASPSSN command to check which node is the current source of the metro mirror, as shown in Figure 7-47.

```
                  Display ASP Session                   NODE3

 Session . . . . . . . . . . . . . . . . . . . :   METROSSN
   Type  . . . . . . . . . . . . . . . . . . . :   *METROMIR




                                                                  Botto
                              Copy Descriptions

 ASP                                                          Data
 device        Name                 Role          State       State     Node
 IASPMETRO     METROLOCAL           SOURCE        VARYOFF     USABLE    NODE3
 IASPMETRO     METROREM             TARGET        ACTIVE      UNUSABLE  NODE4
```

Figure 7-47 DSPASPSSN command

2. On the Disk pools panel (we showed how to get here in 7.6.5, “Disk GUI” on page 226), select the double arrow against the metro mirror iASP and from the drop-down menu select **Sessions** → **Open**, as illustrated in Figure 7-48.

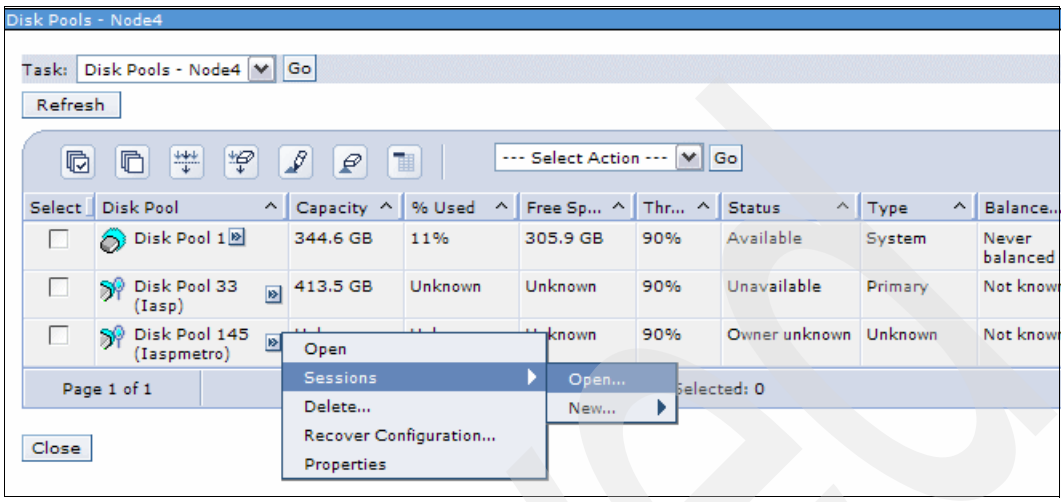


Figure 7-48 Working with sessions

3. Select the **Metro Mirror** session, then from the drop-down menu choose **Detach** and click **Go**, as shown in Figure 7-49.

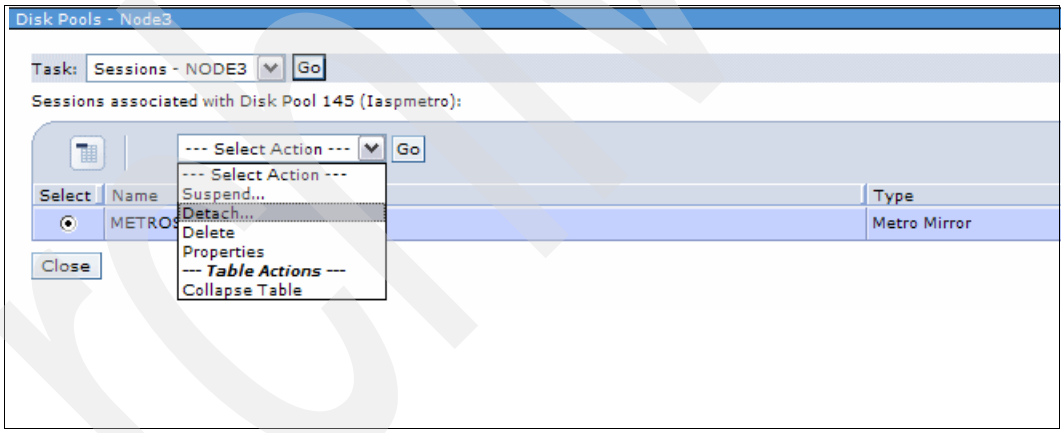


Figure 7-49 Detach action

The next panel shows the warning message that you get if you are trying to detach an iASP that is still available (Figure 7-50).

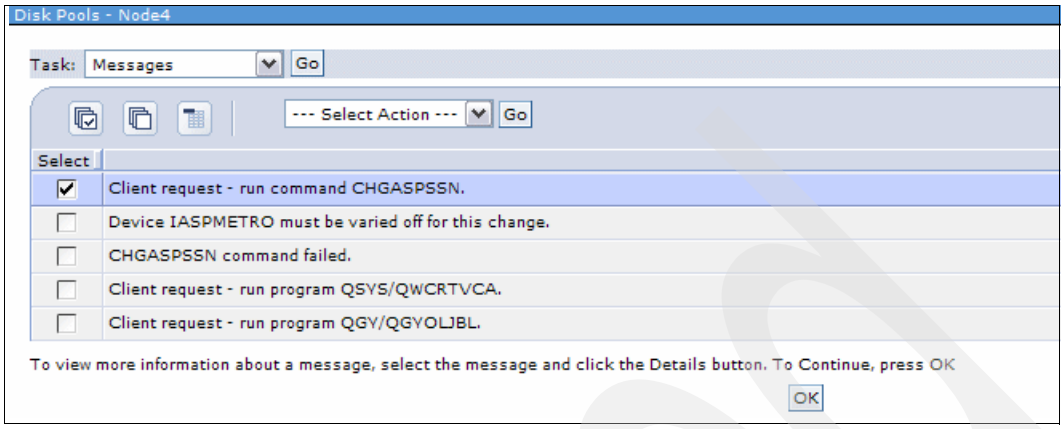


Figure 7-50 Detach warning message

4. The next panel shows that the action has been successfully completed. Click **OK** to exit (Figure 7-51).

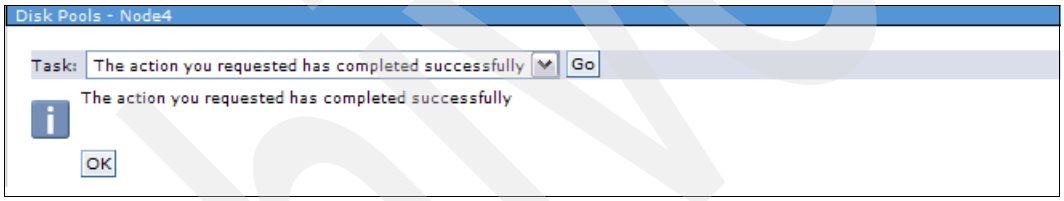


Figure 7-51 Detach completed

5. The DSPASPSNN command run from primary node shows the target ASP device detached and in suspended state, as shown in Figure 7-52.

```
                    Display ASP Session                    NODE3

Session . . . . . . . . . . . . . . . . . . :   METROSSN
Type  . . . . . . . . . . . . . . . . . . . :  *METROMIR




                                                                Bot
                          Copy Descriptions

ASP                                                        Data
device       Name              Role          State        State    Node
IASPMETRO    METROLOCAL      SOURCE        VARYOFF        USABLE   NODE3
IASPMETRO    METROREM      DETACHED      SUSPENDED        USABLE   NODE4
```

Figure 7-52 DSPASPSSN from source site after detach

6. The DSPASPSSN command run from the backup node shows the ASP device UNKNOWN for both the source and the target copy (Figure 7-53).

```
                          Display ASP Session                       NODE4
Session . . . . . . . . . . . . . . . . . . . :   METROSSN
Type  . . . . . . . . . . . . . . . . . . . . :     *METROMIR




Copy Descriptions

 ASP                                                          Data
 device        Name               Role        State         State    Node
 IASPMETRO     METROLOCAL         SOURCE      UNKNOWN       UNKNOWN  NODE3
 IASPMETRO     METROREM           TARGET      UNKNOWN       UNKNOWN  NODE4
```

Figure 7-53 DSPASPSSN from target site after detach

The next window shows the suspended state of the source metro mirror.

```
dscli> lspprc -l -dev IBM.2107-75AY032 1800-1803 1900-1903
Date/Time: 27 maggio 2008 19.58.29 CEST IBM DSCLI Version: 5.2.2.372 DS: IBM.2107-75AY032
ID        State     Reason      Type         Out Of Sync Tracks Tgt Read Src Cascade
=============================================================================================
1800:1400 Suspended Host Source Metro Mirror 5                  Disabled Disabled
1801:1401 Suspended Host Source Metro Mirror 3                  Disabled Disabled
1802:1402 Suspended Host Source Metro Mirror 3                  Disabled Disabled
1803:1403 Suspended Host Source Metro Mirror 3                  Disabled Disabled
1900:1500 Suspended Host Source Metro Mirror 3                  Disabled Disabled
1901:1501 Suspended Host Source Metro Mirror 3                  Disabled Disabled
1902:1502 Suspended Host Source Metro Mirror 3                  Disabled Disabled
1903:1503 Suspended Host Source Metro Mirror 3                  Disabled Disabled
```

Figure 7-54 lspprc from DSCLI on source copy

7.3.4 Reattaching

The reattach must be done from the current target node of the metro mirror.

The iASP on the target node must be varied off to perform the reattach, while the iASP on the source node can be available.

Start on the Disk pools panel (which we showed how to get to in 7.6.5, “Disk GUI” on page 226):

1. Select the iASP that you want to reattach, then click the double arrow and select **Sessions** → **Open**. See Figure 7-55.

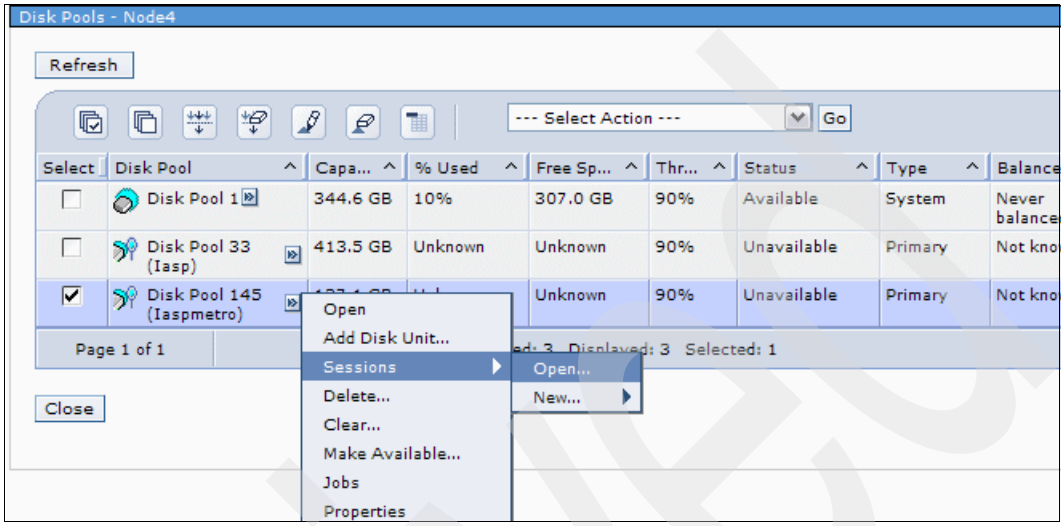


Figure 7-55 Working with sessions

2. Select the **Metro Mirror** session and from the drop-down menu choose **Reattach,** as shown Figure 7-56.

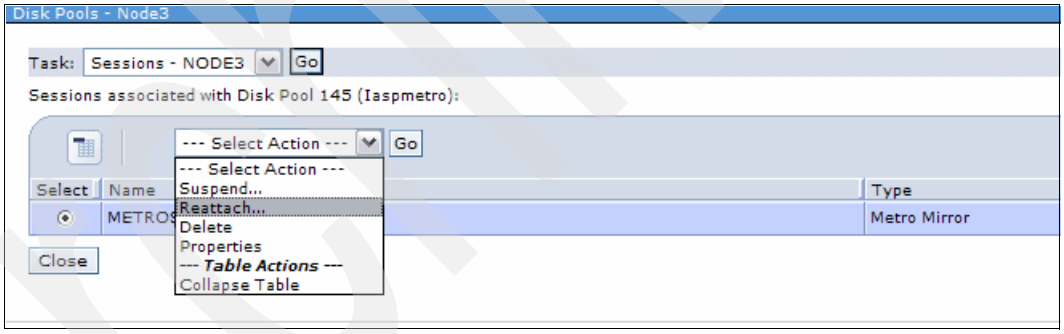


Figure 7-56 Reattach action

3. Click **Continue**, as shown in Figure 7-57.

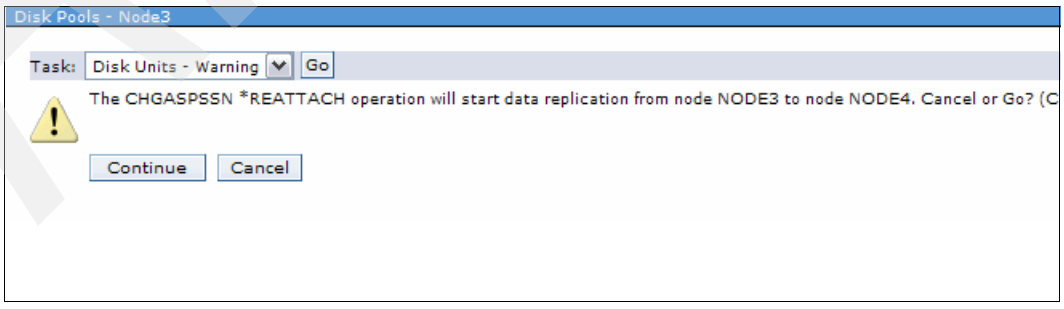


Figure 7-57 Reattach action

The next window (Figure 7-58) shows the error that you get if you try to reattach the metro mirror copy from the source node of the metro mirror instead of from the target node.

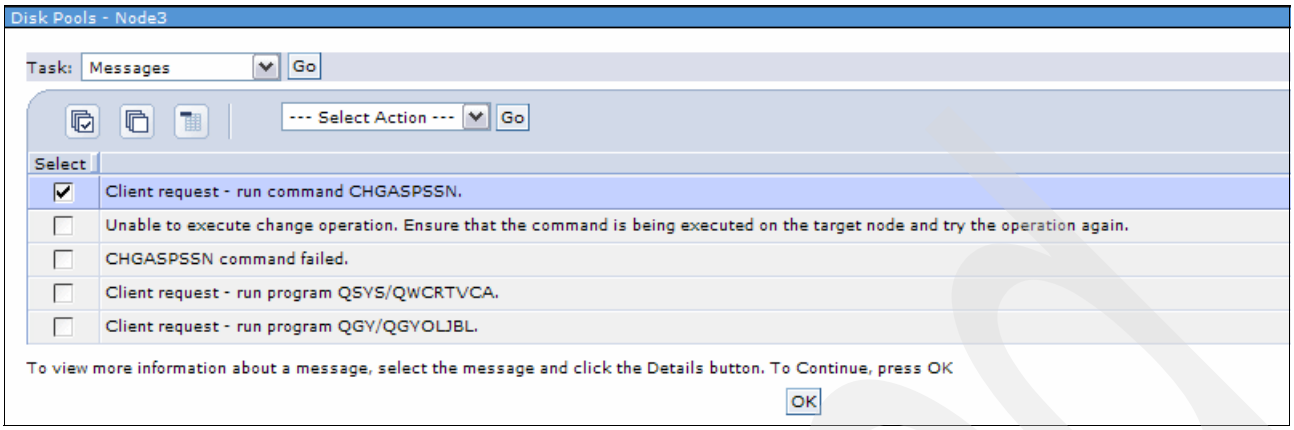


Figure 7-58 Reattach done from source node failed

If you attempt to reattach the metro mirror copy from the target node but the iASP is not varied off you receive another warning message saying that the iASP device must be varied off to complete the CHGASPSSN command, as shown in Figure 7-59.

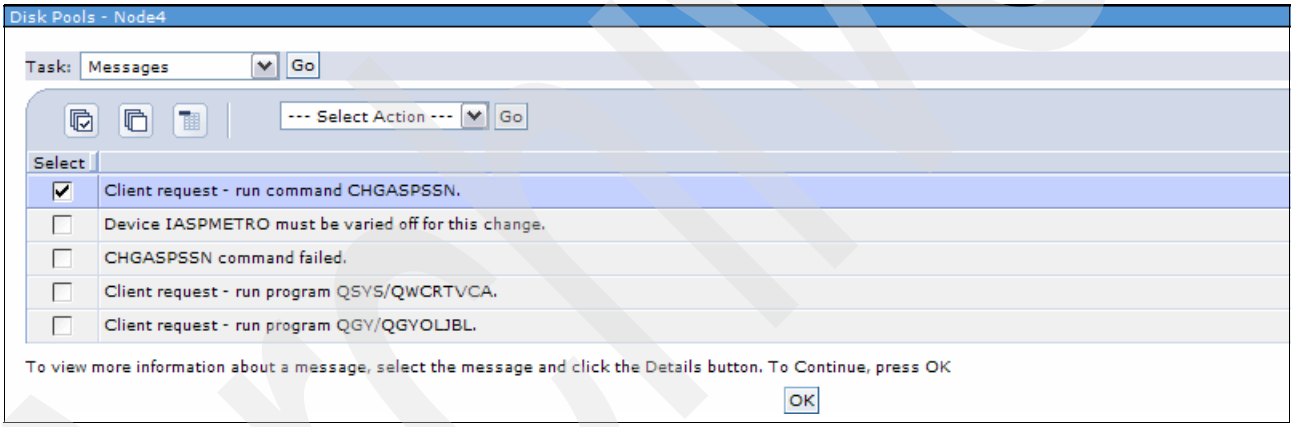


Figure 7-59 Reattach done from target node with iASP available failed

4. You will see that the action completed successfully. Click **OK** to exit, as shown in Figure 7-60.

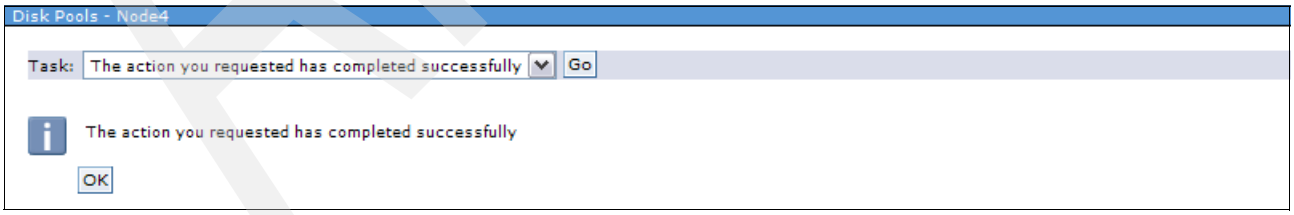


Figure 7-60 Reattach complete

In Figure 7-61 you can see the command **lspprc** after reattach from source mirror copy.

```
dscli> lspprc -l -dev IBM.2107-75AY032 1800-1803 1900-1903
Date/Time: 27 maggio 2008 21.00.45 CEST IBM DSCLI Version: 5.2.2.372 DS: IBM.2107-75AY032
ID        State       Reason Type         Out Of Sync Tracks Tgt Read
==============================================================================
1800:1400 Full Duplex -      Metro Mirror 0                  Disabled
1801:1401 Full Duplex -      Metro Mirror 0                  Disabled
1802:1402 Full Duplex -      Metro Mirror 0                  Disabled
1803:1403 Full Duplex -      Metro Mirror 0                  Disabled
1900:1500 Full Duplex -      Metro Mirror 0                  Disabled
1901:1501 Full Duplex -      Metro Mirror 0                  Disabled
1902:1502 Full Duplex -      Metro Mirror 0                  Disabled
1903:1503 Full Duplex -      Metro Mirror 0                  Disabled
```

Figure 7-61 lspprc after reattach from source mirror copy

In Figure 7-62 you can see the command **lspprc** after reattach from target mirror copy.

```
dscli> lspprc -l -dev IBM.2107-75AY031 1400-1403 1500-1503
Date/Time: 27 maggio 2008 21.21.01 CEST IBM DSCLI Version: 5.2.2.372 DS: IBM
ID        State              Reason Type        Out Of Sync Tracks Tgt Read
==============================================================================
1800:1400 Target Full Duplex -      Metro Mirror 0                 Disabled
1801:1401 Target Full Duplex -      Metro Mirror 0                 Disabled
1802:1402 Target Full Duplex -      Metro Mirror 0                 Disabled
1803:1403 Target Full Duplex -      Metro Mirror 0                 Disabled
1900:1500 Target Full Duplex -      Metro Mirror 0                 Disabled
1901:1501 Target Full Duplex -      Metro Mirror 0                 Disabled
1902:1502 Target Full Duplex -      Metro Mirror 0                 Disabled
1903:1503 Target Full Duplex -      Metro Mirror 0                 Disabled
```

Figure 7-62 lspprc after reattach from target mirror copy

7.3.5 Switching

In a metro mirror environment to perform a switchover you can use the command CHGCRGPRI or the switch option from the GUI interface (as shown in Figure 7-63).

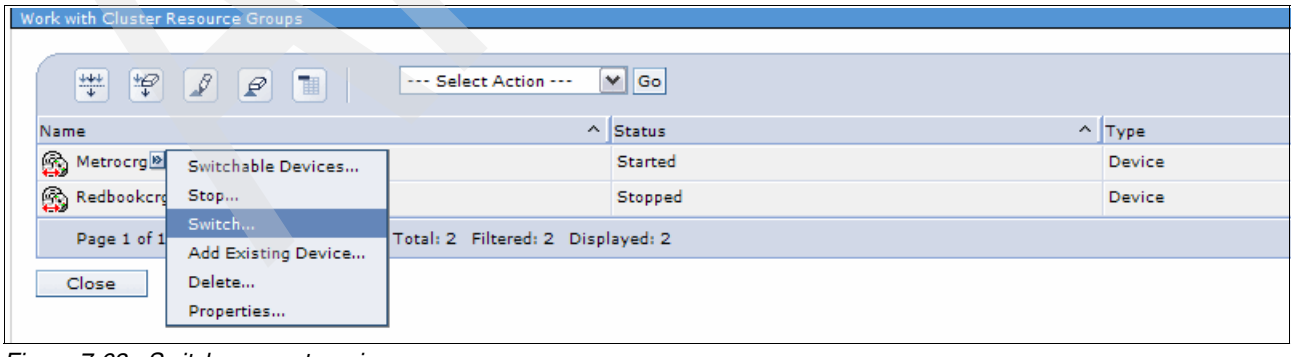


Figure 7-63 Switchover metro mirror

After a switchover is completed the original backup system becomes the new primary system and the source node of the ASP metro mirror session. The data flow direction (reverse direction) is also changed from the original target to the original source node.

To run a switchover from the GUI interface, click **Switch** from the drop-down menu of the metro mirror CRG device, as shown in Figure 7-63 on page 194.

Before the switchover or a failover is performed a health check of the Storage system is performed. This also happens in regular intervals during normal operation. If the health check finds a problem in the Storage system (MetroMirror is disabled for one or more disks of the iASP) then the remote node is set to ineligible and switchover and failover cannot occur.

From command line run the DSPASPSSN to verify that the NODE ROLE in the asp metro mirror session is changed. See Figure 7-64.

```
Display ASP Session                                        NODE4

Session . . . . . . . . . . . . . . . . . . :   METROSSN
Type  . . . . . . . . . . . . . . . . . . . :     *METROMIR




                                                                Bot
                              Copy Descriptions

ASP                                                        Data
device        Name                Role          State      State     Node
IASPMETRO     METROREM          SOURCE      AVAILABLE      USABLE    NODE4
IASPMETRO     METROLOCAL        TARGET        ACTIVE    UNUSABLE    NODE3
```

Figure 7-64 DSPASPSSN window

Run DSPCRGINF to verify that the role of the nodes in the cluster is also changed. The backup is now the primary node.

7.3.6 Deleting the metro mirror environment

If you want to delete your metro mirror environment, this has to be done in several steps on your IBM i as well as on your Storage system:

1. Delete the ASP session that represents your metro mirror environment.
2. Delete the ASP copy descriptions that were used by this session unless they are used by some other ASP session.
3. Remove the metro mirror setup on your Storage system. This must be done using the Storage system itself.

7.3.7 Storage system analysis

There is a log available of the communication between IBM i and the attached Storage system. It can be found in the directory /ibm, which was created as part of the installation of DSCLI. It is named xsm.log. An example of this log can be seen in Figure 7-65.

```
 Browse : /ibm/xsm.log
  Record :       1   of    2042 by  18                      Column :    1    329
by 131
 Control :

....+....1....+....2....+....3....+....4....+....5....+....6....+....7....+....
8....+....9....+....0....+....1....+....2....+....3.
 ************Beginning of data**************
<*>05/21/2008 08:14:43 8F0B7F52CEB84001 <*INFO > : QYASPPPRC : switchPPRC :
Start of PPRC switch. CRG: METROCRG
<*>05/21/2008 08:14:53 8F0B7F52CEB84001 QDSCLI/DSCLI SCRIPT(*NONE)
PROFILE('/ibm/XSMDS.profile') USER(admin) DEV('IBM.2107-75AY032')
1800:1400,Full Duplex,-,Metro Mirror,18,60,Disabled,Invalid
1801:1401,Full Duplex,-,Metro Mirror,18,60,Disabled,Invalid
1802:1402,Full Duplex,-,Metro Mirror,18,60,Disabled,Invalid
1803:1403,Full Duplex,-,Metro Mirror,18,60,Disabled,Invalid
1900:1500,Full Duplex,-,Metro Mirror,19,60,Disabled,Invalid
1901:1501,Full Duplex,-,Metro Mirror,19,60,Disabled,Invalid
1902:1502,Full Duplex,-,Metro Mirror,19,60,Disabled,Invalid
1903:1503,Full Duplex,-,Metro Mirror,19,60,Disabled,Invalid

<*>05/21/2008 08:15:03 8F0B7F52CEB84001 QDSCLI/DSCLI SCRIPT(*NONE)
PROFILE('/ibm/XSMDS.profile') USER(admin) DEV('IBM.2107-75AY031')
1800:1400,Target Full Duplex,-,Metro Mirror,18,unknown,Disabled,Invalid
1801:1401,Target Full Duplex,-,Metro Mirror,18,unknown,Disabled,Invalid
1802:1402,Target Full Duplex,-,Metro Mirror,18,unknown,Disabled,Invalid
1803:1403,Target Full Duplex,-,Metro Mirror,18,unknown,Disabled,Invalid
1900:1500,Target Full Duplex,-,Metro Mirror,19,unknown,Disabled,Invalid
1901:1501,Target Full Duplex,-,Metro Mirror,19,unknown,Disabled,Invalid
```

Figure 7-65 Example of /ibm/xsm.log

7.4 Setting up an environment using FlashCopy

In this section we walk you through the steps to set up an environment that uses FlashCopy of an iASP together with PowerHA for i. For more information about FlashCopy see 4.4, “FlashCopy” on page 51. As many of the steps are similar to those shown in 7.2, “Setting up an environment using metro mirror” on page 164, we do not repeat all the steps in detail here. In our example, we have two nodes that are using metro mirror between them. We create a FlashCopy of the iASP of the backup node and then attach this new copy of the iASP to an additional partition to save its data to tape. You could also decide to create a FlashCopy from your production iASP.

If you want to use FlashCopy of an IAPS together with PowerHA for i, you again must make the system that uses the production copy of the iASP and the system that you want to attach the flashed copy of the iASP part of the same cluster. This setup is described in 7.2.3, “Preparing the scenario” on page 167. Once you have created the cluster and device domain

and added the nodes to the device domain, proceed with the steps outlined in the following section. Note that the node that you wish to attach the flashed copy of your iASP to is *not* a member of a device CRG.

Remember that you must install DSCLI on all the nodes that are part of your FlashCopy environment.

Create device description of iASP on node that houses copy

On the node that houses the flashed copy of your production iASP you must create the device description of iASP with the same device description name and resource name as are used for iASP on the production node. To do this:

1. Open the IBM i session on the backup node.
2. Type the command:

   ```
   CRTDEVASP
   ```

3. Press F4.
4. Insert the device description name and the resource name. Press Enter. I
5. f you want, type in the description of iASP and press Enter again.

Figure 7-66 shows an example of creating a device description.

```
                    Create Device Desc (ASP) (CRTDEVASP)

Type choices, press Enter.

Device description . . . . . > IASPMETRO     Name
Resource name  . . . . . . . > IASPMETRO     Name
Relational database  . . . . .   *GEN
Message queue  . . . . . . . .   *SYSOPR     Name
  Library  . . . . . . . . . .               Name, *LIBL, *CURLIB
Text 'description' . . . . . .   Decsription of IASP for FlashCopy

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

Figure 7-66 Create device description of an iASP

Note that at this time only the device description of the iASP is needed on the backup node, while the disk pool itself does not need to be created.

Create copy description for the iASPs on production and backup

As in the scenario described for metro mirror, we require an ASP session description, as well as a local and a remote ASP copy description. These are again created and assigned using IBM System Director Navigator. Make sure that you follow these steps from the target system of your FlashCopy environment (in our case, node5):

1. Choose **Configuration and Service**, then on the right pane select **Disk Pools** and then select the double arrows beside your iASP. Choose **Session** → **New** → **FlashCopy**, as shown in Figure 7-67.

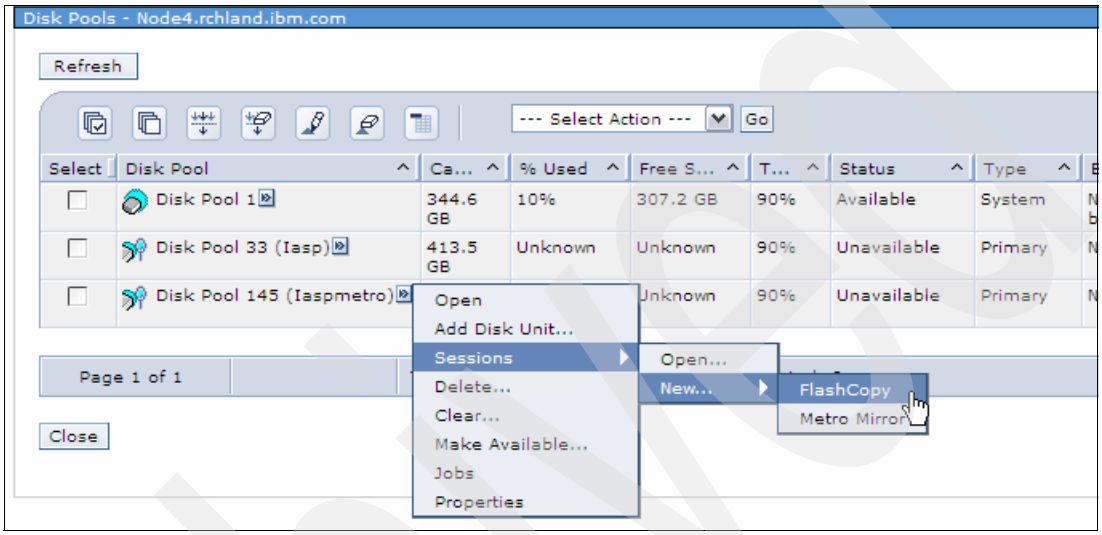


Figure 7-67 Create FlashCopy session

2. Make sure that the default settings for your FlashCopy Session correspond to your needs. These settings define how the actual FlashCopy process on your Storage system is implemented. Figure 7-68 show the defaults.

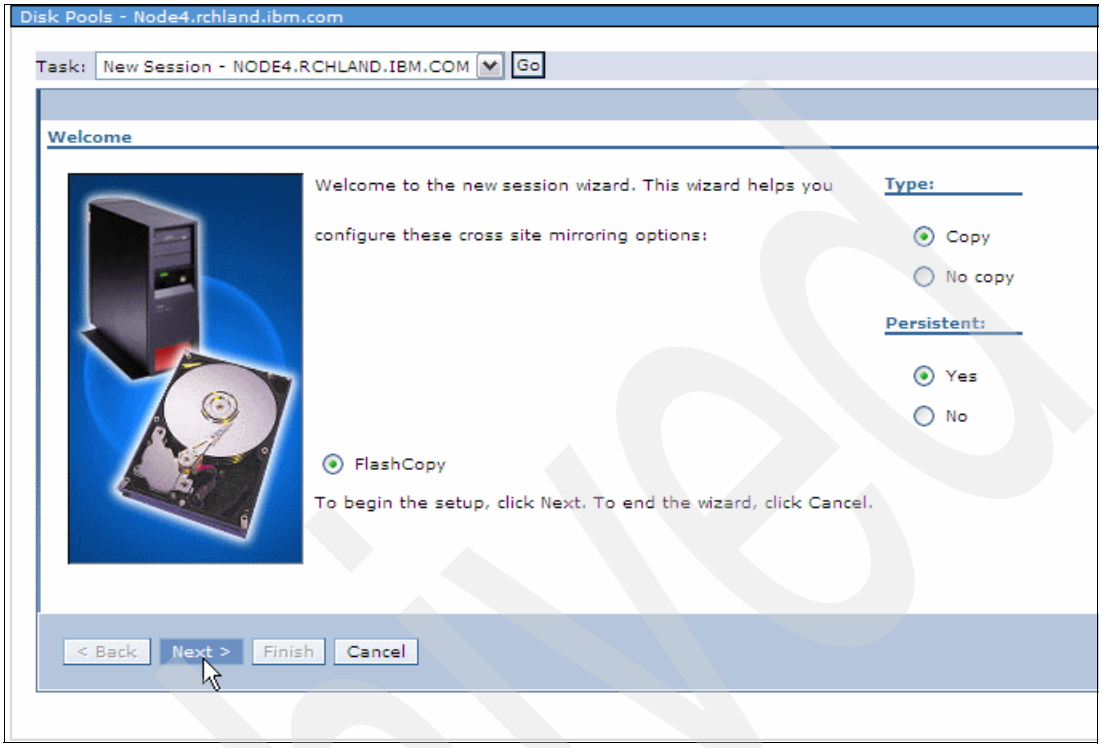


Figure 7-68 FlashCopy session defaults

3. In Figure 7-69 we have changed the defaults for type and persistent to the values that we want to work with. Click **Next** to continue.

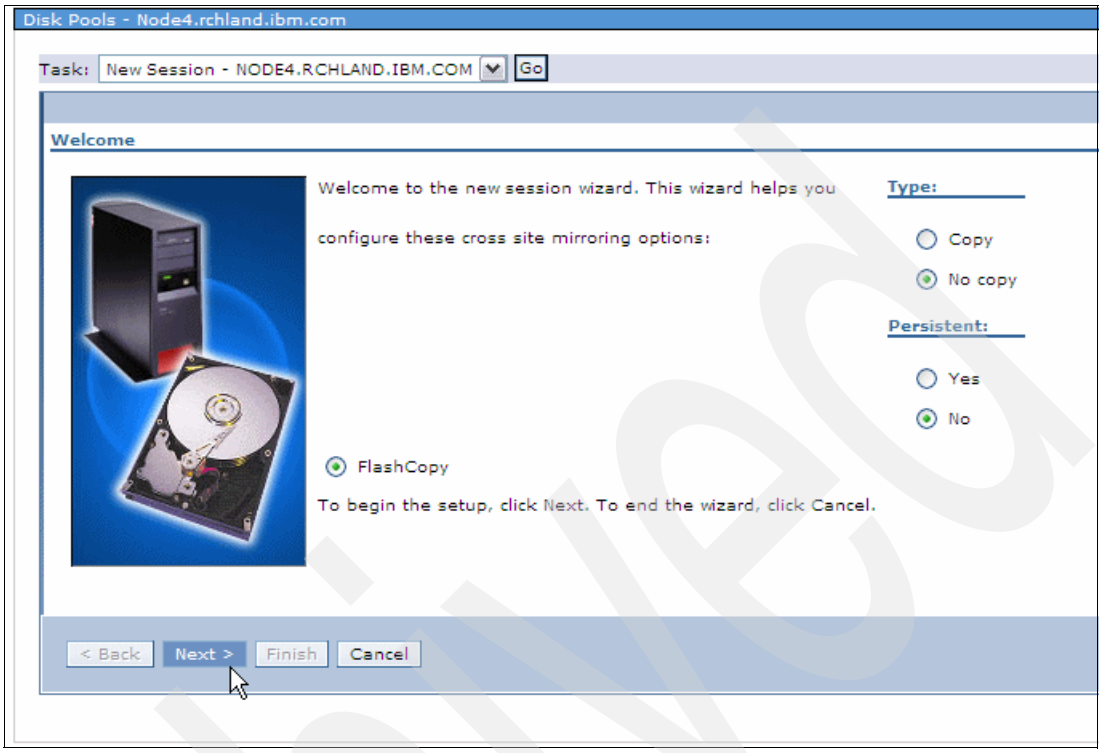


Figure 7-69 FlashCopy session defaults changed

4. Select an ASP copy description for the source copy of your iASP.

> **Attention:** The GUI here talks about the local and the remote copy. This does *not* pertain to the view from the node that you logged onto using the GUI. This pertains to the logical setup of your environment. In our example for a FlashCopy setup, the local copy is the FlashCopy *source* system and the remote system is the FlashCopy *target* system.

5. Note that one ASP copy description can be used in different session descriptions. As we have already created a copy description for the target of our metro mirror environment, we can use this copy description as the source copy description for our FlashCopy environment. The same would be true if you decided to take a FlashCopy from your production iASP. You could the use the copy description for the iASP on that node in your metro mirror session description as well as in your FlashCopy session description. We therefore simply choose the existing copy description METROLOC by checking the box next to it, as shown in Figure 7-70. Click **Next** to continue.

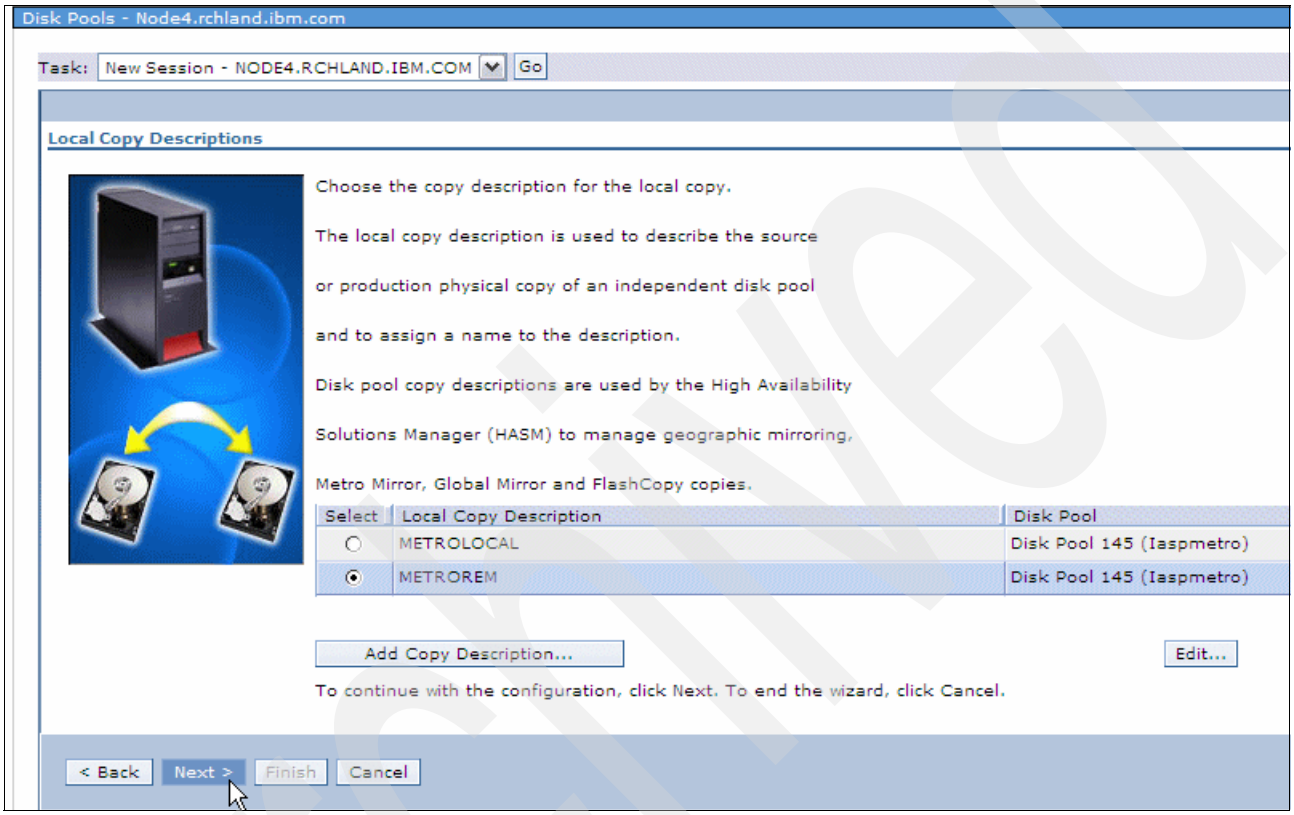


Figure 7-70 Select local ASP copy description

6. Create a copy description for the FlashCopy target.

> **Attention:** The GUI here talks about the local and the remote copy. This does *not* pertain to the view from the node that you logged onto using the GUI. This pertains to the logical setup of your environment. In our example for a FlashCopy setup, the local copy is the FlashCopy *source* system. The remote system is the FlashCopy *target* system.

This one does not already exist. Therefore, choose **Add Copy Description**, as shown in Figure 7-71.

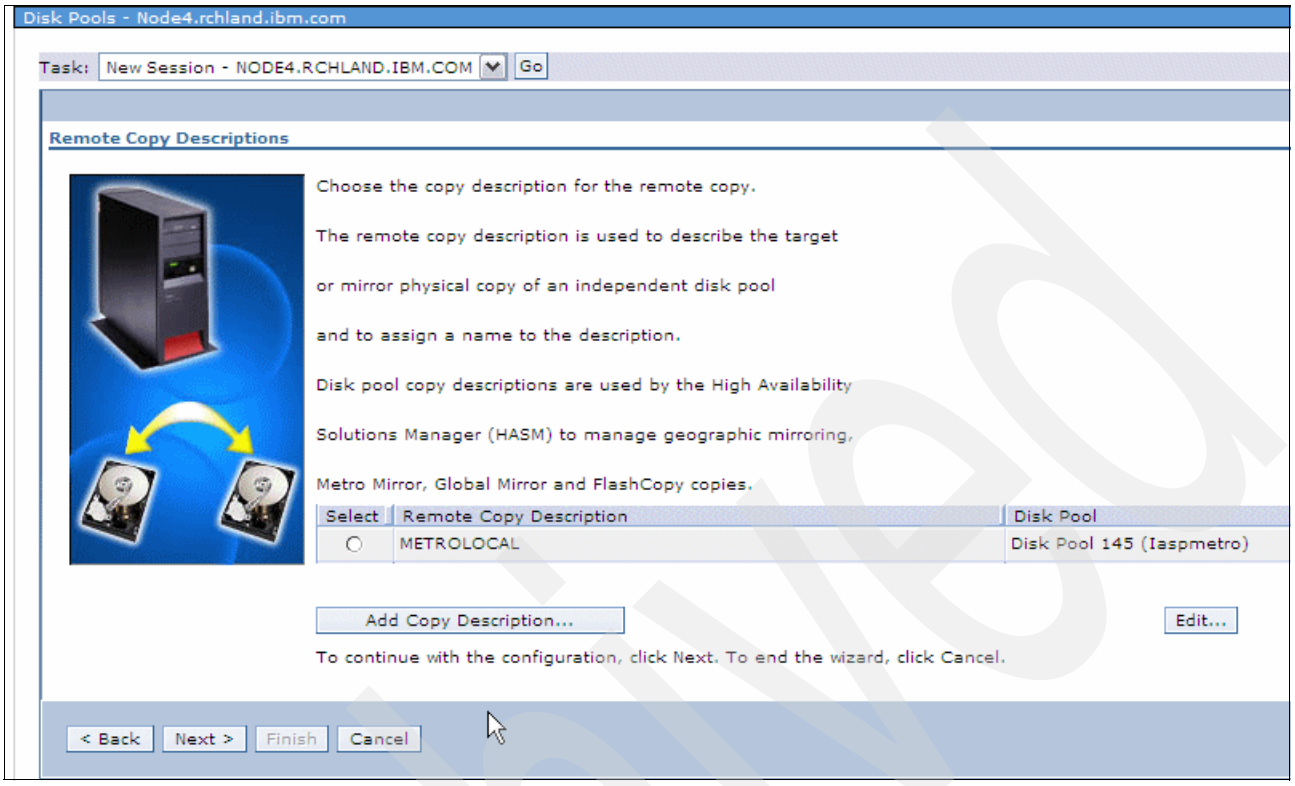


Figure 7-71 Add copy description for FlashCopy target

7. You now must fill in the details for this new copy description, as shown in Figure 7-72. Note that for the FlashCopy target copy description you do *not* specify a device CRG or a site name. That information is not required for a FlashCopy session. To fill in the required information about your Storage system click **Add**, as shown in Figure 7-72.

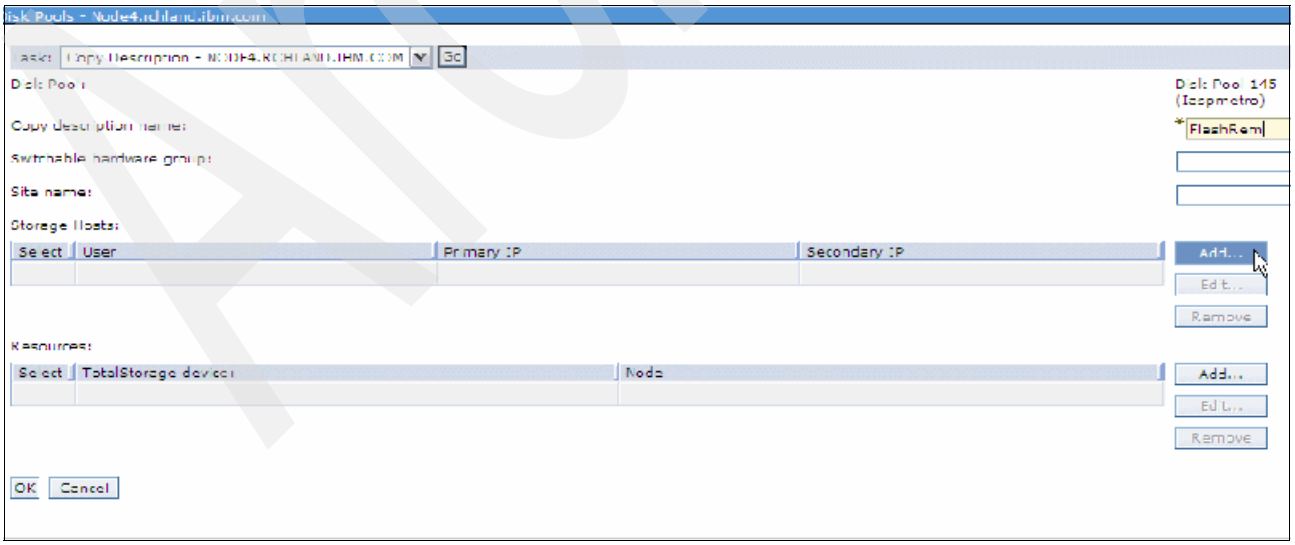


Figure 7-72 Details for FlashCopy target copy description

8. You then must enter a user ID and password, as well as an IP address to access your Storage system, as shown in Figure 7-73. Clicking **OK** brings you back to the window shown in Figure 7-72 on page 202. There select **Add** on the right-hand side of Resources.

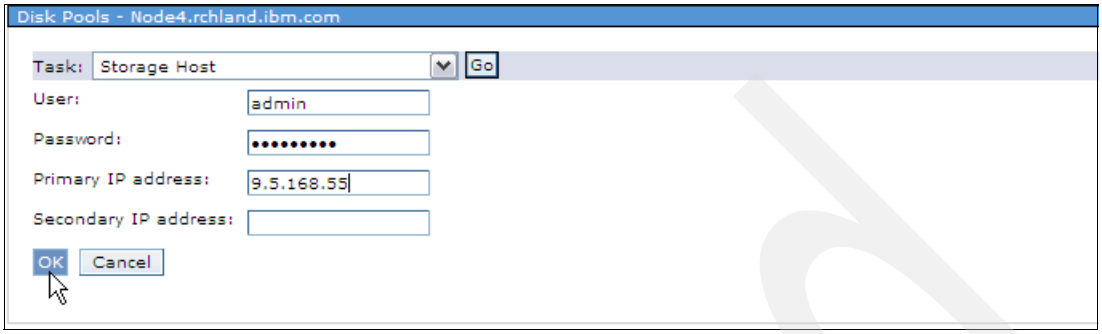


Figure 7-73 Storage host details

9. The previous step opens the window shown in Figure 7-74. To add the LUN ranges of the FlashCopy target select **Add**.

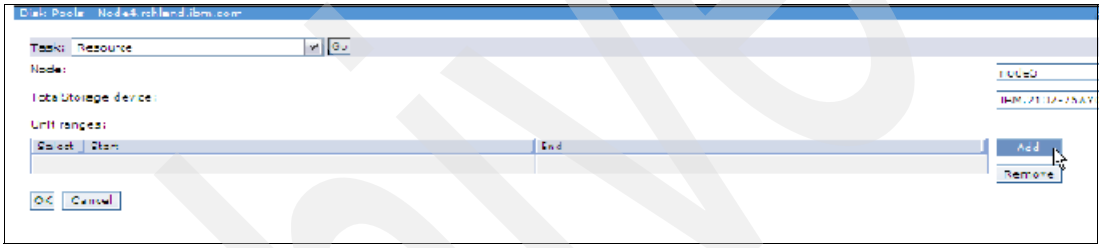


Figure 7-74 Add resources

10. Once you have entered all LUN ranges that the FlashCopy target consists of click **OK**, as shown in Figure 7-75.

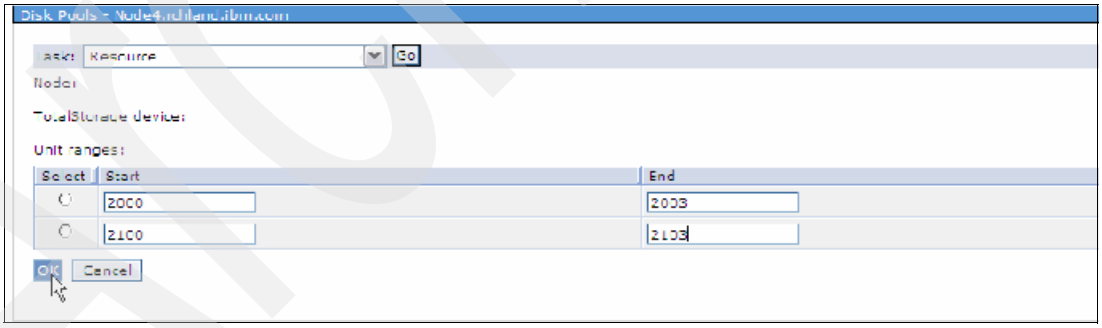


Figure 7-75 Resources added

11. On the overview of the copy description check again that all data was entered correctly and click **OK**, as shown in Figure 7-76.

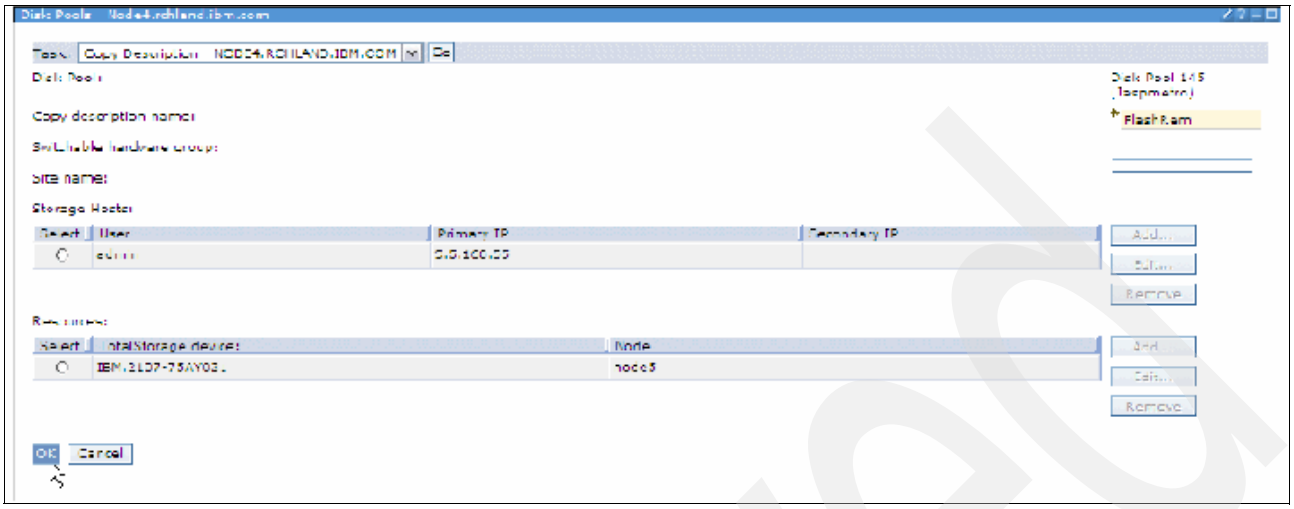


Figure 7-76 Remote copy description overview

12. On the next panel make sure to select the newly created copy description by checking the box next to it, as shown in Figure 7-77. Click **Next** to continue.

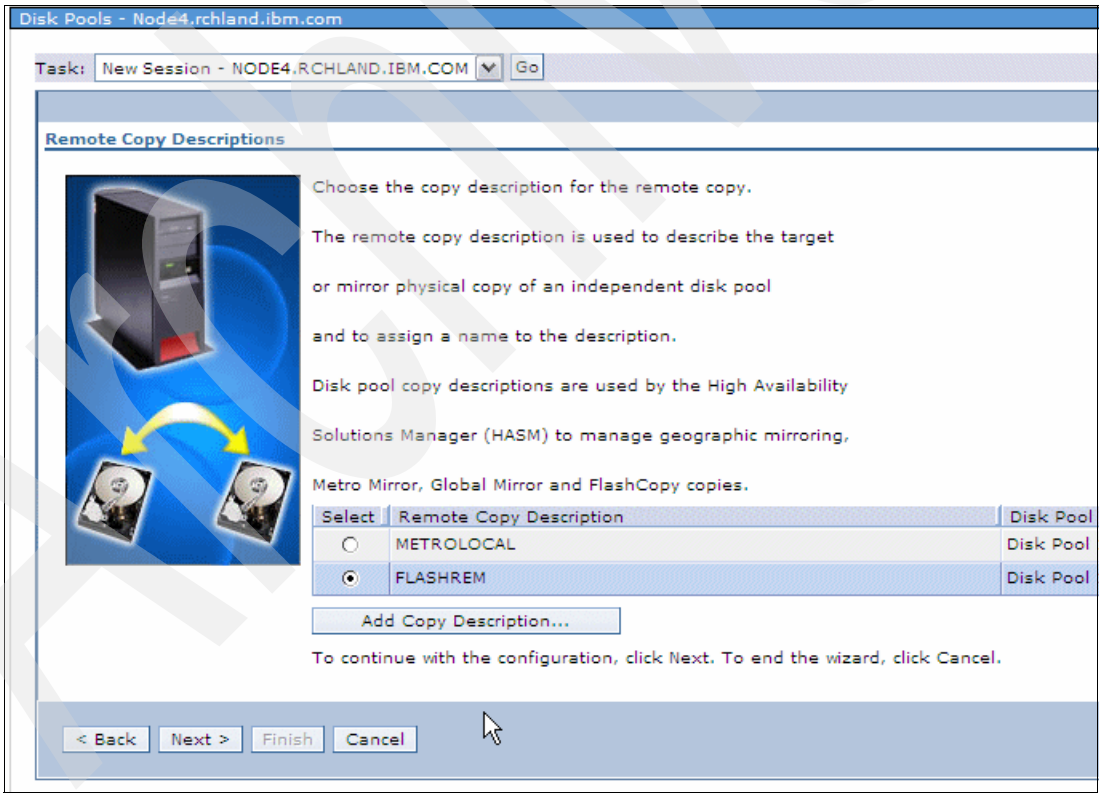


Figure 7-77 Select target copy description

13. You now must provide a name for the new session description, as shown in Figure 7-78. Click **Next** to continue again.

Figure 7-78 Name of session description

14. In the last step you are presented with an overview of the session descriptions. Make sure the everything is correct and click **Finish**, as shown in Figure 7-79.

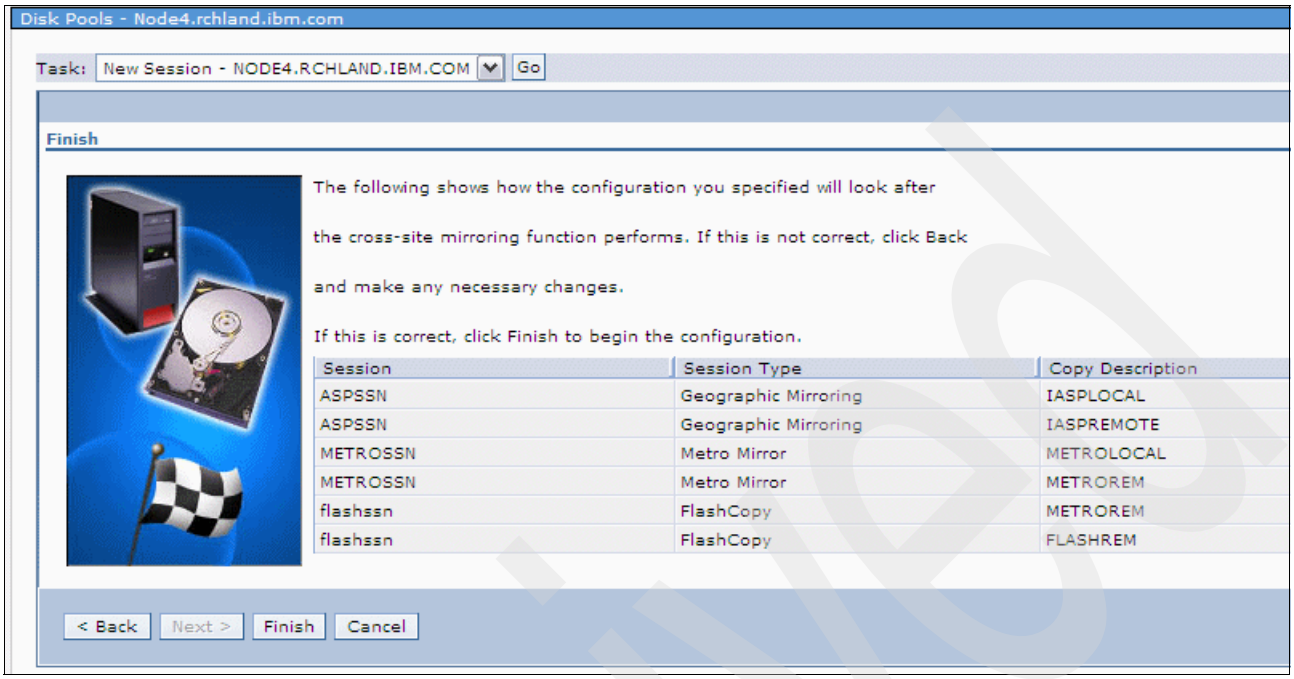


Figure 7-79 New session creation: Overview

15. Once you click the **Finish** button, the session gets created and is also automatically started. Note that this start of a FlashCopy session also starts the FlashCopy process on your Storage system. Be aware that you must manually vary on the target iASP. This is not done automatically by the start of your ASP session. You can either use the GUI interface or a 5250 session to do so.

7.5 Working with the FlashCopy environment

This section shows you how to work with your FlashCopy environment.

Attention: Do *not* try to vary on the flashed copy of an iASP after you have ended the ASPSSN that belongs to this iASP. Once the ASPSSN is ended the FlashCopy relationship between source and target is also ended. Unless you were doing a FlashCopy *COPY and that background copy has completely finished, the iASP on the target side does not really contain all its data. Most or part of the data in reality was being read from the source side. If only one page was not really copied over to the target system, then the copy of the iASP is *unusable* as soon as the ASP session is ended (using the command ENDASPSSN or deleting the session using the GUI interface). Varying on this iASP leads to unpredictable results.

Quiesce the application data

If you are not able to vary off your production iASP before taking a FlashCopy, we recommend that you at least quiesce the application data in your iASP. This ensures that database activity, as well as activity in the IFS, get suspended. This suspend can be done by issuing the command CHGASPACT with parameter *SUSPEND. Doing so will achieve the following things:

- Most data is flushed from main memory to the iASP disks.
- Database and IFS activities are suspended.
- Other activities may continue to occur.

This will lead to a nicer abnormal vary-on of the flashed copy with a shorter recovery time. Most database activity should be at commit boundaries, thus eliminating the need for extensive rollbacks.

Note that you can specify a suspend time out. This parameter determines how long the suspend operation itself is allowed to run. The Suspend Timeout Action parameter then determines what should happen after the time-out has been reached. The suspend state can either continue (default behavior) or be automatically ended. If a time out occurs, a spoolfile is generated that contains information about which transactions were still outstanding.

Note also that as long as the ASP is in suspended status, new database or IFS transactions are not allowed to start. This means that your applications are waiting until the command CHGASPACT *RESUME is issued.

The order of the steps you should take (if your application environment can afford them) is:

1. If you still have an existing ASP session description for your FlashCopy environment, you must end it using ENDASPSSN.
2. Suspend access to your production iASP using CHGGASPACT *SUSPEND.
3. Start the FlashCopy session and thereby also the FlashCopy itself using STRASPSSN *FLASHCOPY.
4. Once the STRASPSSN command is finished, resume access to your production iASP using CHGASPACT *RESUME.
5. Vary on the flashed copy of the iASP to start using it.

7.6 Other functions of the CRS GUI

The Cluster Resource Services GUI interface allows you to configure and manage clustered resources and environments. Unlike the HASM GUI discussed in Chapter 6, “High Availability Solutions Manager GUI” on page 89, this cluster interface is based on task-oriented goals. This interface allows you to:

- Create and manage a cluster.
- Create and manage cluster nodes.
- Create and manage cluster resource groups (CRGs).
- Create and manage cluster administrative domains.
- 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.

Everything related to your ASP session description (such as, suspend, resume, detach, reattach, create, or delete) is done from the disk GUI of IBM System Director Navigator.

7.6.1 Cluster Resource Services GUI

When you access IBM System Director Navigator and access cluster resource services, you are presented with the window shown in Figure 7-80.

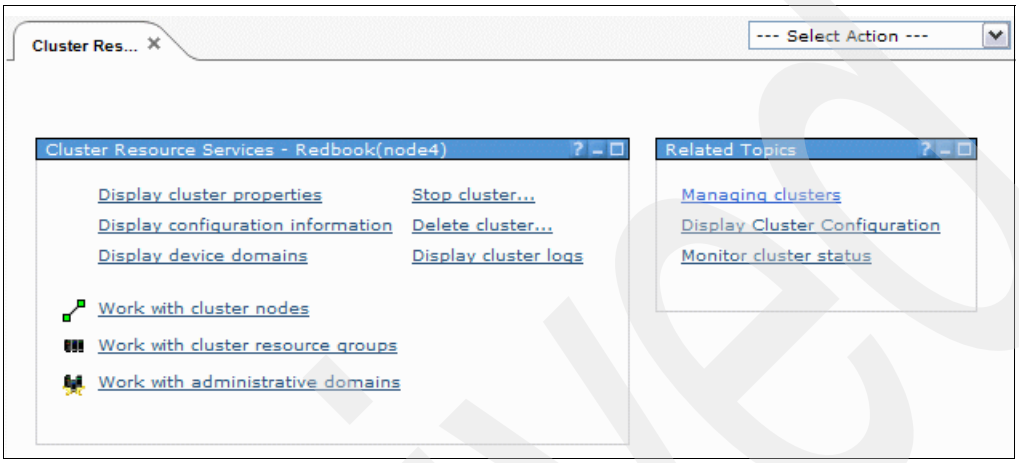


Figure 7-80 First cluster panel

From here, you can start the following actions:

- Display cluster properties.
- Display configuration information.
- Display device domains.
- Work with cluster nodes.
- Work with cluster resource groups.
- Work with administrative domain.
- Stop cluster.
- Delete cluster.
- Display cluster logs.

In the following sections you can find an overview of the functionalities that you can find in these areas.

7.6.2 Cluster nodes

When you select **Work with cluster nodes**, as shown in Figure 7-80 on page 208, you will come to the panel illustrated in Figure 7-81. This shows you what node names there are, what IP addresses they use for cluster heartbeating, and their current status.

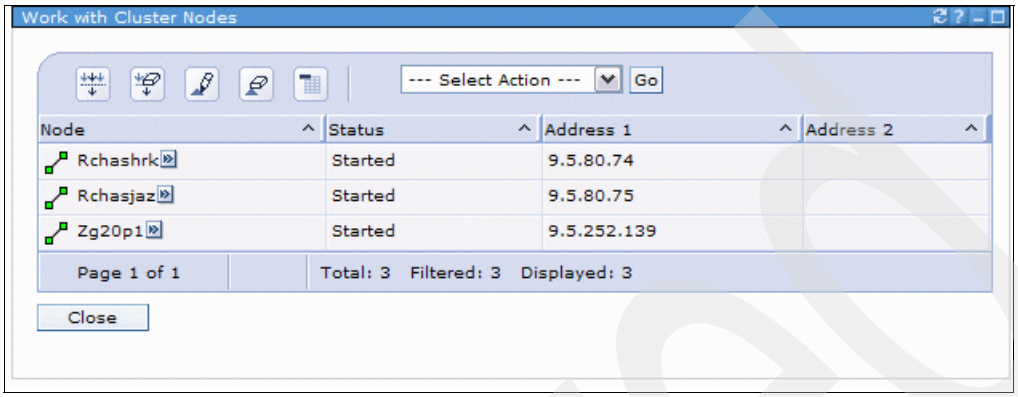


Figure 7-81 Work with cluster nodes

Stop a cluster node

From the panel shown in Figure 7-82, you can choose a node. Select the double arrow to the right of it, illustrated by the green circle in Figure 7-82, then choose the **Stop** option, shown by the red circle.

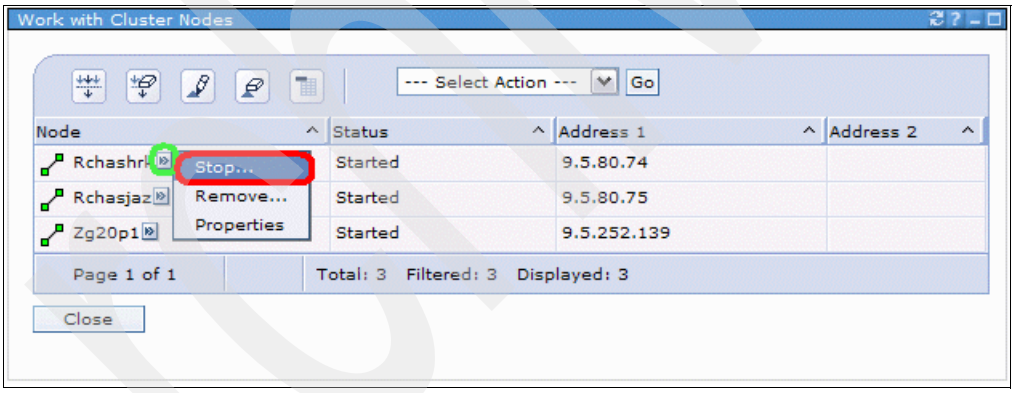


Figure 7-82 Cluster node options

If the cluster node is stopped when you select the double arrow, then this window gives you the opportunity to start the cluster node.

The remove option lets you remove a node from your cluster. Properties shows you the node properties (device domain that this node belongs to, potential cluster version, IP addresses used for heartbeating) and also allows you to change the heartbeat IP addresses by simply typing over them.

7.6.3 Work with cluster resource groups

The Work with cluster resource group option provides you with the tasks needed in the CRG area. These consist of starting and stopping the CRG, performing a manual switchover, adding new disk pools or new switchable devices to the CRG, as well as looking at and

changing CRG properties like the failover message queue, the recovery domain, or exit programs.

Some of these options are available only when the CRG is in a certain status, for example, started or stopped. To access the options, select the double arrow beside your CRG, as shown in Figure 7-83

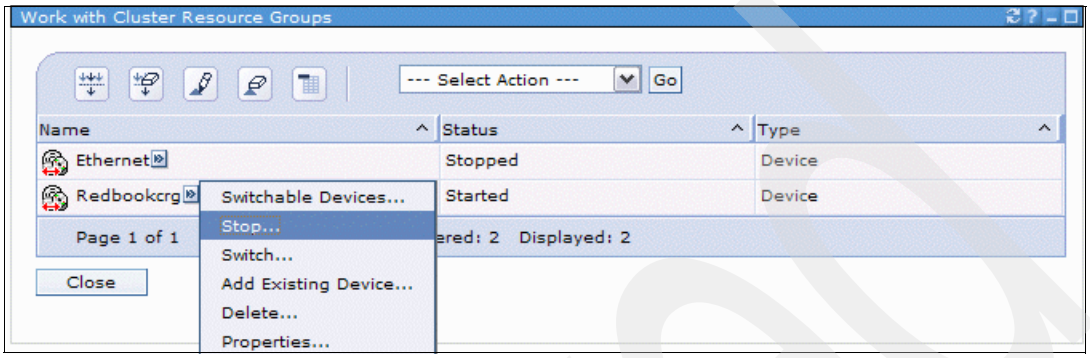


Figure 7-83 CRG options

7.6.4 Administrative domains

The cluster administrative domain, represented by a peer cluster resource group object on IBM i operating system, allows the IBM i Administrator to manage replication of the objects that are part of their IBM i High Availability solution, but are not able to be stored on an independent auxiliary storage pool. Table 7-1 describes all the resources and attributes that can be monitored inside a cluster administrative domain with IBM i 6.1.

Table 7-1 Monitored resource entry type supported with HASM on IBM i 6.1

Object or attribute description	Type
Device description for an auxiliary storage pool device	*ASPDEV
Class description (*)	*CLS
System environment variable (*)	*ENVVAR
Line description for an Ethernet line	*ETHLIN
Job description (*)	*JOBD
Network attribute (*)	*NETA
Network server configuration	*NWSCFG
Network server description	*NWSD
Device description for a network server host adapter	*NWSHDEV
Network server storage space	*NWSSTG
Device description for a optical device	*OPTDEV
Subsystem description	*SBSD
System value (*)	*SYSVAL
Device description for a tape device	*TAPDEV
TCP/IP attribute (*)	*TCPA

Object or attribute description	Type
Line description for an token-ring network line	*TRNLIN
User profile (*)	*USRPRF

(*) Available with IBM i 5.4

To manage the cluster administrative domain and the monitored resources with the new GUI interface delivered by the license product PowerHA for i in IBM Systems Director for IBM i:

1. When connected to the console select **Cluster Resource Service** (in the red circle in Figure 7-84).
2. Select **Work with Administrative Domains** (the green circle in Figure 7-84).

Figure 7-84 Work with cluster administrative domain

3. If the user ID that you use to log in to the IBM Systems Director for i5/OS console does not exist on the remote node, you can use this panel to log into the remote system that is defined as another node in the cluster (Figure 7-85).

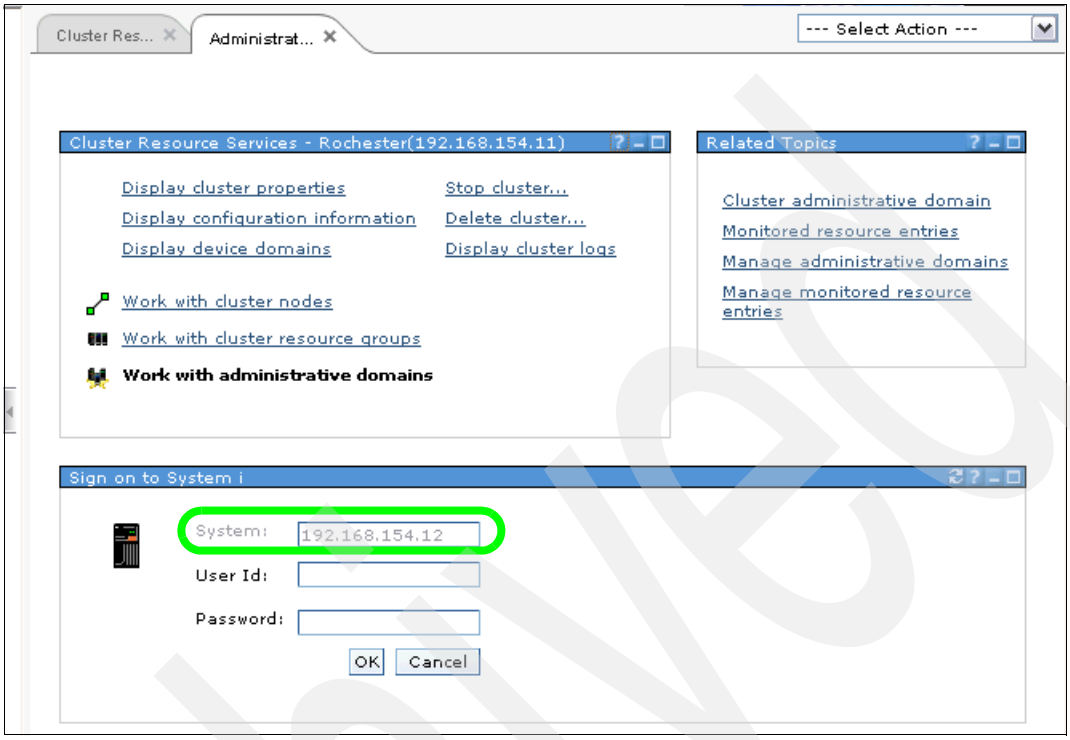


Figure 7-85 Sign-on window

4. After you log in to the second node using a valid user profile and password (or if the user profile that you use already exists on the other node) you will receive the next panel (Figure 7-86), allowing you to manage your cluster administrative domain.

   By selecting the appropriate action from the list (as shown in Figure 7-86), you will be able to create a new cluster administrative domain. Select this option and click **GO** on the right.

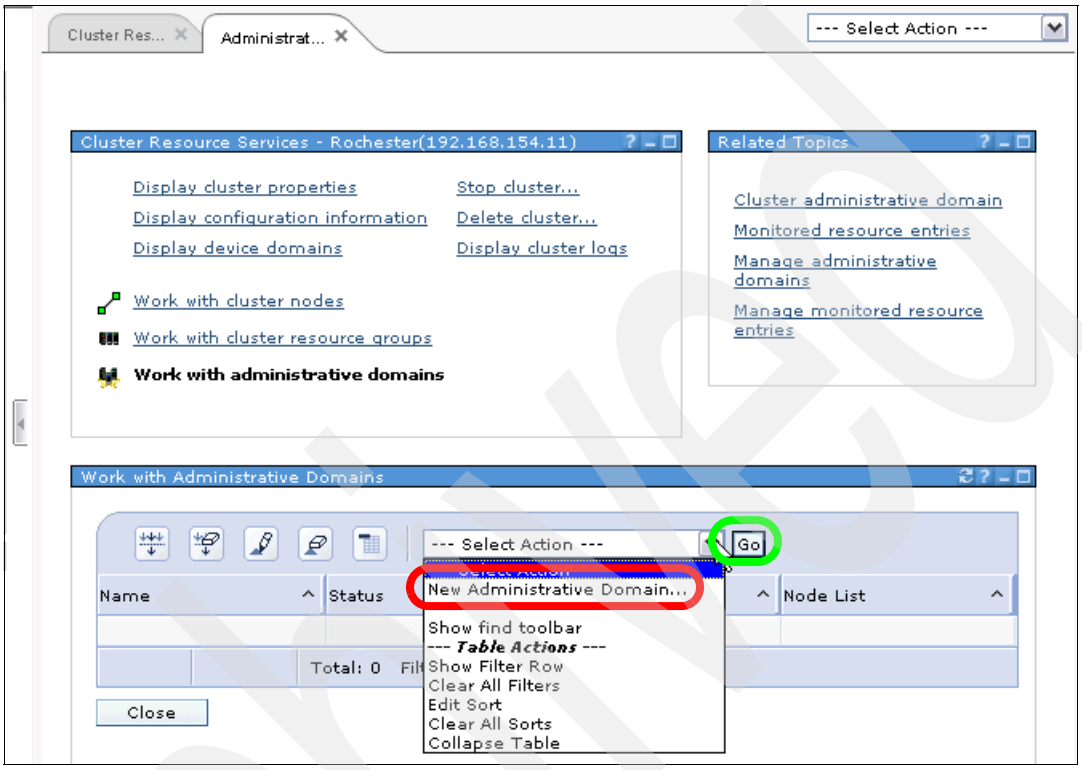


Figure 7-86 Create a new cluster administrative domain

5. You are prompted on the next panel to enter the name of your new cluster administrative domain and other attributes and parameters of the cluster administrative domain, as shown in Figure 7-87.

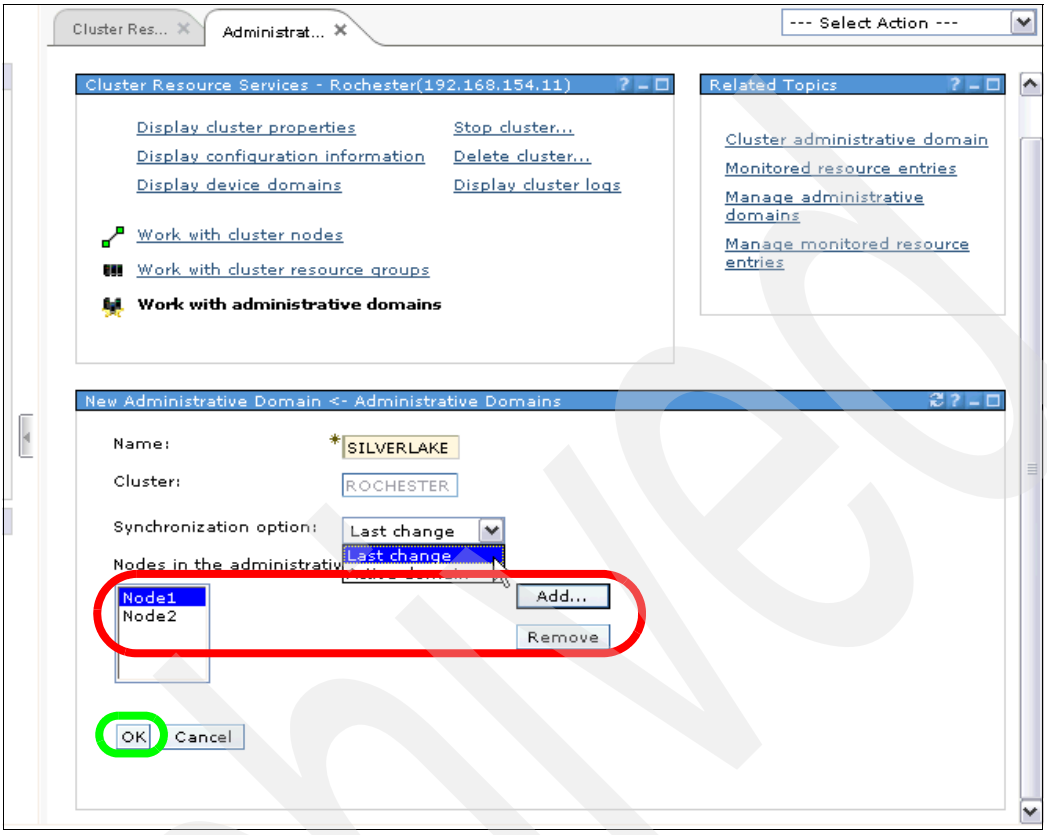


Figure 7-87 Enter cluster administrative domain parameters and attributes

The synchronization option allows you to specify:

– Last change

  The last change that was made before the node joined the cluster administrative domain is processed by all nodes in the active domain. The last change could have been made in the active domain or on the joining node while it was inactive.

– Active domain

  Only changes made on active nodes in an active cluster administrative domain are processed. Changes made on a node while it was inactive are not passed to the active domain. When a node joins the cluster administrative domain, it will be synchronized with the values from the active domain.

You can also add other nodes to this new cluster administrative domain.

6. Click **OK** (green circle in Figure 7-87 on page 214) to create your new cluster administrative domain. Then you get the panel shown in Figure 7-88 with the list of your cluster administrative domains that are defined and can be managed.
7. By selecting the double arrow (green circle in Figure 7-88) to the right of the cluster administrative name that you want to manage, you can select the appropriate action that you want to perform. If you cannot display the panel above, you need to close this window, then come back to it by selecting the **Work with administrative domains** option (red circle in Figure 7-88).

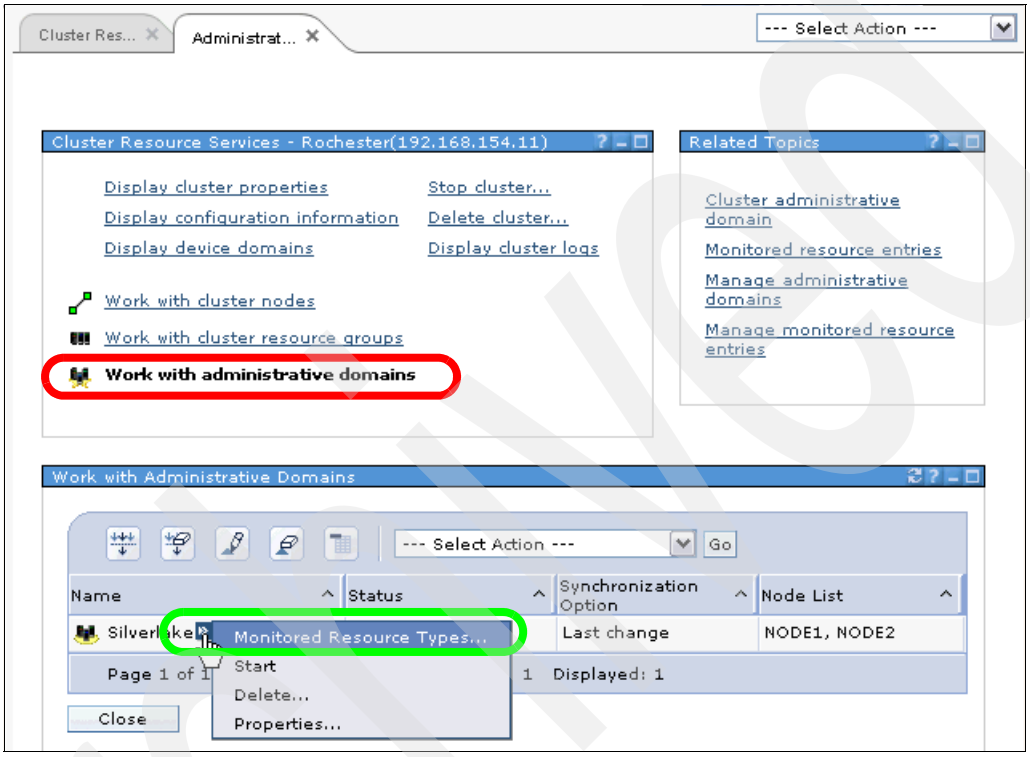


Figure 7-88 Manage a cluster administrative domain

Add monitored resource entries

To add monitored resource entries:

1. If you select the option **Monitored Resource type** (green circle in Figure 7-88 on page 215), you will get the following panel (Figure 7-89) that allows you to choose which types of resources you want to monitor in your high availability environment. For instance, to add a user profile to your cluster administrative domain, simply select the double arrow on the right of the User Profile option**,** as shown by the red circle in Figure 7-89, then select **Add Monitored Resource Entry**, as shown by the green circle.

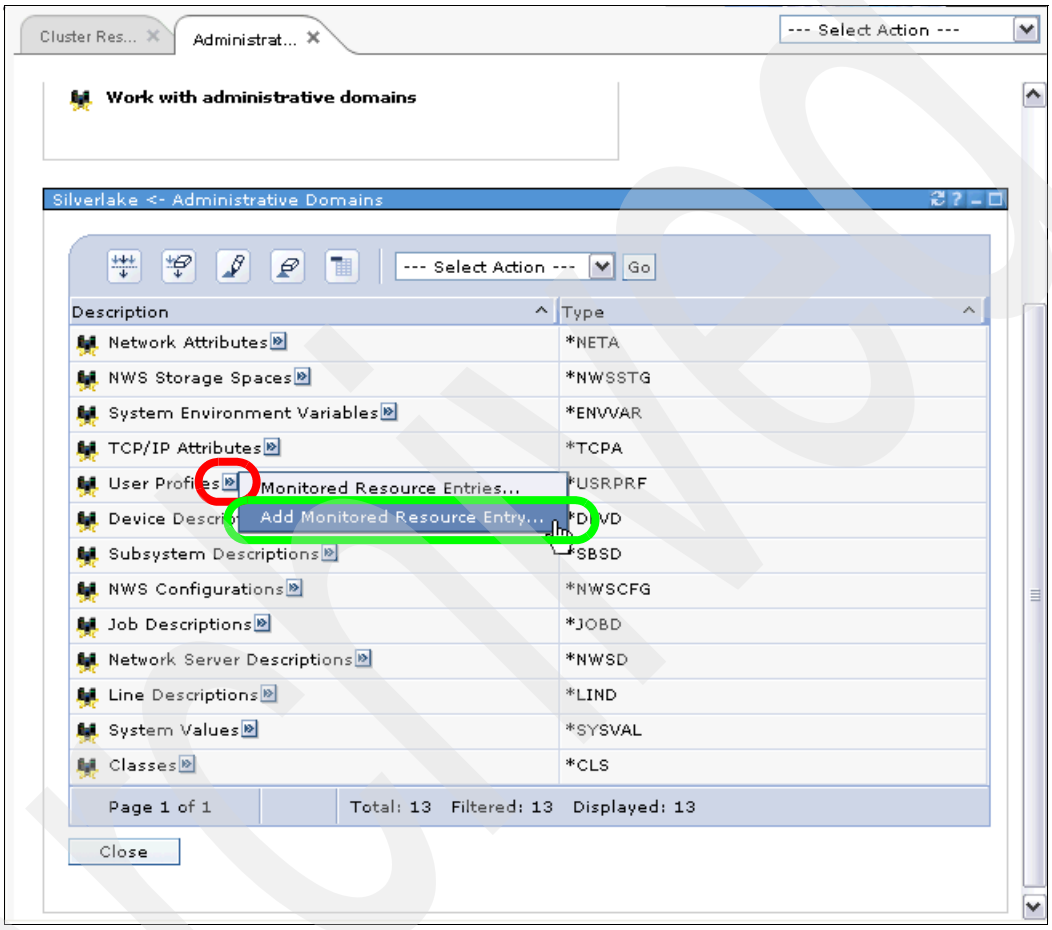


Figure 7-89 Manage monitored resource entries in the cluster administrative domain

You will get the panel shown in Figure 7-90 that allows you enter all the user profiles (red circle) that you must keep synchronized between all the nodes in your cluster resource group.

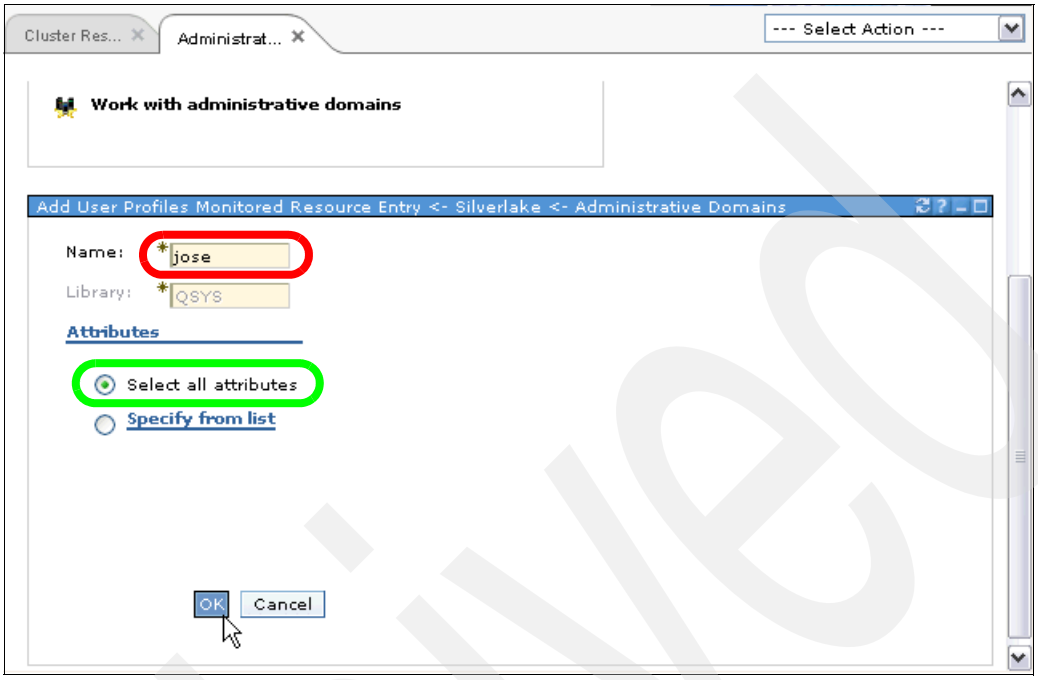


Figure 7-90 Add a user profile to the monitored resource entries

2. You can either maintain all attributes of the user profile by keeping the default option Select all attributes or you can select individual user profile attributes according to your needs. See Figure 7-91 for an example.
3. When you add a user profile to the cluster administrative domain that does not exist on the other nodes, the system creates it for you on all the other nodes.

Important: With independent auxiliary storage pools it is very important to synchronize the user profiles between all nodes involved in the cluster resource group representing the iASP. Otherwise, the vary-on process can take a lot of extra time.

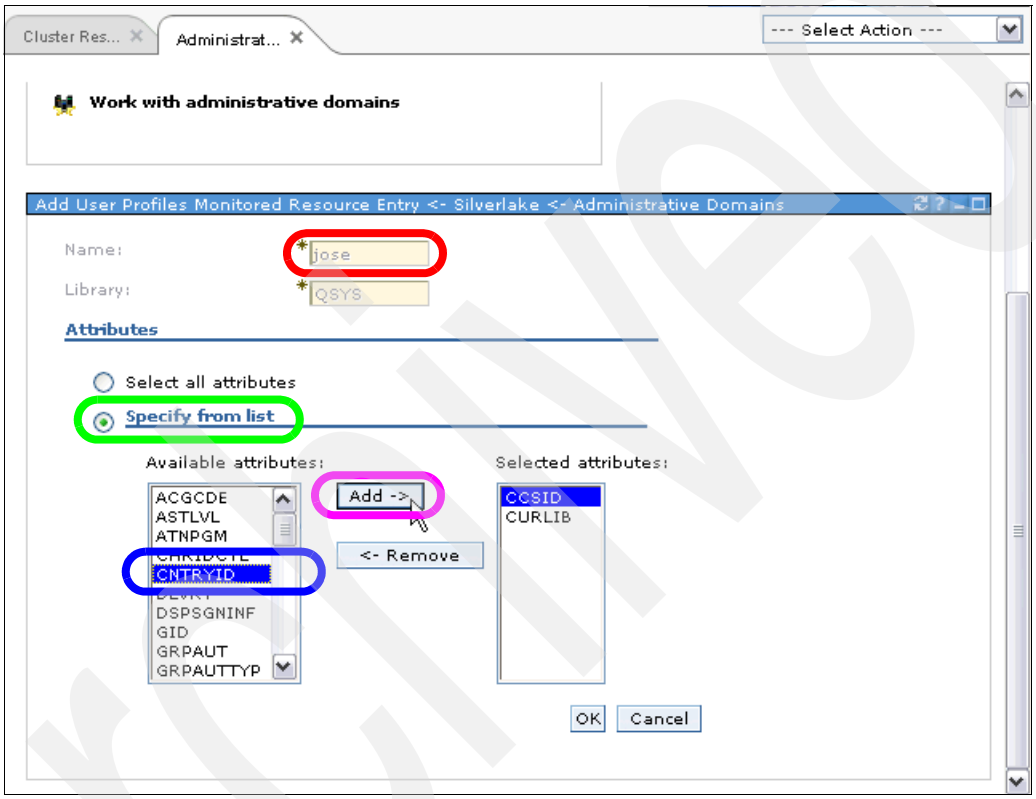


Figure 7-91 Enter monitored resource entry attributes

Note: From the cluster resource service GUI you can simply manage your monitored resources that will be replicated to all nodes that are part of the cluster administrative domain.

To help you in the management of user profiles or other resources, you can also write your own program using the CL commands to add all your existing or newly created user profiles to the cluster administrative domain automatically.

When you want to add a resource, the objects representing this resource in the IBM i operating system must exist. The object must already be created on the node that you use to manage your cluster administrative domain. Otherwise, you will get an error message, as shown in Figure 7-92.

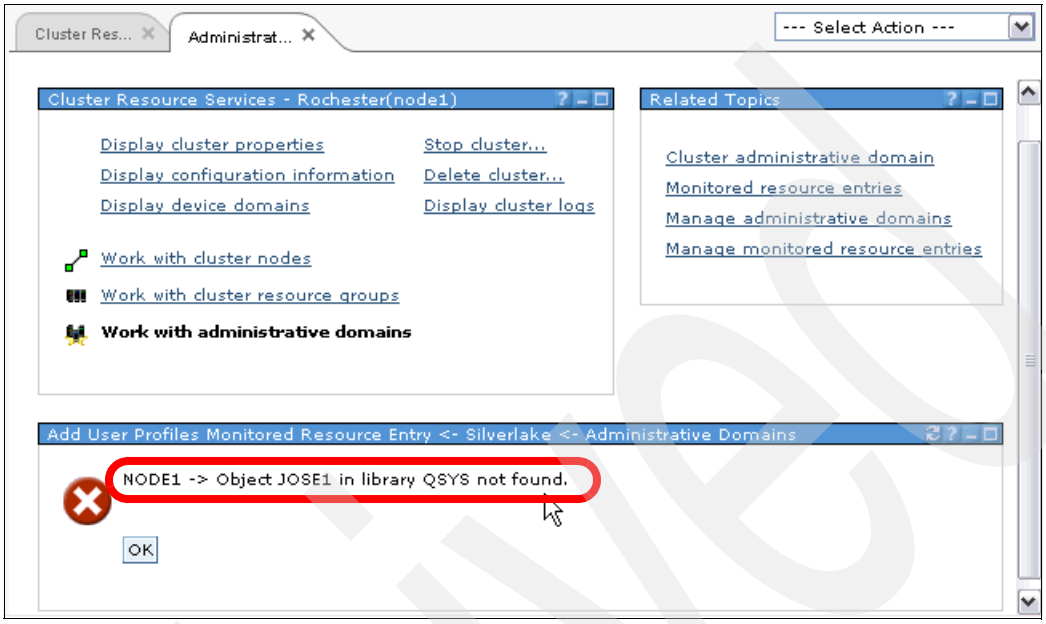


Figure 7-92 Object does not exist error

4. Selecting **OK** takes you back to the panel to add the monitored resource, as shown in Figure 7-93.

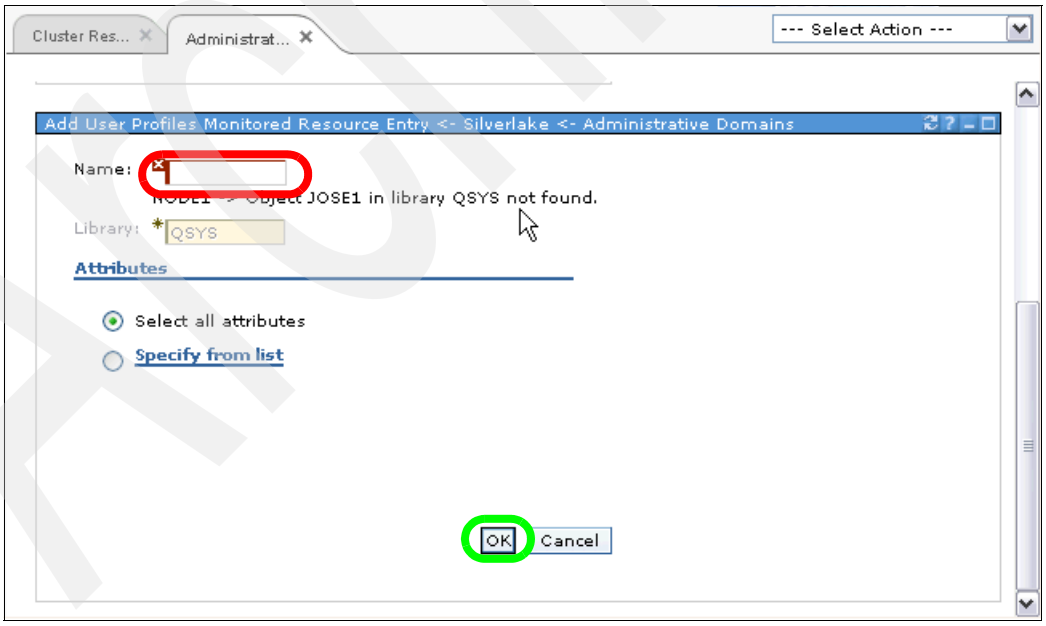


Figure 7-93 Error message if user profile does not exist

5. You can also manage system values, system environment variables, network attributes, and TCP/IP attributes, with the possibility either to select all values (or attributes) (as shown in Figure 7-94) or select specific systems values from a list (as shown in Figure 7-95.)

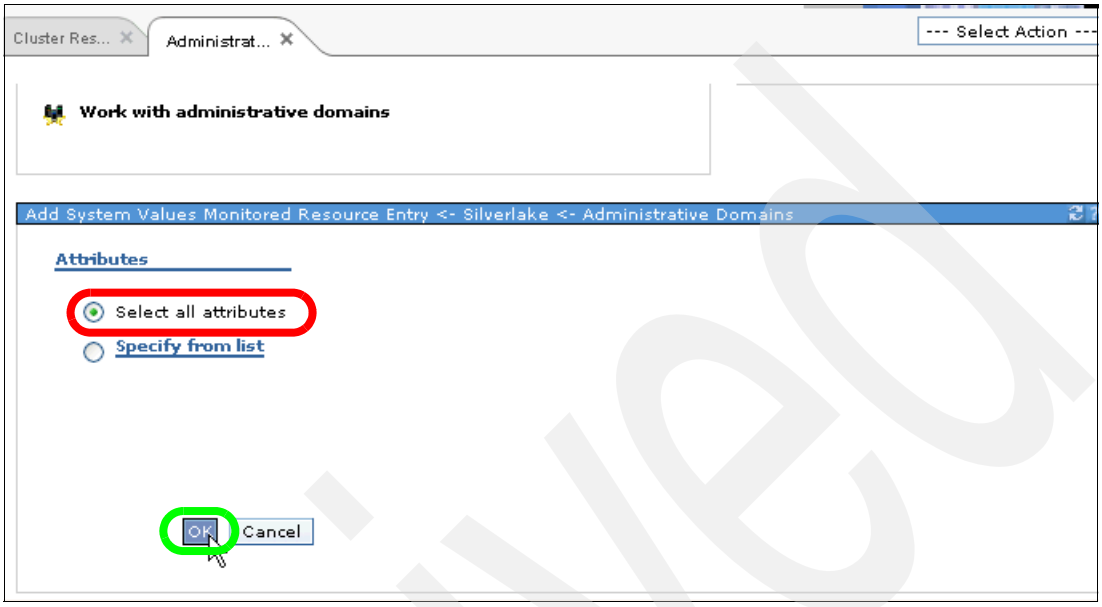


Figure 7-94 Select all system values

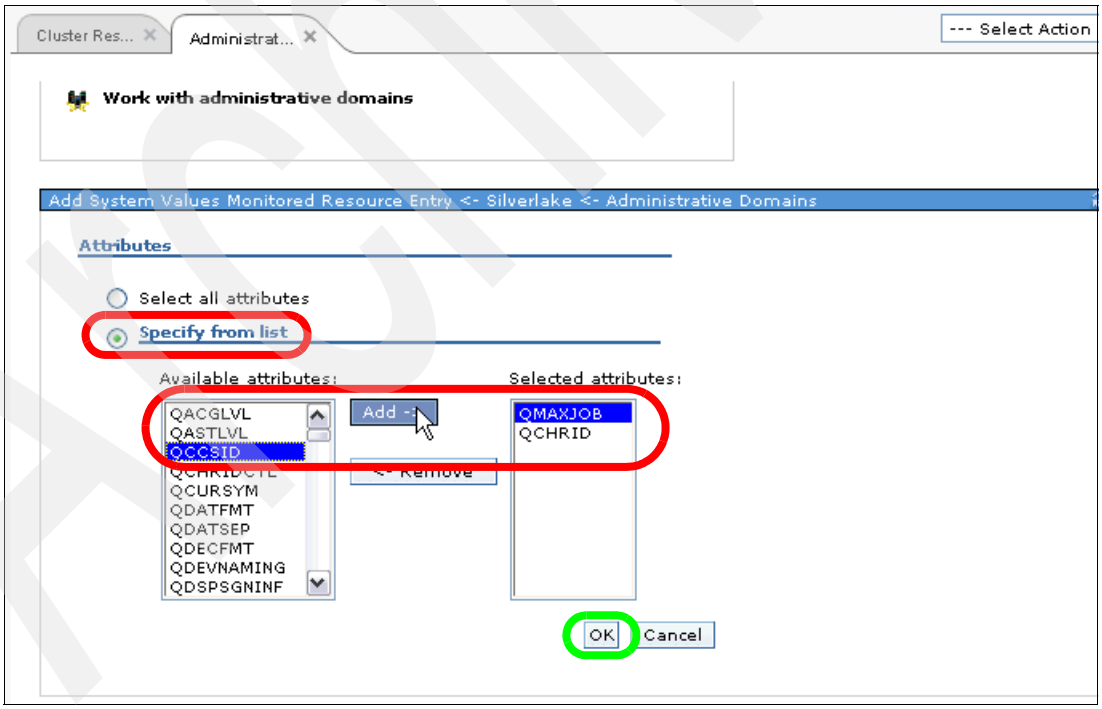


Figure 7-95 Specify individual system values to be monitored

The following device descriptions can also be monitored in the cluster administrative domain to help you by automatically synchronizing them in your high availability environment:

- Auxiliary storage pool
- Network server host adaptor
- Optical devices
- Tape devices

See Figure 7-96.

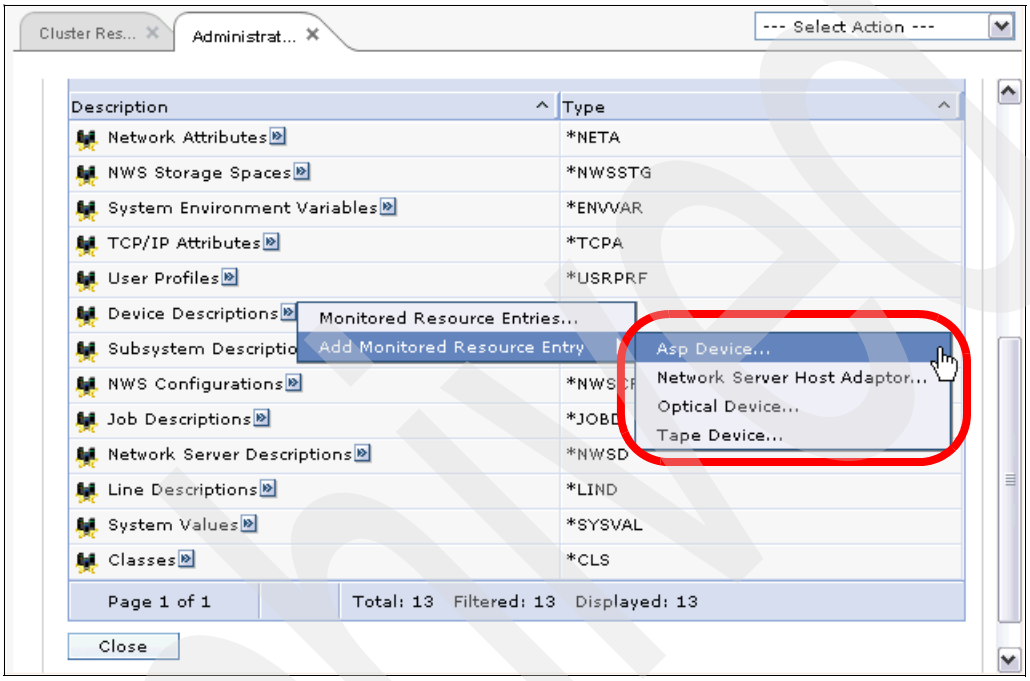


Figure 7-96 Add device description entries

Ethernet or token-ring line descriptions that help you to maintain synchronization between nodes for your network configuration can also be monitored, as shown in Figure 7-97.

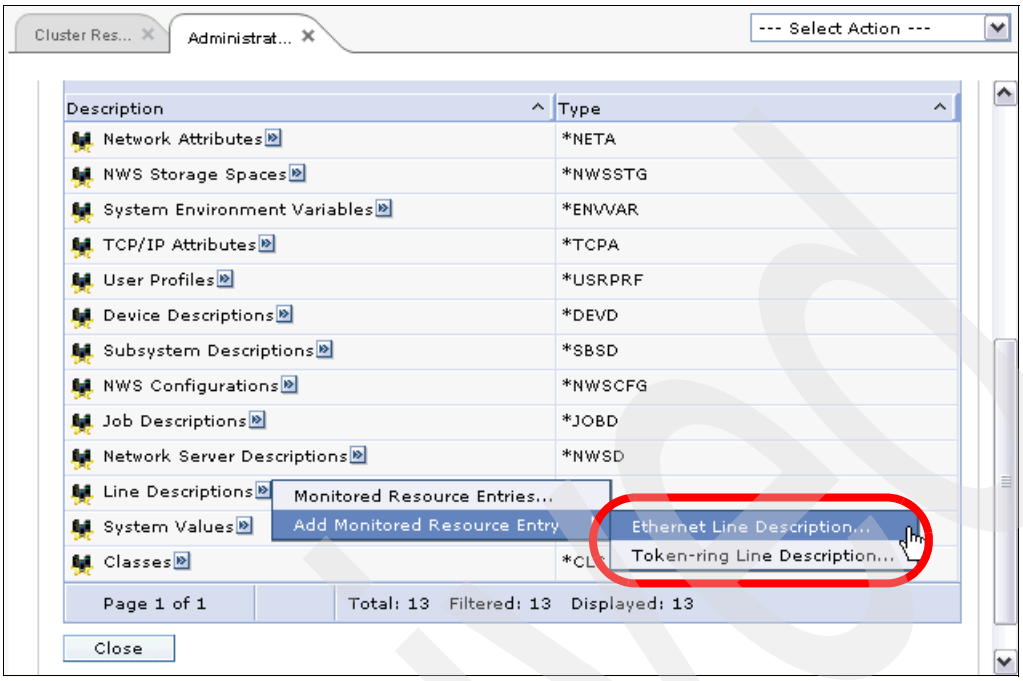


Figure 7-97 Add communication resource entries

Managing monitored resource entries

Once you have added the monitored resources to the cluster admin domain you can display, change, or remove them from the Work with administrative domains panel:

1. Select the double arrow and then the **Monitored Resources Entries** option, as shown in Figure 7-98.

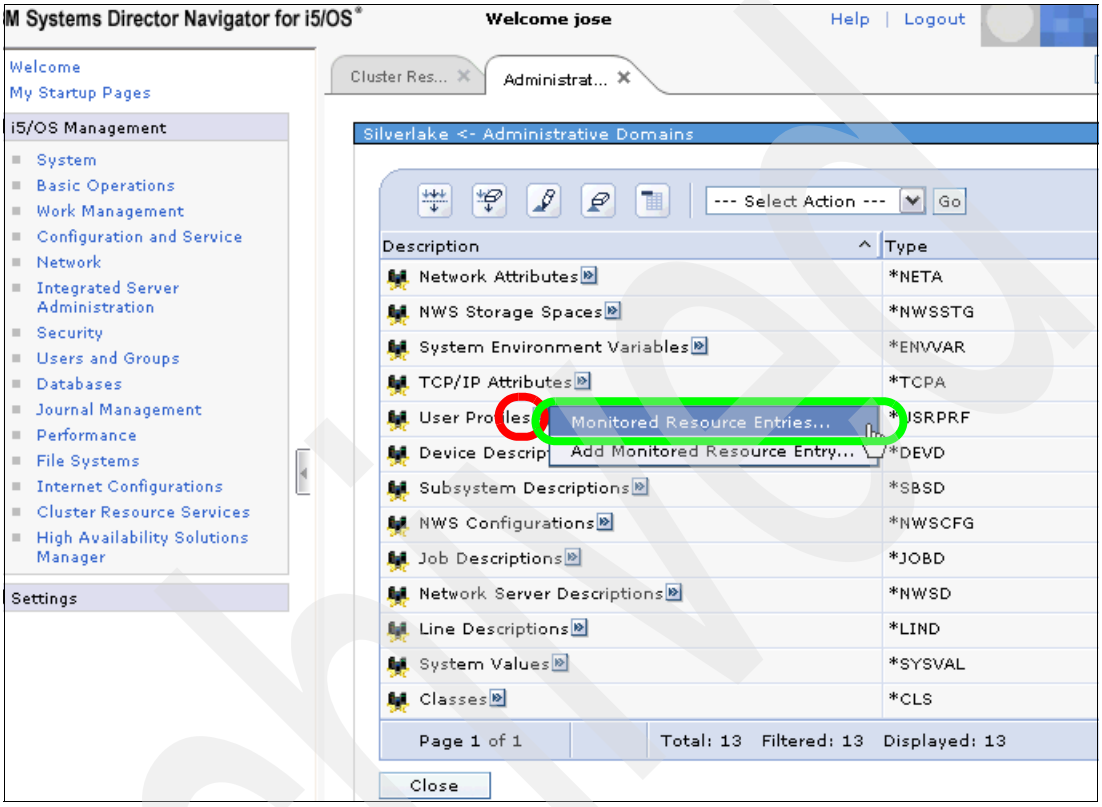


Figure 7-98 Manage monitored resource entries

2. In the window shown in Figure 7-99 you are be able to manage the monitored resource entries already added for the selected type by displaying the attributes or the messages, or by removing this monitored resource entry.

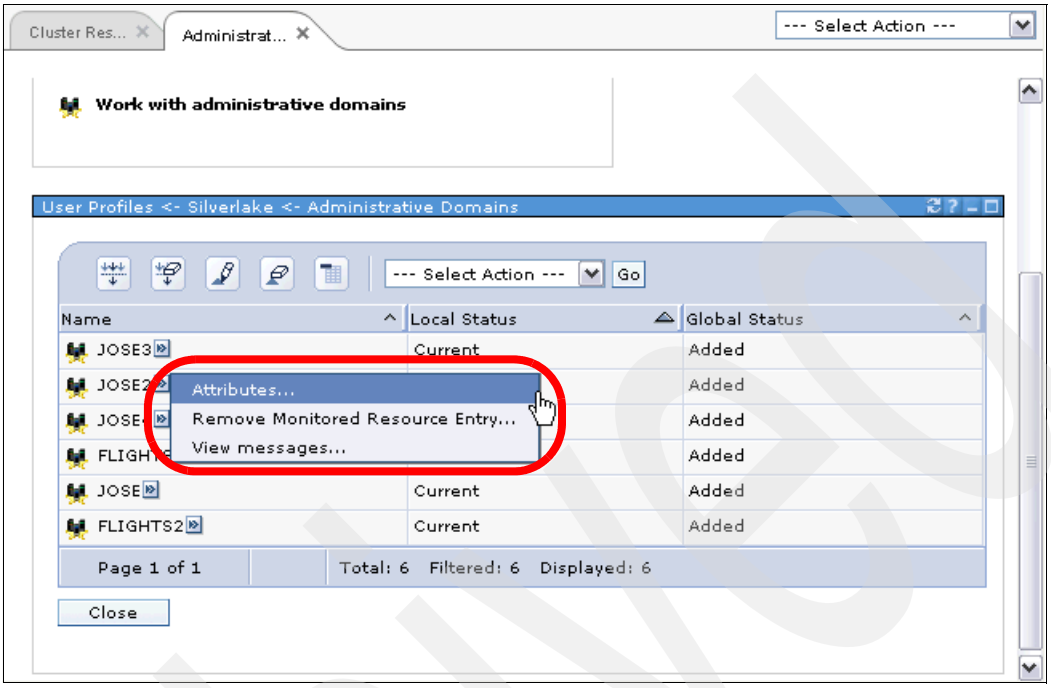


Figure 7-99 Work with monitor resource entry

You can also sort or filter the list using the respective buttons (see the colored circles in Figure 7-100).

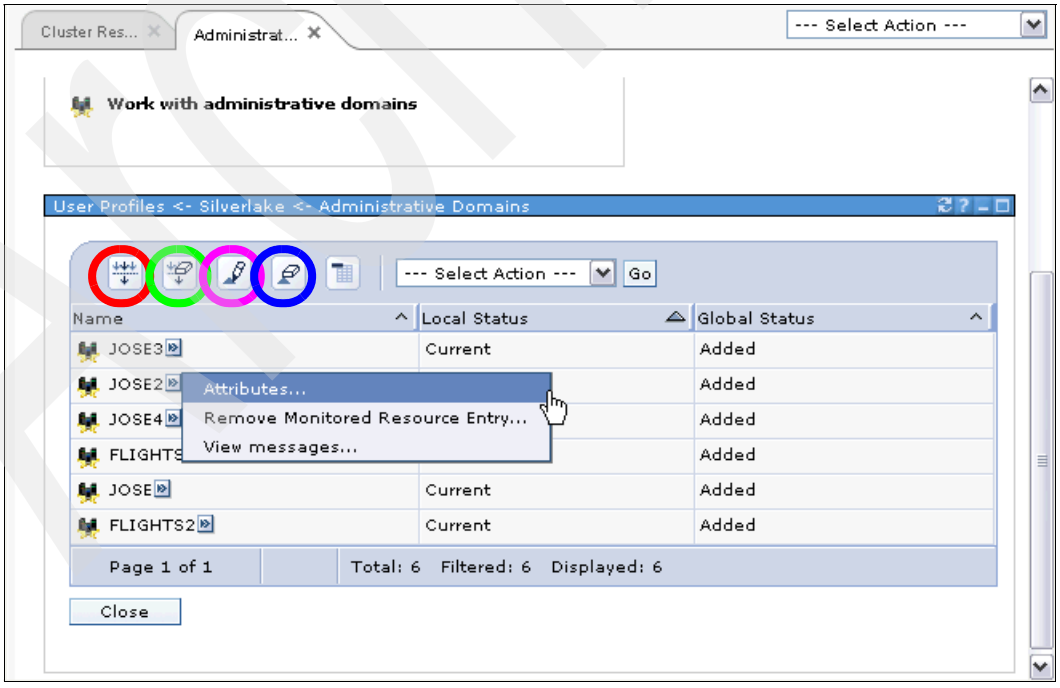


Figure 7-100 Filter and sort displayed monitored resource entries

For instance, if you click the Sort button you will be prompted to select the criteria and the order to in which to sort the list, as in Figure 7-101.

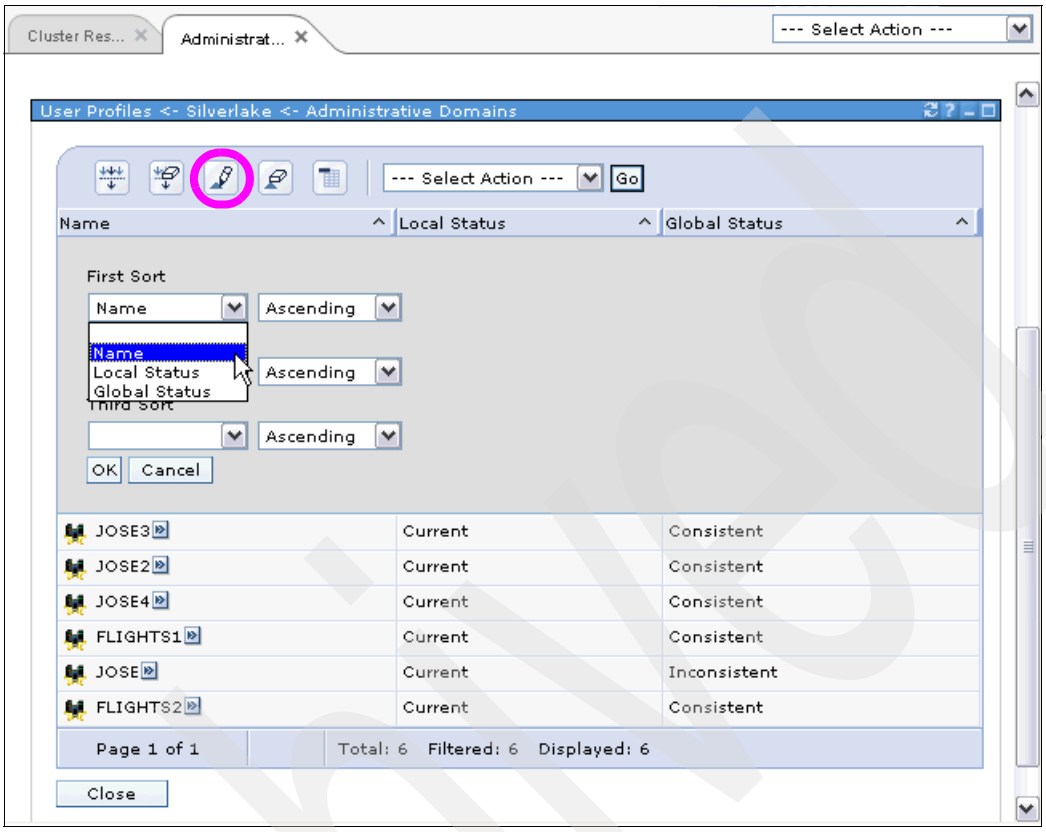


Figure 7-101 Sort displayed monitored resource entries

This gives the result shown in Figure 7-102.

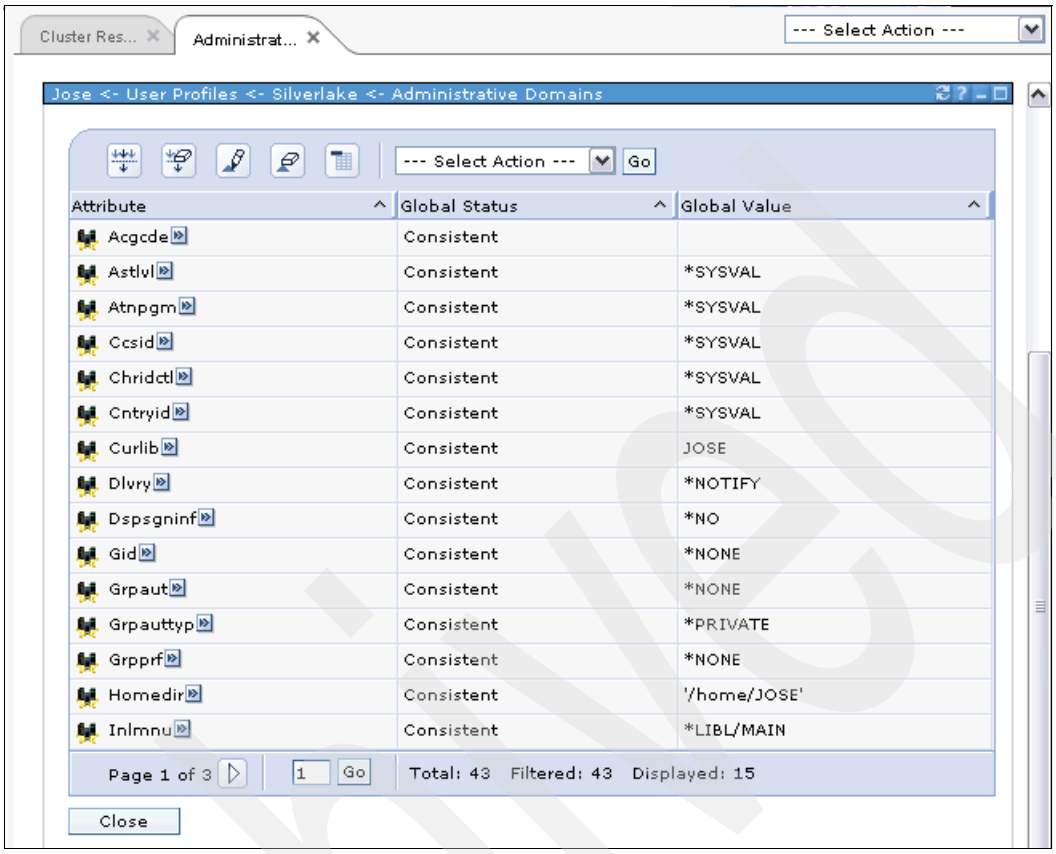


Figure 7-102 Example of displaying user profiles sorted by name

7.6.5 Disk GUI

The disk GUI provides you with the tools that you need to:

- Vary on and off an iASP.
- Work with ASP sessions and from there with ASP copy descriptions.
- Add disks to an existing ASP.
- Find out which jobs are currently connected to an iASP.

Figure 7-103 shows you the available options.

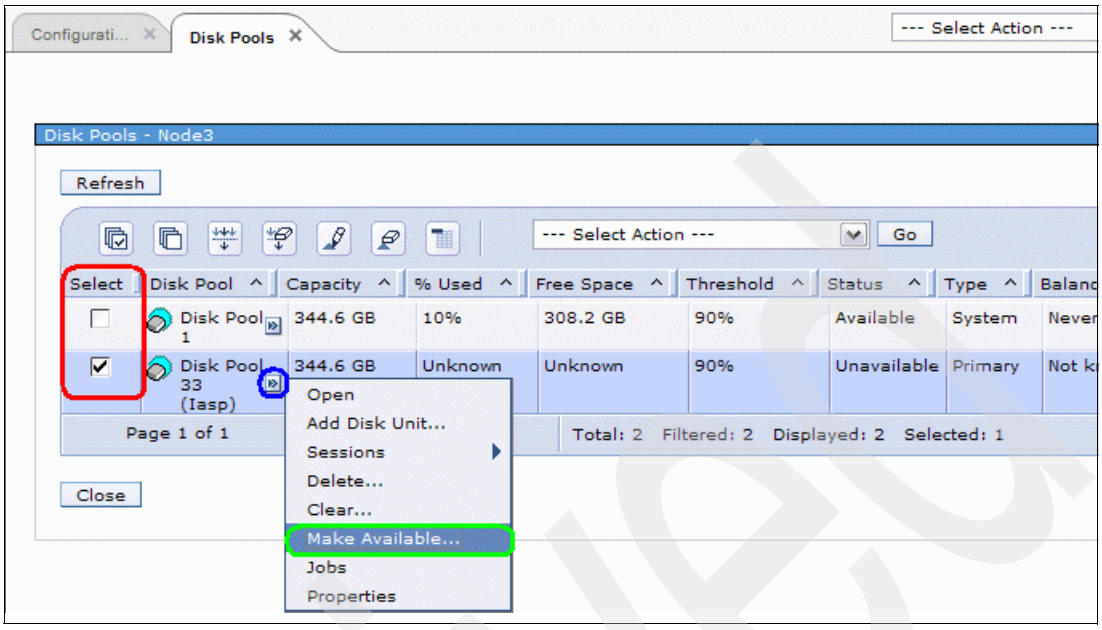


Figure 7-103 Work with disk pools

If you choose Session from the Disk Pool view, another window opens, as shown in Figure 7-104. When you select the session that you want to work with, you can then select the action that you want to take from the Select Action drop-down menu. These actions include:

- Suspend with tracking.
- Suspend without tracking.
- Resume.
- Detach with tracking.
- Detach without tracking.
- Reattach.
- Delete the session.
- Work with session properties.

Some of these actions are only available for certain types of sessions. Some actions are only available if your session is in a certain status (for example, resume is only shown when the session is suspended).

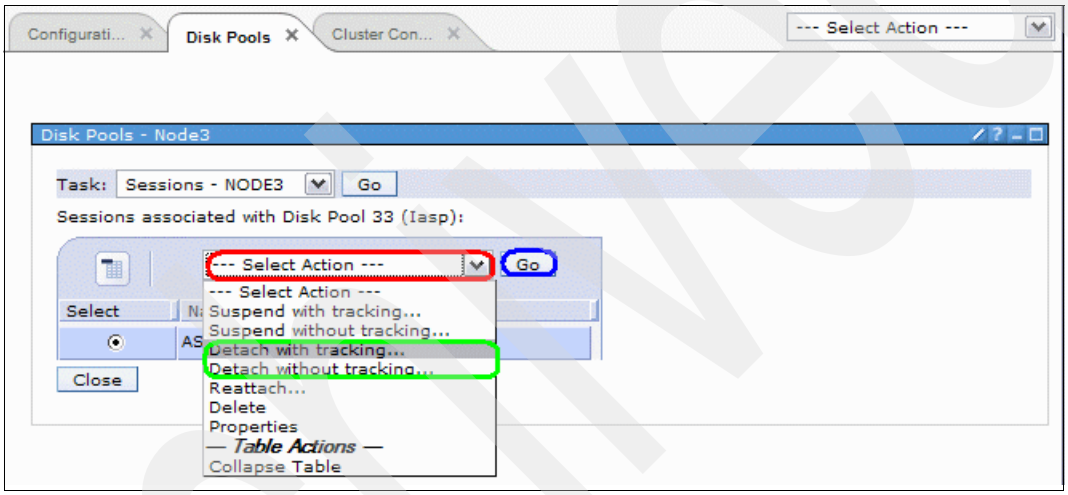


Figure 7-104 Session actions

8

Commands

In this chapter we show you how to set up a cluster environment with geographic mirroring and an administrative domain using the available commands. Note that setting up other environments that PowerHA for i offers would be quite similar to the example given in this chapter. We also provide information and descriptions of the cluster and independent auxiliary storage pool commands that were changed for 6.1.

8.1 Cluster command history

Clustering technology came out in i5/OS V4R4M0. At first there was no command-line access, but we could write our own commands with the supplied APIs.

At V5R1M0 big changes were made with some example cluster commands that could be compiled and used in the QUSRTOOL library.

V5R2M0 heralded a great day when we had fully supported commands in QSYS.

Since then there have been enhancements to commands, but no major changes. However, at 6.1 we see an entirely new generation of cluster commands.

8.2 Setting up a cluster environment using commands

In this section we provide the steps to set up a cluster environment with two nodes that will use geographic mirroring. This cluster environment will then serve as the basis for setting up an administrative domain and adding monitored resource entries to it. Though some of the commands are now part of PowerHA for i instead of the operating system, the command syntax has remained the same in many cases. Nevertheless, the following sections describe a complete cluster setup for geographic mirroring, including the setup of an administrative domain in order to provide you with all necessary steps.

8.2.1 Creating a cluster with geographic mirroring

The first step is creating a two-node cluster, as shown in Figure 8-1. In our example, we use two connections for heartbeating. We also make use of the newly available cluster message queue.

1. Configured as shown in Figure 8-1, in case of a failover, a message would be sent to the QSYSOPR message queue of the backup node. If no answer is received to this message within 10 minutes then an automatic failover occurs.

```
                        Create Cluster (CRTCLU)

Type choices, press Enter.

Cluster  . . . . . . . . . . . . > REDBOOK       Name
Node list:
  Node identifier  . . . . . . . > NODE3        Name
  IP address . . . . . . . . . . > '10.0.3.13'
                                   > '10.0.4.13'

  Node identifier  . . . . . . . > NODE4        Name
  IP address . . . . . . . . . . > '10.0.3.14'
                                   > '10.0.4.14'
               + for more values
Start indicator  . . . . . . . .   *YES         *YES, *NO
Target cluster version . . . . .   *CUR         *CUR, *PRV
Cluster message queue  . . . . . > QSYSOPR      Name, *NONE
  Library  . . . . . . . . . . . >   QSYS       Name
Failover wait time . . . . . . .   10           Number, *NOWAIT, *NOMAX
Failover default action  . . . .   *PROCEED     *PROCEED, *CANCEL

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

Figure 8-1 Create Cluster command

2. Start both cluster nodes with the command shown in Figure 8-2. Make sure that you issue this command for each of the cluster nodes.

```
                      Start Cluster Node (STRCLUNOD)

Type choices, press Enter.

Cluster  . . . . . . . . . . . .   redbook       Name
Node identifier  . . . . . . . .   node3         Name

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

Figure 8-2 Start Cluster Node command

3. Add these two nodes to a device domain. This can be done with the command shown in Figure 8-3. Again, make sure to issue this command for both nodes in your cluster.

```
                        Add Device Domain Entry (ADDDEVDMNE)

Type choices, press Enter.

Cluster  . . . . . . . . . . . .   redbook       Name
Device domain  . . . . . . . . .   redbook       Name
Node identifier  . . . . . . . .   node3         Name

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

Figure 8-3 Add Device Domain Entry command

4. You are now ready to create the independent auxiliary storage pool (iASP) (if this has not been available on your system before). The creation of an iASP must be done using one of the graphical interfaces. It cannot be done with any or command or using SST. You can find an example on how to do this in 7.2.1, “Create an iASP on the production node” on page 164.
5. Create a device description for the iASP on the backup system. Make sure that this device description and relational database name match the values that you used on your production system. Also, the resource name must be the same as the device description name. See Figure 8-4 for the command.

```
                    Create Device Desc (ASP) (CRTDEVASP)

Type choices, press Enter.

Device description . . . . . . . > IASP          Name
Resource name  . . . . . . . . . > IASP          Name
Relational database  . . . . . .   *GEN
Message queue  . . . . . . . . .   *SYSOPR       Name
  Library  . . . . . . . . . . .                 Name, *LIBL, *CURLIB
Text 'description' . . . . . . .   *BLANK


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

Figure 8-4 Create iASP device description on backup system

6. You now have all the prerequisites for creating a cluster resource group. In our example in Figure 8-5 on page 233 the CRG is set up to control the switching and failover process for the iASP. Failover may occur between the two nodes of the cluster. And, as we want to make use of geographic mirroring, the CRG also holds information about site names as well as on data port addresses that are used for mirroring data from the production iASP to the backup iASP. After a failover or a switchover occurs, the iASP will be automatically varied-on on the target system. As we already made use of the cluster message queue,

there is no need to also define a CRG-level failover message queue. This message queue would be ignored and the cluster message queue would be used.

```
                  Create Cluster Resource Group (CRTCRG)

Type choices, press Enter.

Cluster  . . . . . . . . . . . . > REDBOOK       Name
Cluster resource group . . . . . > REDBOOKCRG    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  . . . . . . . > NODE3         Name
  Node role  . . . . . . . . . . > *PRIMARY      *CRGTYPE, *PRIMARY...
  Backup sequence number . . . .   *LAST         1-127, *LAST
  Site name  . . . . . . . . . . > NODE3         Name, *NONE
  Data port IP address . . . . . > '10.0.1.13'
             + for more values > '10.0.2.13'

  Node identifier  . . . . . . . > NODE4         Name
  Node role  . . . . . . . . . . > *BACKUP       *CRGTYPE, *PRIMARY...
  Backup sequence number . . . .   *LAST         1-127, *LAST
  Site name  . . . . . . . . . . > NODE4         Name, *NONE
  Data port IP address . . . . . > '10.0.1.14'
             + for more values > '10.0.2.14'
             + for more values
Exit program format name . . . .   EXTP0100      EXTP0100, EXTP0200

Exit program data  . . . . . . .   *NONE
Distribute info user queue . . .   *NONE         Name, *NONE
  Library  . . . . . . . . . . .                 Name
Configuration object list:
  Configuration object . . . . . > IASP          Name, *NONE
  Configuration object type  . .   *DEVD         *DEVD, *CTLD, *LIND, *NWSD
  Configuration object online  . > *ONLINE       *OFFLINE, *ONLINE, *PRIMARY
  Server takeover IP address . .   *NONE
             + for more values
Text description . . . . . . . .   *BLANK

Failover message queue . . . . .   *NONE         Name, *NONE
  Library  . . . . . . . . . . .                 Name

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

Figure 8-5 Create Cluster Resource Group command

7. For test purposes you might want to start the CRG using the command shown in Figure 8-6.

```
                    Start Cluster Resource Group (STRCRG)

Type choices, press Enter.

Cluster  . . . . . . . . . . . . > REDBOOK       Name
Cluster resource group . . . . . > REDBOOKCRG    Name
Exit program data  . . . . . . .   *SAME

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

Figure 8-6 Start Cluster Resource Group command

As this was done for test purposes only, make sure to end the CRG using the ENDCRG command before moving on.

8. Create the mirror copy of the iASP. This cannot be done with any commands, but must be done using a graphical interface like IBM System Director Navigator. In IBM System Director Navigator choose **i5/OS Management** → **Configuration and Service** and then on the right panel select **Disk Pools**. You then must click the double arrow beside your disk pool and choose **Sessions** → **New** → **Geographic Mirroring- Configure Geographic Mirroring**, as shown in Figure 8-7

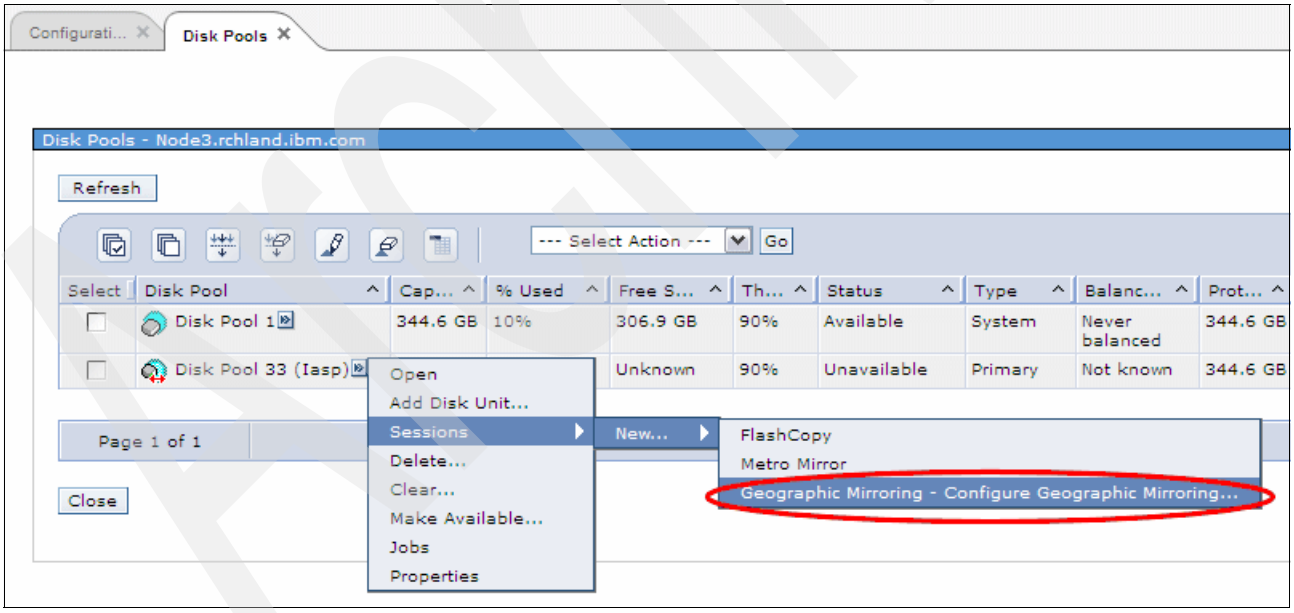


Figure 8-7 Configure geographic mirroring

Doing so presents you with the welcome panel shown in Figure 8-8. Click **Next**.

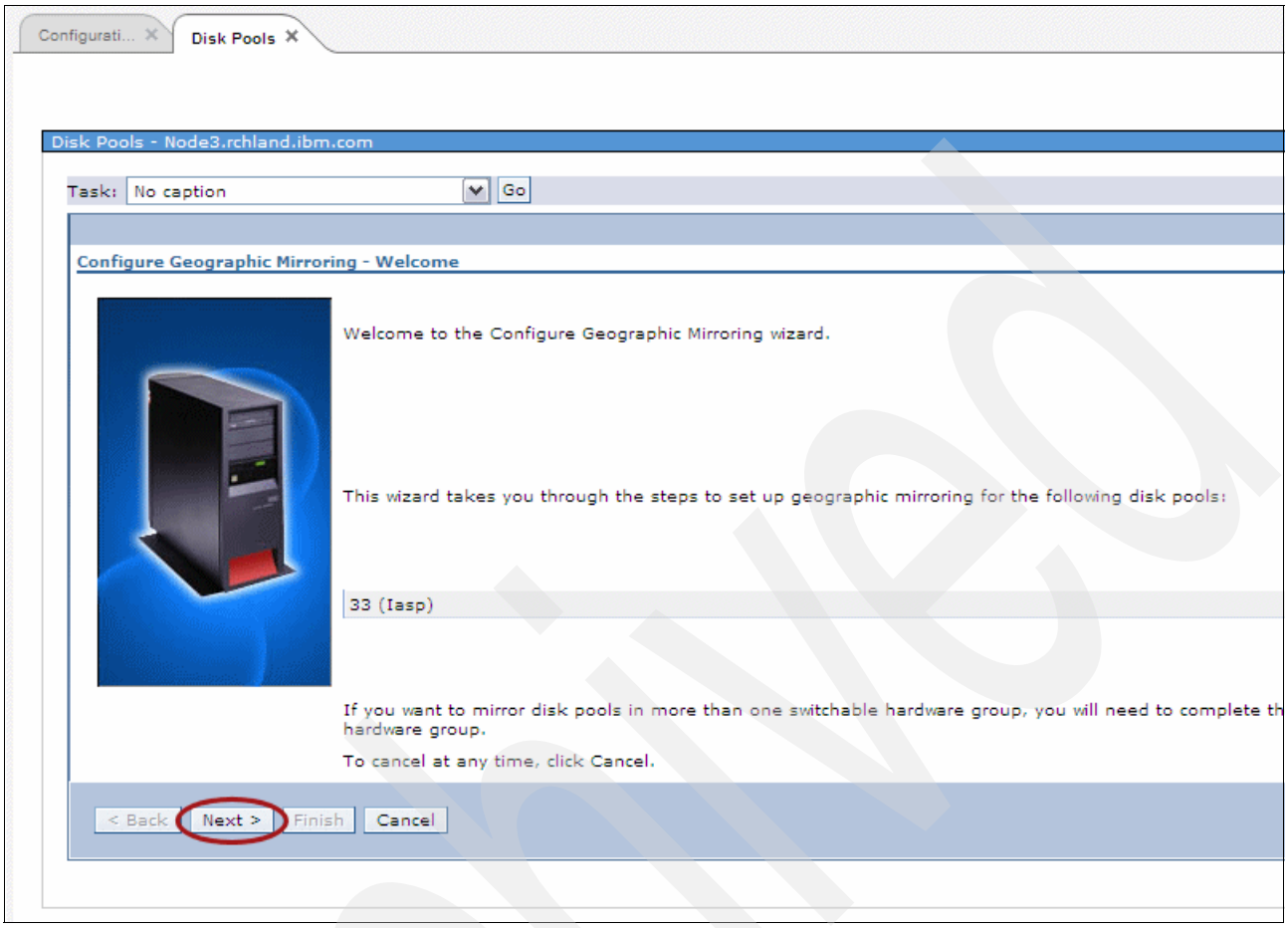


Figure 8-8 Configure geographic mirroring: Welcome window

9. The next panel shows you the properties that geographic mirroring uses by default. This includes the mode (synchronous or asynchronous), the recovery time-out parameter, and the resynchronization priority. To change any of theses values and to also have a closer look at additional parameters click **Edit**, as shown in Figure 8-9.

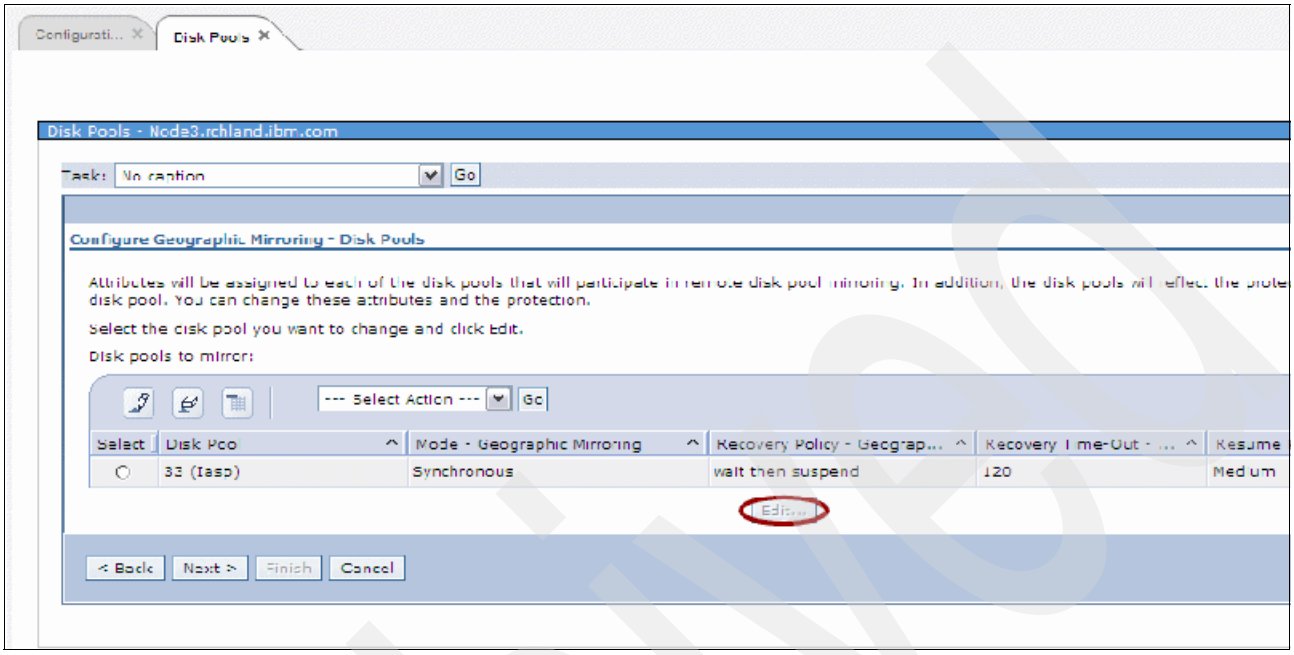


Figure 8-9 Geographic mirroring properties

10. The next panel also gives you the opportunity to change the size of the tracking space used for the iASP. Also make sure that you check the box next to **Protect data in this disk pool** if the disk that you want to put into the mirror copy of the iASP is using RAID5. Look at Figure 8-10 for details.

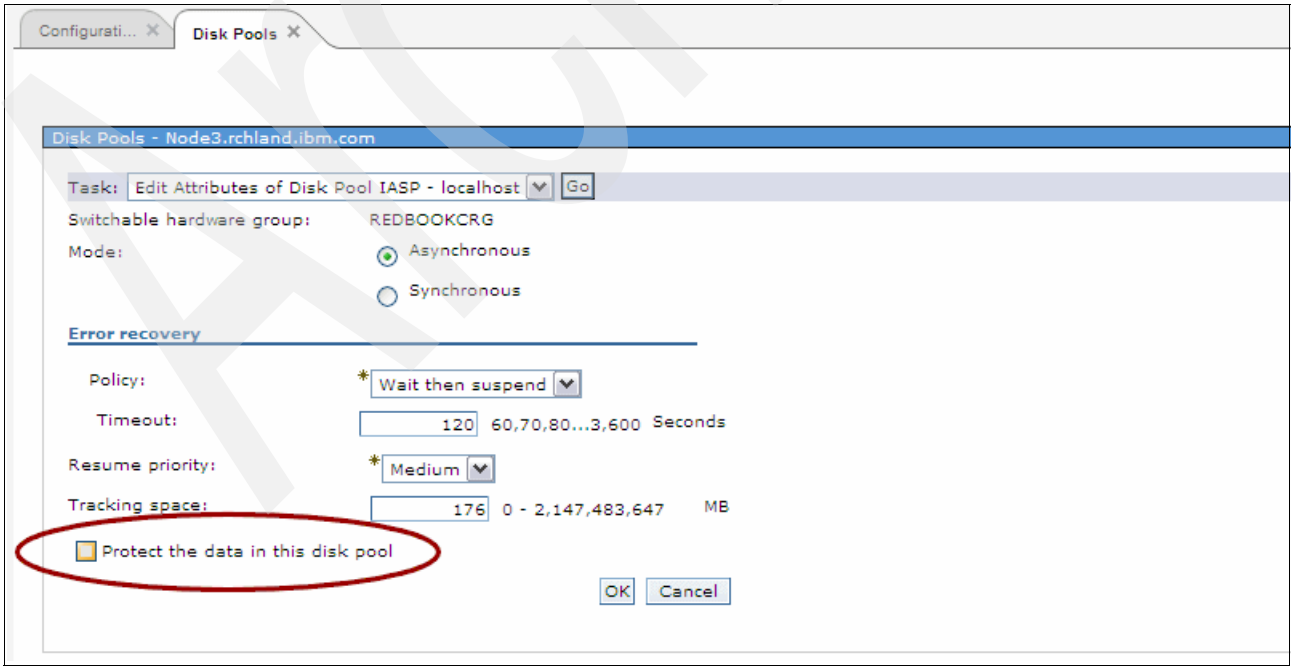


Figure 8-10 Geographic mirroring: Edit properties

11. Tell the GUI on which remote system you want to create the mirror copy of your iASP, as shown in Figure 8-11. Fill in the data correctly either using IP names or addresses. Then click **Next**.

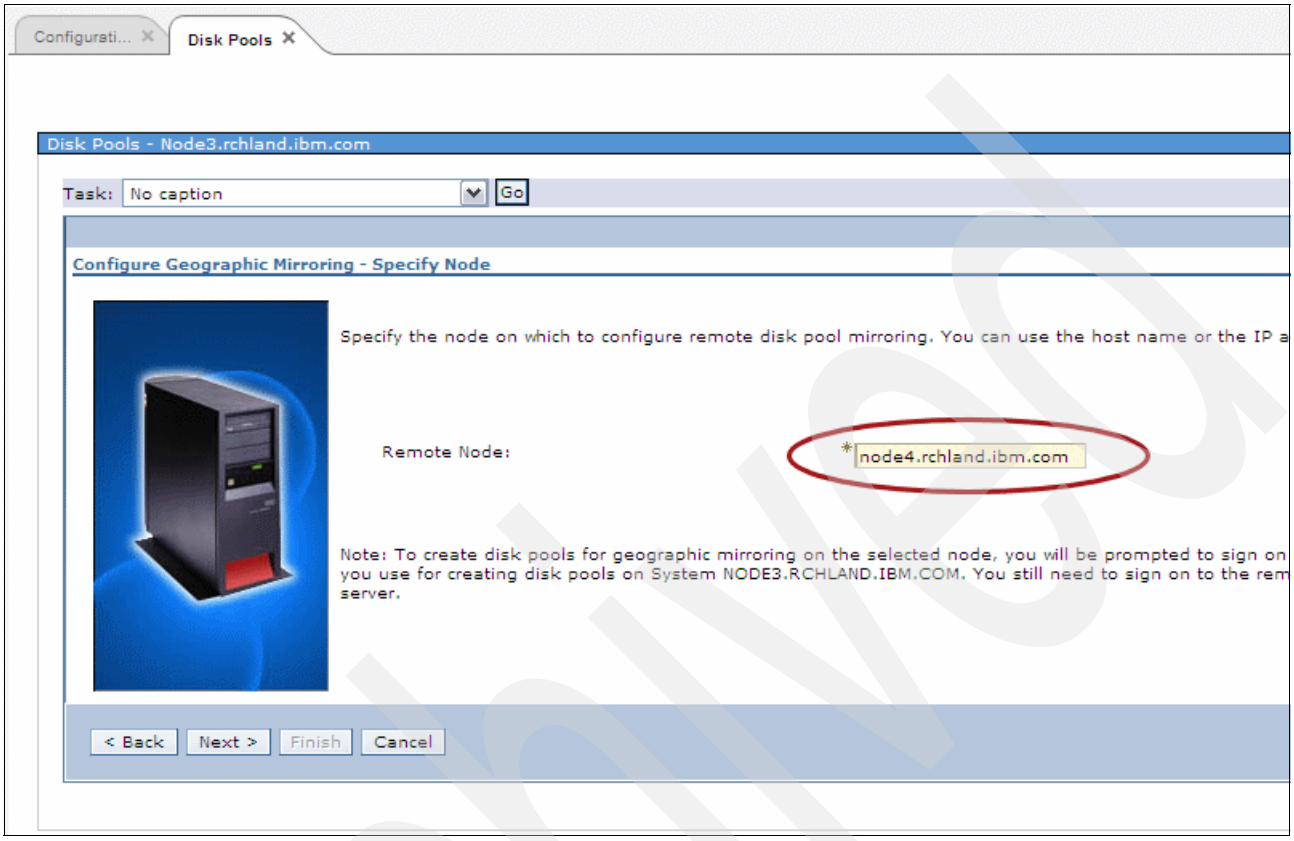


Figure 8-11 Geographic mirroring: Define remote system

12. Sign on to SST of the remote system, as shown in Figure 8-12.

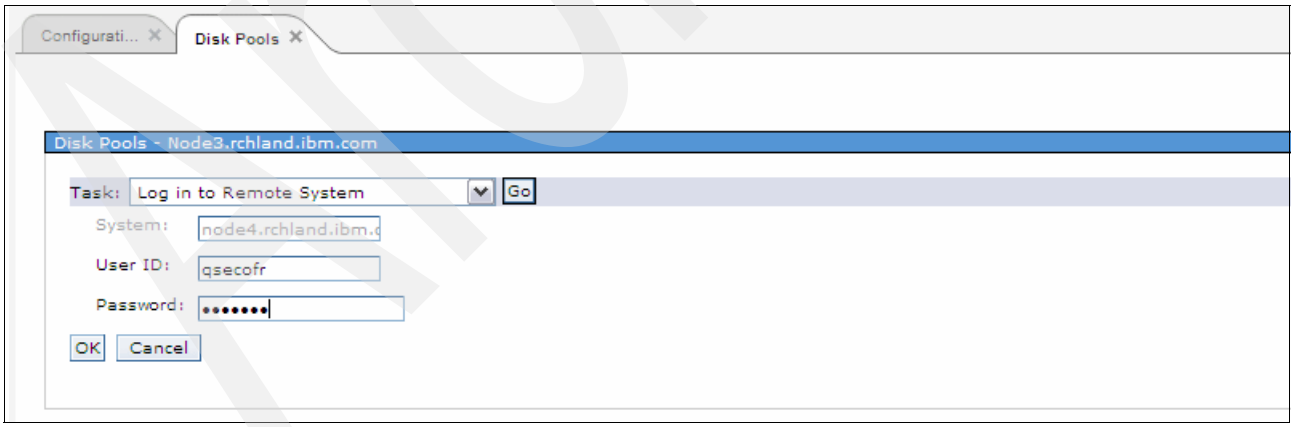


Figure 8-12 Geographic mirroring: SST sign on to remote system

13. The next panel shows you the current setup of your mirror copy of the iASP. It currently contains no disks. Therefore, click **Add Disks**, as shown in Figure 8-13.

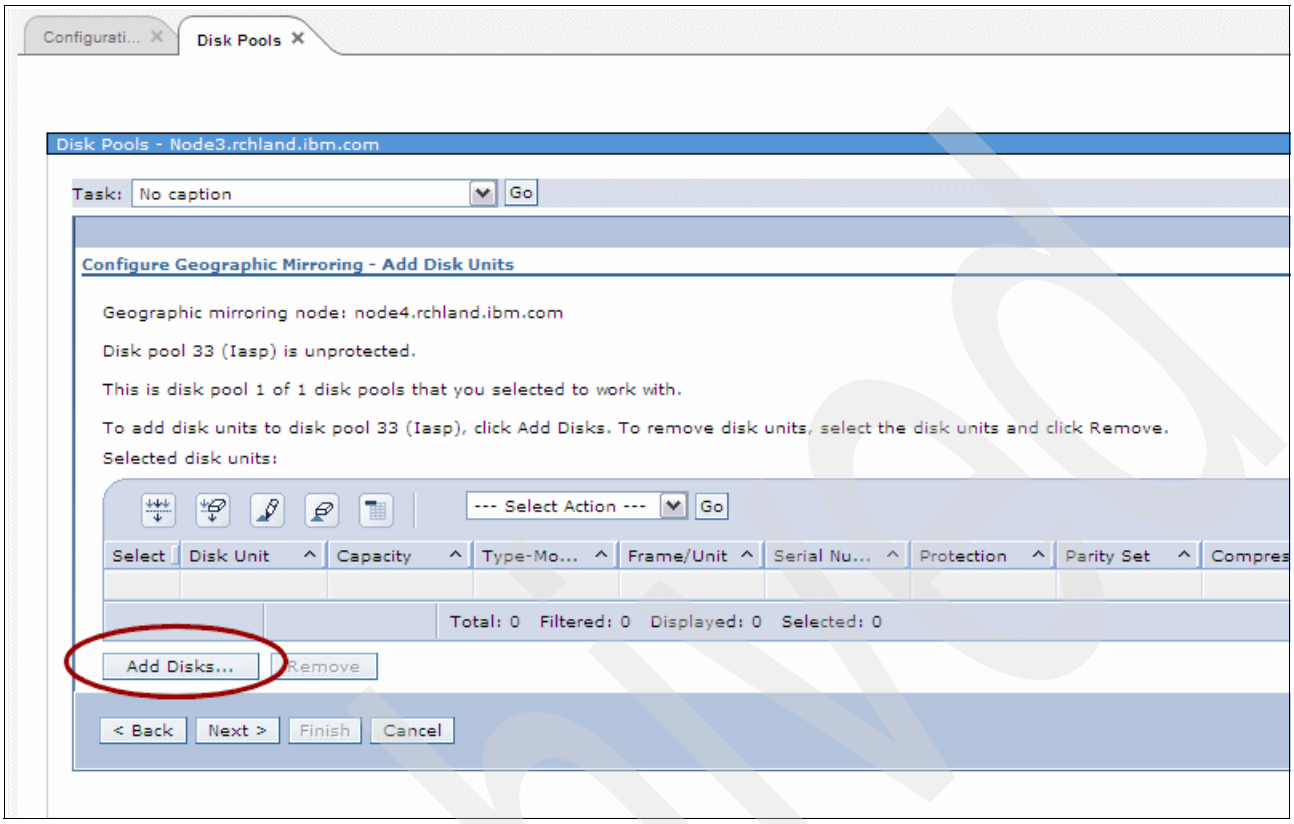


Figure 8-13 Geographic mirroring: Add Disks

14. The next panel presents you with all disks that are in the status unconfigured on your backup system. Select the disk units that you want to include in the mirror copy of your iASP, as shown in Figure 8-14, and click **Add**. Make sure that the resulting capacity is about the same as for the iASP on your production system.

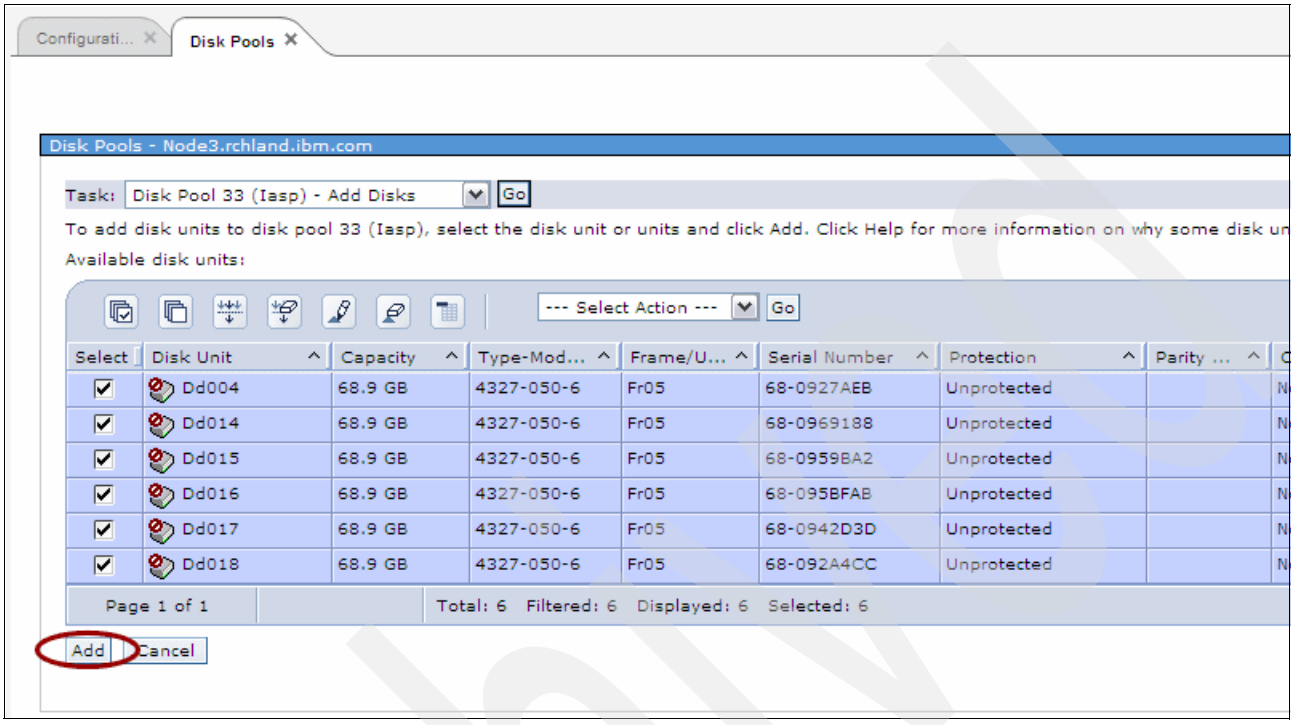


Figure 8-14 Geographic mirroring: Select disks to add

15. The configuration wizard then presents you the configuration resulting from your selection, as shown in Figure 8-15. If this looks like you planned it click **Finish**. If you need to make changes click **Back**.

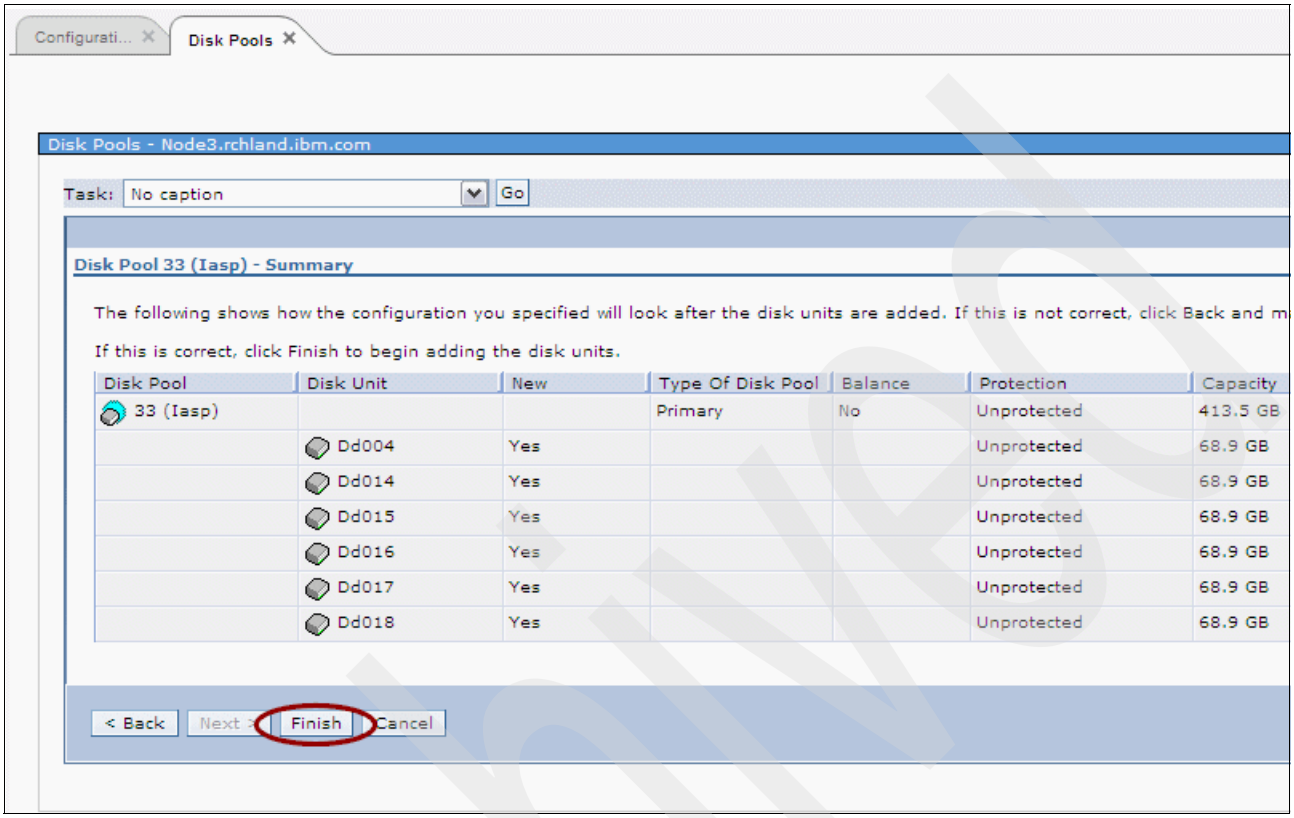


Figure 8-15 Geographic mirroring: Final configuration

16. While the disk configuration is taking place, you can see the panel shown in Figure 8-16. The percent complete information is regularly updated.

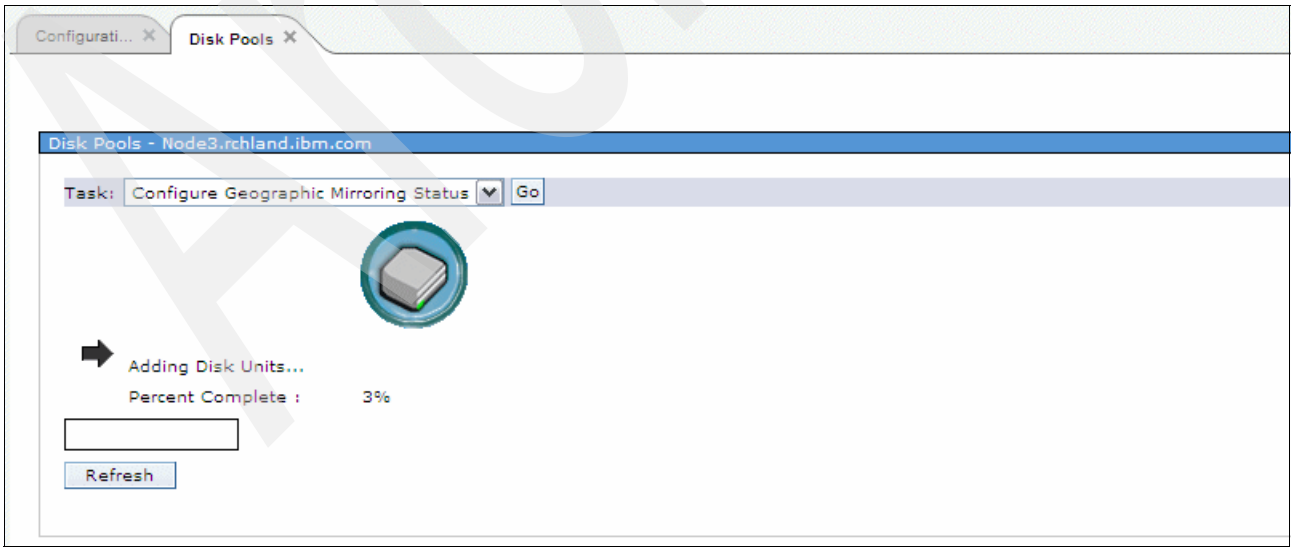


Figure 8-16 Geographic mirroring: Disks being added

17. After the configuration process is finished you must start your CRG using the STRCRG command and vary on the iASP using the command VRYCFG. In the next step, you must configure copy descriptions for the iASP that the CRG is working on. Notice that for each iASP in the CRG you need two copy descriptions—one pointing to the local node and one pointing to the remote node. Figure 8-17 shows the copy description for the production system. The information about storage host, location, and logical unit names is only needed if you are using metro mirror, global mirror, or FlashCopy.

```
                    Add ASP Copy Description (ADDASPCPYD)

Type choices, press Enter.

ASP copy . . . . . .   iasplocal     Name
ASP device . . . . . .   iasp          Name
Cluster resource group . . . . .   redbookcrg    Name, *NONE
Cluster resource group site  . .   node3         Name, *NONE
Storage host:
  User name  . . . . . .   *NONE
  Password . . . . . .
  Internet address . . . . . .

Location . . . . . .   *NONE         Name, *DEFAULT, *NONE
Logical unit name:
  TotalStorage device  . . . . .   *NONE
  Logical unit range . . . . . .                Character value
              + for more values

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

Figure 8-17 Add ASP Copy Description for production system

Figure 8-18 shows the command to create the corresponding copy description for the backup system.

```
                    Add ASP Copy Description (ADDASPCPYD)

Type choices, press Enter.

ASP copy . . . . . > IASPremote    Name
ASP device . . . . . > IASP        Name
Cluster resource group . . . . . > REDBOOKCRG    Name, *NONE
Cluster resource group site  . . > NODE4         Name, *NONE
Storage host:
  User name  . . . . . .   *NONE
  Password . . . . . .
  Internet address . . . . . .

Location . . . . . .   *NONE         Name, *DEFAULT, *NONE
Logical unit name:
  TotalStorage device  . . . . .   *NONE
  Logical unit range . . . . . .                Character value
              + for more values

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

Figure 8-18 Add ASP Copy Description for backup system

18. The last step to complete the setup of your geographic mirroring environment is to start the ASP session that links the two copy descriptions together. This is done with the command STRASPSSN, as shown in Figure 8-19. Note that the parameter's consistency source and consistency target are only valid for a global mirror environment. They have to be set to *NONE for all other environments.

```
                        Start ASP Session (STRASPSSN)

Type choices, press Enter.

Session  . . . . . .   aspssn        Name
Session type . . . . . .   *geomir       *GEOMIR, *METROMIR...
ASP copy:
  Preferred source . . . . . .   iasplocal     Name
  Preferred target . . . . . .   iaspremote    Name
  Consistency source . . . . . .   *NONE         Name, *NONE
  Consistency target . . . . . .   *NONE         Name, *NONE
              + for more values

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

Figure 8-19 Start ASP Session command

8.2.2 Setting up an administrative domain

With the clustering and geographic mirroring environment set up and started you are now ready to create an administrative domain that will help you to synchronize objects residing in the system ASP between your production and your backup side. This is also possible if you are using any of the storage-based solutions for high availability that PowerHA for i offers.

1. The first step you need to take is to create the administrative domain. This can be done with a new command that is part of the PowerHA for i license program. An example is shown in Figure 8-20. Make sure that you put all required nodes in the administrative domain node list. The synchronization option lets you decide how you want to handle changes on nodes that occurred while this node was not part of the active administrative domain. For more information see 14.1.4, “Resource synchronization” on page 354.

```
                    Create Cluster Admin Domain (CRTCAD)

Type choices, press Enter.

Cluster  . . . . . .   redbook       Name
Cluster administrative domain  .   redbookadm    Name
Admin domain node list . . . . .   node3         Name
              + for more values    node4
Synchronization option . . . . .   *LASTCHG      *LASTCHG, *ACTDMN

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

Figure 8-20 Create Cluster Administrative Domain command

2. Once created you then have to start the administrative domain using the STRCAD, as shown in Figure 8-21.

```
                    Start Cluster Admin Domain (STRCAD)

Type choices, press Enter.

Cluster  . . . . . .   redbook       Name
Cluster administrative domain  .   redbookadm    Name


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

Figure 8-21 Start Cluster Admin Domain command

3. Begin to add monitored resources by using the command ADDCADMRE. Figure 8-22 gives an example on how to do this. The panel content may vary slightly depending on the type of resource that you choose.

```
                    Add Admin Domain MRE (ADDCADMRE)

Type choices, press Enter.

Cluster  . . . . . . . . . . . . > REDBOOK       Name
Cluster administrative domain  . > REDBOOKADM    Name
Monitored resource . . . . . . . > jordan        Character value
Monitored resource type  . . . . > *USRPRF       *ASPDEV, *CLS, *ENVVAR...
Monitored attributes . . . . . .   *ALL          Name, *ALL
               + for more values

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

Figure 8-22 Add administrative domain monitored resource entry command

4. You can check the status of your monitored resource entries by using iSeries Navigator, IBM Systems Director Navigator for i5/OS, or by following these steps in a 5250 session:
   a. In the command line type WRKCLU and select option 8 (Work with administrative domains).
   b. Select option 7 on the administrative domain that you want to work with.

   This will lead you to the panel in Figure 8-23.

```
                      Work with Monitored Resources

Administrative domain  . . . . . . . . . :   REDBOOKADM
Consistent information in cluster  . . . :   Yes
Domain status  . . . . . . . . . . . . . :   Active

Type options, press Enter.
  1=Add   4=Remove   5=Display details   7=Display attributes

                Resource                Global           Local
Opt  Resource   Type        Library     Status           Status

     QASTLVL    *SYSVAL                 Consistent       Current
     JORDAN     *USRPRF     QSYS        Consistent       Current
     TESTADM    *USRPRF     QSYS        Consistent       Current
     TESTSJ1    *USRPRF     QSYS        Consistent       Current

                                                                      Bottom
Parameters for option 1 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve
F11=Order by global status      F12=Cancel   F24=More Keys
```

Figure 8-23 Work with Monitored Resources

8.3 Commands in QSYS

Most of the cluster commands that we have used in the past have their new versions moved into the PowerHA for i LPP and their own library named QHASM. During install of the PowerHA for i LPP, the commands will also be moved into the QSYS library. There are also important iASP-related commands in QSYS.

8.3.1 Cluster commands

The following cluster commands were viewed as important enough to leave in QSYS. These commands will be available regardless of ´whether the PowerHA for i LPP is installed.

Delete Cluster Resource Group

The Delete Cluster Resource Group (DLTCRG) command is used to delete a CRG on a partition. The DLTCRG command will only work when clustering is not active.

There are no parameter changes from the previous release.

Deleting a device CRG will not cause devices to switch back to the preferred primary system. Therefore, you must use care with this command or the hardware associated with the device CRG could be left on the wrong partition. This is especially a danger with HSL switchable hardware towers, where the hardware could be left with the wrong SPCN owner.

If there is a takeover IP address used with the CRG that was created by cluster, and the IP is not active at the time of the delete, then the IP address will be removed.

There is also a version of this command that can be used when the cluster is active. For more information see “Delete Cluster” on page 251.

Dump Cluster Trace

The Dump Cluster Trace (DMPCLUTRC) command is used to dump out internal cluster information for later problem debug purposes. The information will be written to a physical file. There are various levels of information that can be chosen.

There are no parameter changes from the previous release.

When a cluster node is ended, the trace information is gone and can no longer be gathered through this command. The important thing to remember is that if you actually have a problem, this command would need to be used before any action that would end the cluster on this node.

Change Cluster Recovery

The Change Cluster Recovery (CHGCLURCY) command can be used to force certain types of problem recovery. This command is not intended for general use, but for use in certain circumstances when under direction of IBM support.

There are no parameter changes from the previous release.

This command can be used to cancel or initiate specific cluster recovery actions.

Start Clustered Hash Table Server

The Start Clustered Hash Table Server (STRCHTSVR) command is used to create a clustered hash table server on each node that is specified. The clustered hash table server

will also be started on each node and a job will start to service the clustered hash table on this node.

There are no parameter changes from the previous release.

The clustered hash table server will allow data to be shared between the different nodes. Note that the data stored in the clustered hash table is not in permanent storage. When the clustered hash table ends, the data is gone.

See “End Clustered Hash Table Server” on page 246 for information about how to end the clustered hash table.

You can find more information about clustered hash tables in the Information Center by following the path **Implementing high availability → Planning your high availability solution → Security planning for high availability → Distributing cluster-wide information**.

End Clustered Hash Table Server

The End Clustered Hash Table Server (ENDCHTSVR) command is used to end the declared clustered hash table server on the nodes specified. If the last node is removed from the clustered hash table then the job will be ended on all nodes and the clustered hash table server will be deleted from the cluster. Any further requests to the clustered hash table server would fail.

There are no parameter changes from the previous release.

See “Start Clustered Hash Table Server” on page 245 for information about how to start the clustered hash table.

You can find more information about clustered hash tables in the Information Center by following the path **Implementing high availability → Planning your high availability solution → Security planning for high availability → Distributing cluster-wide information**.

8.3.2 iASP commands

In the list below we included independent auxiliary storage pool (iASP) specific commands. Other commands such as VRYCFG can be used with an iASP, as well as other devices.

Display Auxiliary Storage Pool Status

The Display ASP Status (DSPASPSTS) command shows the vary-on steps and the progress that has been made for an iASP device when it is being varied on or off.

There are no parameter changes from the previous release.

This command lets you view the same steps that a system takes when it IPLs up or down that you would normally see from the control panel, virtual control panel, or Hardware Management Console (HMC). You can see a similar list from iSeries access if you use that to vary on the iASP.

Note that even if the vary on/off is finished, you can use this command to see the timing for the last steps. So if a vary-on takes longer than normal, you can use this command to see which steps the iASP vary on/off took the most time in.

Set Auxiliary Storage Pool Group

The Set Auxiliary Storage Pool Group (SETASPGRP) command will set the auxiliary storage pool (ASP) group for the current thread. This will allow that thread to have access to the objects inside of the different independent auxiliary storage pools within that ASP group.

There are no parameter changes from the previous release.

The command will also let you change the libraries in the library list for the current thread.

You can only have one ASP group associated with your thread at a time, so if you previously had a different ASP group associated with your thread and use this command the newly specified ASP group will replace the previous one.

Change Auxiliary Storage Pool Activity

The Change ASP Activity (CHGASPACT) command can be used to suspend database transactions and Integrated File System (IFS) changes for the system and basic ASPs or for iASPs. The command can also be used to resume these transactions and force changes to disk.

V6R1M0 is the first release to have this command. Figure 8-24 shows an example of the command.

```
                      Change ASP Activity (CHGASPACT)

Type choices, press Enter.

ASP device . . . . . .                  Name, *SYSBAS
Option . . . . . .                      *SUSPEND, *RESUME, *FRCWRT
Suspend timeout  . . . . . .            Number
Suspend timeout action . . . . .  *CONT *CONT, *END

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

Figure 8-24 Example of the CHGASPACT command

The obvious reason to use the functions in CHGASPACT would be to make a more optimal environment for performing a FlashCopy or geographic mirroring detach.

Important: This command will not prevent all activity and is not the equivalent of having the iASP varied off, so the vary-on of the FlashCopy target or detached Geographic Mirrored iASP will still result in an abnormal vary-on path being followed.

For the best data protection and the shortest vary-on for the target of FlashCopy or a detached mirror copy, vary off the iASP on the production node first. If you cannot do a vary-off, CHGASPACT OPTION(*SUSPEND) is the next best option.

*OPTION(*FRCWRT)*

The OPTION(*FRCWRT) option flushes eligible pages from memory out to DASD.

*OPTION(*SUSPEND)*

The OPTION(*SUSPEND) does the option (*FRCWRT) automatically. It suspends what DB transactions (such as commitment control) it can in the specified time. It suspends all DB

options (non-commitment control) that it can within 10 seconds. It flushes the journal cache, if using journal caching. It also suspends IFS database type activity.

If the OPTION(*RESUME) option is not run to free everything up, OPTION(*SUSPEND) automatically ends after 20 minutes.

OPTION(*RESUME)

OPTION(*RESUME) ends OPTION(*SUSPEND) and frees up database activity again.

OPTION(*RESUME) does not restore to memory what was moved to disk. It lets normal system operations page into memory what is needed.

Work Auxiliary Storage Pool Jobs

The Work with ASP Jobs (WRKASPJOB) command can be used to get a list of jobs with access to an ASP group.

There are no parameter changes from the previous release.

From this panel you can end a job (ENDJOB), work with a job (WRKJOB), or send a message to a user (similar to SNDMSG).

This is helpful when you need to bring the iASP to a state where you could vary it off normally, do a reclaim storage (RCLSTG), and so on.

Vary Configuration

The Vary Configuration (VRYCRG) command can be used to vary devices on and off.

There are no parameter changes from the previous release.

For an iASP you would always specify the CFGTYPE parameter as *ASP.

To vary on the iASP you would specify STATUS of *ON. This would start the iASP equivalent of a system IPL up. To vary off the iASP you would specify STATUS of *OFF. This would start the iASP equivalent of an system IPL down. With STATUS of *OFF, you also can specify the FRCVRYOFF parameter. The normal options are *NO and *YES. *YES can cause an abnormal shut down, which would cause an abnormal vary on path the next time that you want to use the iASP and vary it on. *NO will not be able to vary off the iASP if anyone is using it. You can use the WRKASPJOB command to see if any jobs are using the iASP. See “Work Auxiliary Storage Pool Jobs” on page 248 for more information.

To see the vary-on/off status of the iASP you can use the DSPASPSTS command. See “Display Auxiliary Storage Pool Status” on page 246 for more information.

8.4 Commands in PowerHA for i LPP

Most of the cluster commands that were in QSYS in previous releases have had their new versions moved into the PowerHA for i LPP. Moving to the LPP has seen these commands moved into the QHASM library, though they are also copied into QSYS during install.

8.4.1 Base cluster commands

The following commands in the PowerHA for i LPP effect changes on the base cluster environment and allow information about that environment to be viewed.

Add Cluster Node Entry

The Add Cluster Node Entry (ADDCLUNODE) command can be used to add a node to an existing cluster.

There are no parameter changes from the previous release.

You can specify to start clustering on the node automatically, or you can start it later with the STRCLUNOD command. See “Start Cluster Node” on page 252 for more information.

One thing to remember is that the DSPNETA/CHGNETA-ALWADDCLU parameter must be set to allow the add. If you have ALWADDCLU set to *RQSAUT instead of *ANY then it will require validation using X.509 digital certificates, and the node running the command and the node being added must have on:

- Operating System option 34 (Digital Certificate Manager)
- Cryptographic Access Provider Product (AC2 or AC3)

Also, the release of the node must be compatible with the existing cluster.

Change Cluster

The Change Cluster (CHGCLU) command can be used to change cluster configuration parameters.

This is a brand new command at V6R1M0, replacing a similar command at V5R4M0 called CHGCLUCFG. Figure 8-25 gives an example.

```
                          Change Cluster (CHGCLU)

Type choices, press Enter.

Cluster . . . . . .                    Name
Cluster message queue . . . . .  *SAME  Name, *SAME, *NONE
  Library . . . . . .                  Name
Failover wait time . . . . . .   *SAME  Number, *SAME, *NOWAIT...
Failover default action . . . .  *SAME  *SAME, *PROCEED, *CANCEL
Configuration tuning level . . .  *SAME  *SAME, *NORMAL, *MIN, *MAX

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

Figure 8-25 Example of the CHGCLU command

Prior to IBM i 6.1 there was a cluster function called a failover message queue that could be set up for a CRG. If the failover message queue was enabled, during a failover situation a message would be sent to a specified queue and the user could cancel or continue the failover. While this proved to be a useful function, it was discovered to be less than optimal for customers with multiple CRGs that would all failover at the same time, as the user would have to respond to a message for each and every CRG using this function. This function is still available at the CRG level and can be set using the CHGCRG command. see “Change Cluster Resource Group” on page 264 for more information.

Therefore, starting in V6R1M0 there is a new option available that can be chosen from this command. It will give you the option to have one message sent to a queue, instead of one for each CRG. Just like on the CRG option, you can put in a time-out value where you can

specify whether clustering should go ahead and fail over or not fail over if no one answers the failover message within a specified time.

Note that if both the cluster and CRG level failover message queues are defined, only the cluster level message queues will be used.

There is also a section where you can override the normal time-out values for cluster heartbeat. We recommend that you do not change the defaults without testing and consulting with IBM support.

To check current values for cluster that can be changed through CHGCLU, use the DSPCLUINF command. See “Display Cluster Information” on page 252 for more information.

Change Cluster Node Entry

The Change Cluster Node Entry (CHGCLUNODE) command can be used to change the status of a node or the IP address information for a node. The node can be active or not.

There are no parameter changes from the previous release.

CHGCLUNODE can change the status of a node from partitioned to failed. There are certain failure conditions that cluster resource services cannot detect as a node failure. Rather, the problem appears to be a communication problem and the cluster looks like it has become partitioned. This will allow a normal failover to happen if needed.

You can add, remove, or modify an IP address for the cluster node. This is the IP address that the node uses to communicate with other nodes in the cluster.

Change Cluster Version

The Change Cluster Version (CHGCLUVER) command can be used to increment the cluster version by one. This can only be done if all nodes in the cluster have a potential cluster node shown that is higher than the current cluster version. You can check the cluster version through the DSPCLUINF command. See “Display Cluster Information” on page 252 for more information.

There are no parameter changes from the previous release.

You cannot use this command to decrement the cluster version. To decrement the cluster version you must delete the cluster and then recreate it at a lower level.

The cluster version determines the level at which each node talks to the other node. New cluster versions will contain new cluster function that you may want to use.

You can find more information about this in the Information Center following the path **Availability → High availability technologies → i5/OS Cluster technology → Cluster concepts → Cluster version**.

Create Cluster

The Create Cluster (CRTCLU) command can be used to create a brand new cluster.

You can specify to turn clustering on automatically if the cluster has only one node. If you specify to start and there is more than one node, then the parameter is ignored. If you want to start the node later or you need to start multiple nodes, you can do it with the STRCLUNOD command. See “Start Cluster Node” on page 252 for more information.

To later add more nodes to the cluster, you can use the ADDCLUNODE command. See “Add Cluster Node Entry” on page 249 for more information.

There is also a section to define your cluster version. You can specify to use the current version or a previous version. You can find more information about this in the Information Center by following the path **Availability → High availability technologies → i5/OS Cluster technology → Cluster concepts → Cluster version**.

Before V6R1M0 there was a cluster function called a failover message queue that could be set up for a CRG. If the failover message queue was enabled, during a failover situation a message would be sent to the specified queue and the user could cancel or continue the failover. While this proved to be a useful function, this was discovered to be less than optimal for customers with multiple CRGs that would all fail over at the same time, as the user would have to respond to a message for each and every CRG using this function. This function is still available at the CRG level and can be set using the CHGCRG command. See “Change Cluster Resource Group” on page 264 for more information.

Therefore, starting in 6.1 there is a new option that can be chosen from this command that will give us the option to have one message per cluster sent to a queue, instead of one for each CRG. See the example in Figure 8-26. Just like on the CRG option you can put in a time-out value where you can specify whether clustering should fail over or not fail over if no one answers the failover message within a specified time.

```
                        Create Cluster (CRTCLU)

Type choices, press Enter.

Cluster  . . . . . .                Name
Node list:
  Node identifier  . . . . . .       Name
  IP address . . . . . .

              + for more values
Start indicator  . . . . . .   *YES          *YES, *NO
Target cluster version . . . . .   *CUR          *CUR, *PRV
Cluster message queue  . . . . .   *NONE         Name, *NONE
  Library  . . . . . .                         Name
Failover wait time . . . . . .   *NOWAIT       Number, *NOWAIT, *NOMAX
Failover default action  . . . .   *PROCEED      *PROCEED, *CANCEL

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

Figure 8-26 Example of the CRTCLU command

Note that if both the cluster and CRG level failover message queues are defined, only the cluster level message queues will be used.

Delete Cluster

The Delete Cluster (DLTCLU) command will delete a cluster on all nodes currently active in a cluster. All CRGs and device domains assigned to the cluster are also deleted. Cluster nodes in inactive, partitioned, or failed status will not be deleted. If this is run on an inactive or partitioned node then only that node is affected.

There are no parameter changes from the previous release.

A node that is a member of a device domain has internal information related to ASPs, disk units, virtual memory addresses, and so on. After the cluster is deleted this information will still be there until the next IPL. Therefore, this node will not be able to become part of another device domain until after it is IPLed.

Display Cluster Information

The Display Cluster Information (DSPCLUINF) command is used to output information about the cluster to either panel or to print.

It is important to check the consistent information in cluster parameter. If this is *YES then the information shown is current for the cluster. However, if this is *NO, then the information shown is only the state that the cluster was in when this node was last active.

The way that the cluster information is presented has changed with IBM i 6.1. You now always get the full information displayed, as opposed to IBM i 5.4 where you could choose between *Full or *Basic display.

End Cluster Node

The End Cluster Node (ENDCLUNOD) command can be used to end clustering on one or multiple active nodes.

There are no parameter changes from the previous release.

You can use this command to end any active node from any other active node. If there is a cluster partition you can only use the command for nodes on your side of the partition.

If the node being ended is the primary node for an active CRG, ownership of the hardware associated with the CRG will move to a backup node. If the CRG is not active or there are no backup nodes that are active there will be no change.

To start a cluster node you can use the STRCLUNOD command. See “Start Cluster Node” on page 252 for more information.

Remove Cluster Node Entry

The Remove Cluster Node Entry (RMVCLUNODE) command can be used to remove a cluster node from the cluster. The node will also be removed from any device domains that it once belonged to. Note that the CRG objects on the node will only be deleted if the node has a status of active when this command is called.

There are no parameter changes from the previous release.

If the cluster is partitioned, than any node that is removed would have to be removed on the other partitions as well or the cluster partitions will not be able to re-merge.

To run this on a node that is not active, you would have to run this command from that specific node.

A node that is a member of a device domain has internal information related to ASPs, disk units, virtual memory addresses, and. After the node is removed this information will still be there until the next IPL. Therefore, this node will not be able to become part of another device domain until after it is IPLed.

Start Cluster Node

The Start Cluster Node (STRCLUNOD) command can be used to start clustering on a node in the cluster. You can use it to start the node you are on or another node.

There are no parameter changes from the previous release.

If the cluster is partitioned, you can only start the node that you are on.

One thing to remember is that the DSPNETA/CHGNETA ALWADDCLU parameter must be set to allow the start on this node. If you have ALWADDCLU set to *RQSAUT instead of *ANY then it will require validation using X.509 digital certificates and the requesting node and node being added must have on:

- Operating System option 34 (Digital Certificate Manager)
- Cryptographic Access Provider Product (AC2 or AC3)

The potential node version of the node that you are starting must be the same as the current cluster version or one level higher. You can find more information about this in the Information Center by following the path **Availability → High availability technologies → i5/OS Cluster technology → Cluster concepts → Cluster version**.

If the node being started is in a device domain, then Operating System option 41 (HA Switchable Resources) must be installed and a valid license key must exist on that node.

Work Cluster

The Work Cluster (WRKCLU) command is used to show a menu of different cluster options, as shown in Figure 8-27.

At IBM 6.1 we have new options included. Figure 8-27 provides you with an example of the menu provided by this command.

```
                          Work with Cluster
                                                          System:   NODE4
Cluster  . . . . . . . . . . . . . . . . . :   TUNDRA
Consistent information in cluster  . . . :   Yes

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

    20. Dump cluster trace

Selection or command
===>

F1=Help   F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel
```

Figure 8-27 Example of WRKCLU menu panel

In the following sections we explain each of the numbered options on the Work with Cluster display.

Display Cluster Information

Using option 1 off of the WRKCLU OPTION(*SELECT) menu or using WRKCLU OPTION(*CLUINF) will bring up the Display Cluster Information panel. This is the same information that you can get from running the DSPCLUINF command on the first and second panel. For more information see “Display Cluster Information” on page 252.

Display Cluster Configuration Information

Using option 2 off of the WRKCLU OPTION(*SELECT) menu or using WRKCLU OPTION(*CLUINF) brings up the Display Cluster Configuration Information panel. This shows the Configuration and Tuning Parameters that you would normally see on third panel of DSPCLUINF, as well as the cluster configuration tuning level that you can see on the first panel. For more information see “Display Cluster Information” on page 252.

Work with Cluster Nodes

Using option 6 off of the WRKCLU OPTION(*SELECT) menu or using WRKCLU OPTION(*NODE) will bring up the Work with Cluster nodes panel. This is illustrated in Figure 8-28.

```
                          Work with Cluster Nodes

Local node . . . . . . . . . . . . . . . :   NODE4
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add   2=Change   4=Remove   8=Start   9=End   20=Dump trace

                              Potential
                              Node  Mod
Opt   Node       Status       Vers Level   -------Interface Addresses-------

      NODE1      Active          6     0   192.168.154.21
      NODE2      Inactive        5     0   192.168.154.22
      NODE3      Inactive        5     0   192.168.154.23
      NODE4      Active          6     0   192.168.154.24
                                                                      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 8-28 Example of the Work with Cluster Nodes panel

The Work with Cluster Nodes display gives several options for working with the nodes in a cluster. It shows the current members and their status, very similar to DSPCLUINF on the second panel. For more information see “Display Cluster Information” on page 252.

The options shown on the Work with Cluster Nodes panel are:

- Option 1 invokes the ADDCLUNODE command. See “Add Cluster Node Entry” on page 249 for more information.
- Option 2 invokes the CHGCLUNODE command. See “Change Cluster Node Entry” on page 250 for more information.
- Option 4 will remove the node from the cluster. This is very similar to the RMVCLUNODE command. See “Remove Cluster Node Entry” on page 252 for more information.

- Option 8 will start a start a cluster node. This is very similar to the STRCLUNOD command. See “Start Cluster Node” on page 252 for more information.
- Option 9 will invoke the ENDCLUNOD command. See “End Cluster Node” on page 252 for more information.
- Option 20 will invoke the DMPCLUTRC command. See “Dump Cluster Trace” on page 245 for more information.

Work with Device Domains

Using option 7 off of the WRKCLU OPTION(*SELECT) menu or using WRKCLU OPTION(*DEVDMN) will bring up the Work with Device Domains panel. This is illustrated in Figure 8-29.

```
                          Work with Device Domains

Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add   6=Work with nodes   7=Work with switchable hardware

                           Number
Opt    Device Domain      of Nodes    ---------------Nodes---------------

        ARCTIC                   4    NODE1    NODE2    NODE3    NODE4
                                                                    Bottom
Parameters for option 1 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 8-29 Example of the Work with Device Domains panel

This shows the different device domains in the cluster. The options shown on the Work with Device Domains panel are:

- Option 1 invokes the ADDDEVDMNE command. See “Add Device Domain Entry” on page 267 for more information.
- Option 6 will bring you to the Work with Device Domain Nodes panel, as shown in Figure 8-30.

```
                        Work with Device Domain Nodes

Device domain  . . . . . . . . . . . . . . :   ARCTIC
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add   4=Remove   20=Dump trace

Opt     Node         Status

        NODE1        Active
        NODE2        Inactive
        NODE3        Active
        NODE4        Active

                                                                      Bottom
Parameters for option 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 8-30 Example of the Work with Device Domain Nodes panel

The options shown on the Work with Device Domain Nodes panel are:

- Option 1 can be used to add an existing node to this device domain. This is similar to the ADDDEVDMNE command. See “Add Device Domain Entry” on page 267 for more information.
- Option 4 can be used to remove a node from the device domain. This is similar to the RMVDEVDMNE command. See “Remove Device Domain Entry” on page 267 for more information.
- Option 20 invokes the DMPCLUTRC command. See “Dump Cluster Trace” on page 245 for more information.

- Option 7 will work with the switchable hardware, as shown in Figure 8-31. It lists all hardware associated with that device domain.

```
                        Work with Switchable Hardware

Device domain  . . . . . . . . . . . . . . :   ARCTIC
Local node . . . . . . . . . . . . . . . . :   NODE4
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  5=Configuration status

                                      Cluster
        Configuration     Object      Resource       Primary
Opt        Object         Type        Group          Node
         ALASKA           *DEVD       POLARBEAR      NODE1

                                                                      Bottom
Command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 8-31 Example of the Work with Switchable Hardware panel

The option shown on the Work with Switchable Hardware panel is Option 5, which will bring up WRKCFGSTS for the device description shown.

Work with Administrative Domains

Using option 8 off of the WRKCLU OPTION(*SELECT) menu or using WRKCLU OPTION(*ADMDMN) will bring up the Work with Administrative Domains panel. This is illustrated in Figure 8-32.

```
                     Work with Administrative Domains

Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Create   2=Change   4=Delete   6=Nodes   7=Monitored resources   8=Start
  9=End      10=Start job          11=End job   20=Dump trace

     Administrative                            Synchronize            Number
Opt       Domain       Status                  Option                of Nodes

       CANADA         Inactive                Last change                  4

                                                                        Bottom
Parameters for option 1, 2 and 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 8-32 Example of the Work with Administrative Domains

The options shown on the Work with Administrative Domains panel are:

- Option 1 invokes the CRTCAD command. See “Create Cluster Administrative Domain” on page 284 for more information.
- Option 2 invokes the CHGCAD command. See “Change Cluster Administrative Domain” on page 283 for more information.
- Option 4 can be used to delete an administrative domain. This is similar to the DLTCAD command. See “Delete Cluster Administrative Domain” on page 285 for more information.
- Option 6 will bring you to the Work with Administrative Domain Nodes panel, as illustrated in Figure 8-33.

```
                    Work with Administrative Domain Nodes

Administrative domain  . . . . . :   CANADA
Consistent information in cluster  . . . :   Yes
Domain status  . . . . . :   Active

Type options, press Enter.
  1=Add   4=Remove   20=Dump trace

Opt     Node         Status

        NODE1        Active
        NODE2        Inactive
        NODE3        Active
        NODE4        Active

                                                                      Bottom
Parameters for option 20 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 8-33 Example of the Work with Administrative Domain Nodes panel

The options shown on the Work with Administrative Domain Nodes panel are:

- Option 1 invokes the ADDCADNODE command. See “Add Cluster Administrative Domain Node Entry” on page 283 for more information.
- Option 4 invokes the RMVCADNODE command for that node. See “Remove Cluster Administrative Domain Monitored Resource Entry” on page 285 for more information.
- Option 20 invokes the DMPCLUTRC command. See “Dump Cluster Trace” on page 245 for more information.

- Option 7 will bring you to the Work with Monitored Resources panel, as illustrated in Figure 8-34.

```
                        Work with Monitored Resources

Administrative domain  . . . . . :   CANADA
Consistent information in cluster  . . . :   Yes
Domain status  . . . . . :   Active

Type options, press Enter.
  1=Add   4=Remove   5=Display details   7=Display attributes

                Resource                Global          Local
Opt  Resource   Type        Library     Status          Status

     ALASKA     *ASPDEV     QSYS        Inconsistent    Current
     FLIGHTS    *JOBD       QGPL        Consistent      Current
     QRETSVRSEC *SYSVAL                 Consistent      Current
     FLIGHTS1   *USRPRF     QSYS        Inconsistent    Current
     FLIGHTS10  *USRPRF     QSYS        Inconsistent    Current
     FLIGHTS2   *USRPRF     QSYS        Inconsistent    Current
                                                                More...
Parameters for option 1 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve
F11=Order by global status      F12=Cancel   F24=More Keys
```

Figure 8-34 Example of the Work with Monitored Resources panel

The options shown on the Work with Monitored Resources panel are:

- Option 1 invokes the ADDCADMRE command. See “Add Cluster Administrative Domain Monitored Resource Entry” on page 281 for more information.
- Option 4 invokes the RMVCADMRE command. See “Remove Cluster Administrative Domain Monitored Resource Entry” on page 285 for more information.
- Option 5 will display more information about monitored resource details.

– Option 7 will bring you to the Display Monitored Resource Attributes panel, as shown in Figure 8-35.

```
                  Display Monitored Resource Attributes

Resource . . . . . :   ALASKA
  Library  . . . . . :     QSYS
Resource type  . . . . . :   *ASPDEV

Type options, press Enter.
  5=Display values

Opt   Attribute    Global Status     Global Value
      MSGQ         Consistent        QSYS/QSYSOPR
      RDB          Consistent        *GEN
      RSRCNAME     Consistent        ALASKA

Bottom
Command
===>
F1=Help       F3=Exit      F4=Prompt   F5=Refresh   F7=Resource details
F9=Retrieve   F12=Cancel   F13=Work with cluster menu
```

Figure 8-35 Example of the Display Monitored Resource Attributes panel

The option shown on the Display Monitored Resource Attributes panel is Option 5, which will display the attribute details.

- Option 8 invokes the STRCAD command for this specific administrative domain. See “Start Cluster Administrative Domain” on page 287 for more information.
- Option 9 invokes the ENDCAD command for this specific administrative domain. See “End Cluster Administrative Domain” on page 285 for more information.
- Option 10 will start the administrative job on this partition.
- Option 11 invokes the ENDJOB command for this specific administrative domain job. This function is not meant for normal usage but for use when specific PTFs come out that require it to be restarted.
- Option 20 invokes the DMPCLUTRC command. See “Dump Cluster Trace” on page 245 for more information.

Work with Cluster Resource Groups

Using option 9 off of the WRKCLU OPTION(*SELECT) menu or using WRKCLU OPTION(*CRG) will bring up the Work with Cluster Resources groups menu. This is illustrated in Figure 8-36.

```
                      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

          POLARBEAR        *DEV        Active                     NODE1


                                                                    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 8-36 Example of the Work with Cluster Resource Groups panel

The options shown on the Work with Cluster Resource Groups panel are:

- Option 1 invokes the CRTCRG command. See “Create Cluster Resource Group” on page 265 for more information.
- Option 2 invokes the CHGCRG command. See “Change Cluster Resource Group” on page 264 for more information.
- Option 3 invokes the CHGCRGPRI command. See “Change Cluster Resource Group Primary” on page 264 for more information.
- Option 4 deletes that CRG. This is very similar to the DLTCRGCLU command. See “Delete Cluster Resource Group Cluster” on page 266 for more information.
- Option 5 displays the same information (though a slightly different format) as DSPCRGINF (when you specify a specific CRG). For more information see “Display Cluster Resource Group Information” on page 266.

- Option 6 brings you to the Work with Recovery Domain panel, as shown in Figure 8-37.

```
                        Work with Recovery Domain

Cluster resource group . . . . . . . . . . :   POLARBEAR
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

        NODE1        Active         *PRIMARY       *PRIMARY       JUNEAU
        NODE3        Active         *BACKUP  2     *BACKUP  1     KODIAK
        NODE4        Active         *BACKUP  1     *BACKUP  2     KODIAK

                                                                  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 8-37 Example of the Work with Recovery Domain panel

The options shown on the Work with Recovery Domain panel are:

- Option 1 invokes the ADDCRGNODE command. See “Add Cluster Resource Group Node Entry” on page 264 for more information.
- Option 4 invokes the RMVCRGNODE command. See “Remove Cluster Resource Group Node Entry” on page 266 for more information.
- Option 5 displays more information about that nodes recovery domain. See the example in Figure 8-38 for more information.

```
                  Display More Recovery Domain Details

Cluster resource group . . . . . . . . . . :   POLARBEAR
Node identifier  . . . . . . . . . . . . . :   NODE1
Consistent information in cluster  . . . :   Yes

                       More Recovery Domain Details

Data port interface addresses  . . . . . :   10.0.1.11
                                             10.0.2.11
                                             10.0.3.11
                                             10.0.4.11

                                                                  Bottom
Press Enter to continue.

F1=Help   F3=Exit   F5=Refresh   F12=Cancel   F13=Work with cluster menu
```

Figure 8-38 Example of the Display More Recovery Domain Details panel

- Option 20 invokes the DMPCLUTRC command. See “Dump Cluster Trace” on page 245 for more information.

- Option 7 brings you to the Work with Configuration Objects panel, as shown in Figure 8-39.

```
                      Work with Configuration Objects

Cluster resource group . . . . . . . . . . :   POLARBEAR
Consistent information in cluster  . . . :   Yes

Type options, press Enter.
  1=Add   2=Change   4=Remove   5=Configuration status

     Configuration  Object   Device   Device     Vary      Server
Opt   Object Name   Type     Type     Subtype    Online    Ip Address

      ALASKA        *DEVD    *ASP     Primary    *ONLINE   192.168.154.10

                                                                    Bottom
Parameters for options 1 and 2 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
F13=Work with cluster menu
```

Figure 8-39 Example of the Work with Configuration Objects panel

The options shown on the Work with Configuration Objects panel are:

- Option 1 invokes the ADDCRGDEVE command. See “Add Cluster Resource Group Device Entry” on page 268 for more information.
- Option 2 invokes the CHGCRGDEVE command. See “Change Cluster Resource Group Device Entry” on page 268 for more information.
- Option 4 invokes the RMVCRGDEVE command. See “Remove Cluster Resource Group Device Entry” on page 269 for more information.
- Option 5 does a WRKCFGSTS command for that particular object.

- Option 8 invokes the STRCRG command. See “Start Cluster Resource Group” on page 267 for more information.
- Option 9 invokes the ENDCRG command. See “End Cluster Resource Group” on page 266 for more information.
- Option 20 invokes the DMPCLUTRC command. See “Dump Cluster Trace” on page 245 for more information.

Dump Cluster Trace

Using option 20 off of the WRKCLU OPTION(*SELECT) menu or using WRKCLU OPTION(*SERVICE) will invoke the DMPCLUTRC command. See “Dump Cluster Trace” on page 245 for more information.

8.4.2 Cluster Resource Group commands

The following commands in the PowerHA for i LPP effect changes on cluster resource groups and allow information about them to be viewed.

Add Cluster Resource Group Node Entry

The Add Cluster Resource Group Node Entry (ADDCRGNODE) command can be used to add new nodes to an existing cluster resource group.

There are no parameter changes from the previous release.

For the primary-backup type of CRG the node can be added either as a primary, backup, or replicate node. To add a node as the primary, the CRG would have to be inactive. To add a node as a backup or replicate the CRG could also be in active state. If the CRG has a status of active and it has more than one backup node and some backup nodes are active, but not all, then the recovery order might be changed to have active nodes before inactive nodes. For peer CRGs the node can be added as a peer or replicate node.

For peer model cluster resource groups, the node can be added as a peer node or a replicate node. If the cluster resource group has a status of active (10) and a peer node is added, the node will be added as an active access point.

If a node is added to a device CRG without a device entry, the entries must be added with the ADDCRGDEVE command before the CRG can be started. For more information about ADDCRGDEVE see “Add Cluster Resource Group Device Entry” on page 268.

Change Cluster Resource Group

The Change Cluster Resource Group (CHGCRG) command allows changes to some attributes of a CRG. The CRG is changed on all active nodes. Non-active nodes will be changed when they rejoin the cluster.

There are no parameter changes from the previous release. However, the layout has had some modifications.

For the primary-backup type of CRG, changing a node to primary or changing the takeover IP address can only be done when the CRG status is inactive or in doubt. To change an active node to a primary, first change it to become the first backup (if it is not already) and then use the CHGCRGPRI command. See “Change Cluster Resource Group Primary” on page 264 for more information. If the CRG has a status of active and it has more than one backup node and some backup nodes are active, but not all, then the recovery order might be changed to have active before inactive.

If the role of a device CRG is being changed, the ownership of the devices specified in the CRG is switched from the current primary to the nw primary if the current primary has no devices varied on. If any devices *are* varied on, an error message will result. After a switch the devices will *not* be varied on.

For peer-model CRGs the recovery domain role can be changed from peer to replicate, or the reverse, by specifying one or more nodes in the recovery domain. There must be at least one node designated as peer if the CRG is active.

Change Cluster Resource Group Primary

The Change Cluster Resource Group Primary (CHGCRGPRI) command will perform a switchover of the CRG by changing the current roles of nodes. The first backup will become the new primary, the current primary will become the last backup, and any other backup will move up one position (such as backup 2 would become backup 1).

There are no parameter changes from the previous release.

You cannot use this command for peer-model CRGs.

When planning a switchover of multiple CRG types we recommend first moving a device CRG, followed by data CRGs, and then last application CRGs. Applications using the device or data CRG should be ended before the switchover.

If the device CRG entry indicates that the device should be varied on and the vary on fails, then the switchover will not complete successfully, but the devices will now remain on the new primary.

Create Cluster Resource Group

The Create Cluster Resource Group (CRTCRG) command allows us to create a CRG. The CRG is what we can use to control highly available resources. When active there will be a system job running with the same name as the CRG.

There are no parameter changes from the previous release.

The CRG an then be controlled by any node in the cluster, not just nodes that the CRG is a part of.

Clustering must be active on each node that you specify to be a part of this CRG. If clustering configures the takeover IP addresses, then they must all be in the same subnet and the IP address cannot be active already.

If the cluster version is 5 then only auxiliary storage pool devices can be part of a device CRG. At cluster Version 6 or later you can also configure a device CRG without a site name to contain:

- An independent auxiliary storage pool device
- An asynchronous communications device
- A binary synchronous communications device
- A cryptographic device
- A distributed data interface communication line
- An Ethernet communication line
- A facsimile communication line
- A network server device for a guest operating system (Linux) running in a logical partition
- A network server device that uses an iSCSI connection
- An integrated network server device
- A local workstation controller
- A network server host adapter device
- An optical device
- A point-to-point protocol communication line
- A synchronous data link control communication line
- A tape device
- A token-ring line
- A wireless local area network communication line
- An X.25 communication line

Note that all nodes in a device CRG must be in the same device domain. For a device CRG with a site name specified (cross-site mirroring or XSM) you can only specify an ASP device. Remember that a server IP address must be specified if you specify a site name, and the reverse is also true. Devices in a device CRG attached to different IOPs or High Speed Link (HSL) I/O Bridges are grouped in an ASP. All the devices on the affected IOAs and IOPs or HSL I/O Bridges can be specified for only one CRG. If a member of an ASP group is part of a device CRG, than all members of the ASP group need to be part of the device CRG before the CRG can be started. You can add these other iASPs with the ADDCRGDEVE command. See “Add Cluster Resource Group Device Entry” on page 268 for more information.

Delete Cluster Resource Group Cluster

The Delete Cluster Resource Group Cluster (DLTCRGCLU) command will delete a CRG from all of its cluster nodes. Clustering must be active to run this command, but the CRG to be deleted must not be active. The CRG will be deleted on all active nodes, and inactive nodes will have the CRG deleted when they rejoin the cluster.

There are no parameter changes from the previous release.

Be aware that the DLTCRGCLU command will not change ownership of the hardware, so the current owner will be where the hardware stays. This could cause issues when SPCN ownership is also involved.

If clustering configured the takeover IP address for an application CRG and the IP address is not active, then the IP address will be removed by this process. If the address is active the command will not finish successfully.

To delete a CRG when the cluster is not active, use the DLTCRG command. See “Delete Cluster Resource Group” on page 245 for more information.

Display Cluster Resource Group Information

The Display Cluster Resource Group Information (DSPCRGINF) command can be used to display or print information about CRGs. It must be run from a node in the cluster. If the node that the command has called from is active then the information will be current. If the node is not active the information will be the last state of the cluster that this specific node knew.

There are no parameter changes from the previous release.

End Cluster Resource Group

The End Cluster Resource Group (ENDCRG) command will end a specific CRG so that it is no longer highly available.

There are no parameter changes from the previous release.

In device CRGs there are no ownership changes of devices when the CRG is ended. The devices are also left in the same condition. For example, if an iASP is varied on, it will remain varied on.

Ending a peer CRG will end the access point for the cluster resources on all nodes defined as peer.

For an application CRG the IP address will be ended.

Remove Cluster Resource Group Node Entry

The Remove Cluster Resource Group Node Entry (RMVCRGNODE) command is used to remove a node from a CRG. The node being removed does not have to be active at the time it is being removed.

There are no parameter changes from the previous release.

In a primary-backup type CRG, if the CRG has no backup nodes the primary cannot be removed. In this case if you are actually trying to delete the CRG you would want to use the DLTCRGCLU (cluster active) or DLTCRG (cluster not active). See “Delete Cluster Resource Group Cluster” and “Delete Cluster Resource Group” on page 245 for more information.

The recovery lists of nodes will be updated when this command is run. If the primary node is removed, the CRG will switch over to the next node in the list *if* any configuration objects are

not varied on. If you are deleting the primary and configuration objects are varied on, then the command will fail. Also, one node in the recovery domain must be active to become the new primary.

For an application CRG, if clustering configured the takeover IP address, it will be removed.

For a peer model CRG the last node designated as a peer node cannot be removed if the CRG is active.

Start Cluster Resource Group

The Start Cluster Resource Group (STRCRG) command will start a CRG on all of its nodes that are currently active in the cluster. A system job with the same name as the cluster will be started on each node in the cluster that is part of the CRG. This command will only run if the node that invokes it is part of an active cluster.

There are no parameter changes from the previous release.

If a primary-backup type of CRG has a status of active and it has more than one backup node and some backup nodes are active, but not all, then the recovery order might be changed to have active nodes before inactive nodes.

For application CRGs this command will verify that the takeover IP address has been configured on all nodes in the recovery list that are primary or backup nodes, and it will start the IP address on the primary node.

For a device CRG the device descriptions must exist on all active nodes for the CRG and the primary node must have ownership of the hardware associated with the device descriptions. If a member of an ASP group is part of a device CRG then all members of the ASP group need to be part of the device CRG before the CRG can be started.

Note that starting the CRG will not cause device descriptions to vary on or start takeover IP addresses.

For peer type CRGs, nodes with a peer role will be active after this command. At least one node must have a peer role.

8.4.3 Switchable device commands

The following commands in the PowerHA for i LPP allow you to work with and configure switchable devices.

Add Device Domain Entry

The Add Device Domain Entry (ADDDEVDMNE) command can be used to add a cluster node to a device domain. If you use this command to add a node to a device domain that does not exist yet, the domain will be created. When a node is part of a device domain it can be added to a device CRG.

There are no parameter changes from the previous release.

You can run ADDDEVDMNE from any node in the cluster that is active. A node can only be in one device domain at a time. This command will not work if there is a cluster partition.

Remove Device Domain Entry

The Remove Device Domain Entry (RMVDEVDMNE) command can be used to remove a node from a cluster device domain.

There are no parameter changes from the previous release.

After a device is removed from a device domain, it will usually need to be IPLed before it can be added to a new device domain. You can run this command on any active node. This command will fail if you try to remove a node that is still part of a device CRG.

Add Cluster Resource Group Device Entry

The Add Cluster Resource Group Device Entry (ADDCRGDEVE) command can be used to add device descriptions to a device CRG. All devices that are added must be able to switch between nodes.

There are no parameter changes from the previous release.

If a member of an ASP group is part of a device CRG then all members of the ASP group need to be part of the device CRG before the CRG can be started. The different members of the ASP group can be added in separate instances of the command. Clustering must be active on the node where you run this command from.

You cannot have the same device descriptions in more than one CRG. The node listed as the primary must currently own the hardware. The device description must exist on all nodes in the device CRG. The resource name for the device description must be the same on all nodes in the device CRG. Devices on the same IOP/IOA or High Speed link (HSL) bridges can only be specified for one device CRG.

If a server takeover IP address is specified it must exist on all nodes in the device domain if the CRG is active. The takeover IP address must also be unique within the cluster and only associated with one ASP group.

All nodes in the CRG must be active.

You can only add the following device description types to a device CRG:

- An independent auxiliary storage pool device
- An asynchronous communications device
- A binary synchronous communications device
- A cryptographic device
- A distributed data interface communication line
- An Ethernet communication line
- A facsimile communication line
- A network server device for a guest operating system (Linux) running in a logical partition
- A network server device that uses an iSCSI connection
- An integrated network server device
- A local workstation controller
- A network server host adapter device
- An optical device
- A point-to-point protocol communication line
- A synchronous data link control communication line
- A tape device
- A token-ring line
- A wireless local area network communication line
- An X.25 communication line

Change Cluster Resource Group Device Entry

The Change Cluster Resource Group Device Entry (CHGCRGDEVE) command can modify device deceptions in a device CRG. You can change the configuration action to be taken when the CRG is switched over to a backup system.

There are no parameter changes from the previous release.

Clustering must be active on the node running the command. At least one node in the device CRG must be active. If a server takeover IP address is specified it must exist on all nodes in the device domain of the CRG is active. The takeover IP address must also be unique within the cluster and only associated with one ASP group.

Remove Cluster Resource Group Device Entry

The Remove Cluster Resource Group Device Entry (RMVCRGDEVE) command can remove one or more device descriptions from a device CRG. All device descriptions can be removed, but at least one will have to be added before it can be started later. To add a device description you can use the ADDCRGDEVE command. See “Add Cluster Resource Group Device Entry” on page 268 for more information.

There are no parameter changes from the previous release.

Hardware ownership will not be changed by running this command. Clustering must be active on the node where you are running the command. At least one node in the CRG must be iASP-related commands in PowerHA for i LPP

8.4.4 iASP-related commands

In this section we explain the iASP-related commands.

Add Auxiliary Storage Pool Copy Description

The Add Auxiliary Storage Pool Copy Description (ADDASPCPYD) command is used to describe a single physical copy of an independent auxiliary storage pool and assign a name to this copy. So for a switchable iASP, you would not need a copy description, since there is only one copy. With geographic mirroring, there are two copies of the data, so you need two copy descriptions. When using a four-node configuration of geographic mirroring of switchable iASPs you would still only need to copy descriptions, since there are still only two copies of the iASP that go between the four nodes.

The two parameters storage host and location unit name are required if the ASP copy will be used in a metro mirror, global mirror, or FlashCopy session. They specify the storage host name for the ASP copy and the logical units that are associated with the copy description.

The location name must be set to *NONE for all configurations discussed in this book.

ADDASPCPYD is a new command. Figure 8-40 shows an example.

```
                    Add ASP Copy Description (ADDASPCPYD)

Type choices, press Enter.

ASP copy . . . . . .                          Name
ASP device . . . . . .                        Name
Cluster resource group . . . . .   *NONE       Name, *NONE
Cluster resource group site  . .   *NONE       Name, *NONE
Storage host:
  User name  . . . . . .           *NONE
  Password . . . . . .
  Internet address . . . . . .



Location . . . . . .              *NONE       Name, *DEFAULT, *NONE
Logical unit name:
  TotalStorage device  . . . . .   *NONE
  Logical unit range . . . . . .              Character value
              + for more values
                                                                  Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 8-40 Example of ADDASPCPYD

When geographic mirroring is configured the disk units are assigned to a mirror copy, and a copy description for the mirror copy of the iASP can be added either before or after the configuration. In the same fashion, when configuring metro mirror, global mirror, or a FlashCopy target the ASP copy description for it can be added at any time.

Change Auxiliary Storage Pool Copy Description

The Change Auxiliary Storage Pool Copy Description (CHGASPCPYD) command can change an existing iASP copy description. This is how to modify what we created using the ADDASPCPYD command if we change our mind. See “Add Auxiliary Storage Pool Copy Description” for more information.

CHGASPCPYD is a new command. Figure 8-41 shows an example.

```
                Change ASP Copy Description (CHGASPCPYD)

Type choices, press Enter.

ASP copy . . . . . .                    Name
ASP device . . . . . .   *SAME           Name, *SAME
Cluster resource group . . . . .  *SAME  Name, *SAME, *NONE
Cluster resource group site  . .  *SAME  Name, *SAME, *NONE
Storage host:
  User name  . . . . . .  *SAME
  Password . . . . . .
  Internet address . . . . . .



Location . . . . . .  *SAME         Name, *SAME, *DEFAULT, *NONE
Logical unit name:
  TotalStorage device  . . . . .  *SAME
  Logical unit range . . . . . .          Character value, *SAME
              + for more values
                                                              Bottom
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 8-41 Example of CHGASPCPYD

Display Auxiliary Storage Pool Copy Description

The Display Auxiliary Storage Pool Copy Description (DSPASPCPYD) displays an iASP copy description. This is how to view what we created using the ADDASPCPYD command if we change our mind. See “Add Auxiliary Storage Pool Copy Description” for more information.

This is a new command. Figure 8-42 shows an example of the parameter and Figure 8-43 on page 272 shows an example of output.

```
               Display ASP Copy Description (DSPASPCPYD)

Type choices, press Enter.

ASP copy . . . . . .  metroloc      Name
Output . . . . . .    *             *, *PRINT



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

Figure 8-42 Parameters of DSPASPCPYD

```
                    Display ASP Copy Description               NODE4
                                                          05/28/08  17:37:22
ASP copy description . . . . :   METROREM
Device description . . . . . :   IASPMETRO
Cluster resource group . . . :   METROCRG
Cluster resource group site  :   METROREM
Sessions . . . . . . . . . . :   FLASHSSN
User . . . . . . . . . . . . :   admin
  Internet address . . . . . :     9.5.168.55

  Alternate internet address :


                      Display ASP IO Resources                 NODE4
                                                          05/28/08  17:38:09
IO Adapter  . . . . . . . . . . . :   *NONE
Location  . . . . . . . . . . . . :   NODE4
IO Adapter  . . . . . . . . . . . :   *NONE
Location  . . . . . . . . . . . . :   *NONE
TotalStorage device . . . . . . . :   IBM.2107-75AY031

                                LUN ranges

Range
1400-1403
1500-1503



                                                                  Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel
```

Figure 8-43 Display from DSPASPCPYD

Remove Auxiliary Storage Pool Copy Description

The Remove Auxiliary Storage Pool Copy Description (RMVASPCPYD) command can be used to remove an iASP copy description. This is how to delete what we created using the ADDASPCPYD command if we change our mind. See “Add Auxiliary Storage Pool Copy Description” for more information.

This is a new command.

Note that this command will not remove the configured disks.

Work with Auxiliary Storage Pool Descriptions

The Work with Auxiliary Storage Pool Copy Descriptions (WRKASPCPYD) command displays existing iASP copy descriptions and identifies any copy sessions that they have.

This is a new command. Figure 8-44 shows an example of the WRKASPCPYD command.

```
                      Work with ASP Copy Descriptions                NODE5
                                                          12/10/07  14:44:26
Type options, press Enter.
  2=Change copy    4=Remove copy   5=Display copy   22=Change session
  24=End session   25=Display session

        ASP            ASP            ASP               Session
Opt     Device         Copy           Session             Type
        HEAT           HEAT_K         XSM               *GEOMIR
        HEAT           HEAT_P         XSM               *GEOMIR


                                                                    Bottom
Parameters or command
===>
F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve   F12=Cancel
```

Figure 8-44 Example of the Work with Auxiliary Storage Pool Descriptions panel

The following options are offered:

- Option 2 invokes the CHGASPCPYD command. See “Change Auxiliary Storage Pool Copy Description” on page 270 for more information. When you enter option 22, press F4, not just Enter, or you do not get an option of what to change.
- Option 4 invokes the RMCASPCPYD command. See “Remove Auxiliary Storage Pool Copy Description” on page 272 for more information.
- Option 5 invokes the DSPASPCPYD command. See “Display Auxiliary Storage Pool Copy Description” on page 271 for more information.
- Option 22 invokes the CHGASPSSN command. See “Change Auxiliary Storage Pool Session” on page 274 for more information. When you enter option 22, press F4, not just Enter, or you do not get an option of what to change.
- Option 24 invokes the ENDASPSSN command. See “End Auxiliary Storage Pool Session” on page 280 for more information.
- Option 25 invokes the DSPASPSSN command. See “Display Auxiliary Storage Pool Session” on page 277 for more information.

When running the other commands through the WRKASPCPYD interface, the output of errors will not be seen at the bottom of the panel. Once you run a command and it comes back, you should manually run a DSPJOBLOG and check for errors, such as in Figure 8-45. Otherwise, you might have gotten an error and not have known it.

```
                      Additional Message Information

Message ID . . . . . :   CPF9898       Severity . . . . . :   40
Message type . . . . . :   Diagnostic
Date sent  . . . . . :   05/14/08      Time sent  . . . . . :   08:24:43

Message . . . . :   Unable to execute change operation.  Ensure that the
  command is being executed on the target node and try the operation again.
Cause . . . . . :   This message is used by application programs as a general
  escape message.

                                                                      Bottom
Press Enter to continue.

F3=Exit   F6=Print   F9=Display message details   F12=Cancel
F21=Select assistance level
```

Figure 8-45 Example of error

Change Auxiliary Storage Pool Session

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

This is a new command. An example is shown in Figure 8-46. 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 8-46 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 (CHGATTR) 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 a little confusing. It is hard to know which node you have to run which option from as well as what states the iASP copies must be in first. We have provided a quick reference list in Figure 8-47. To simplify the chart we are letting *source* also mean *production copy* and *target* also mean *mirror copy*.

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
*Suspend	Geographic Mirroring	Yes	No	No	Yes
*Resume	Geographic Mirroring	Yes	No	No	Yes

Figure 8-47 What options can be run from where and what status the iASP copies must be in

CHGASPSSN *SUSPEND or *DETACH will offer the tracking *YES or *NO, as shown in Figure 8-48. 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 8-48 CHGASPSSN *DETACH for geographic mirroring*

CHGASPSSN with geographic mirror *will* allow a *DETACH if the iASP is varied on. Since the iASP is in use, this will mean that the mirror copy will have to go through an abnormal vary-on, and there are risks to your data. Using the steps in “Change Auxiliary Storage Pool Activity” on page 247 will help minimize those risks.

Note: When using FlashCopy or geographic mirror detach of a varied on iASP, make sure that you use the steps in “Change Auxiliary Storage Pool Activity” on page 247 to help minimize any risks to your data. Remember that varied off the iASP first is the safest method.

CHGASPSSN with metro mirroring will *not* allow a *DETACH if the iASP is varied on. If you try you will get a CPD26B9, as shown in Figure 8-49.

```
                        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 8-49 msgCPD26B9 example

CHGASPSSN using the *DETACH parameter removes a FlashCopy relation, but it will keep the ASP session. This parameter will not cause a reset/reload of an IOP/IOA.

CHGASPSSN using the *REATTACH parameter recreates a FlashCopy using the parameters in an already created ASP session. This parameter does not cause a reset/reload of an IOP/IOA.

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.

Display Auxiliary Storage Pool Session

The Display Auxiliary Storage Pool Session (DSPASPSSN) command will display an independent auxiliary storage pool session.

This is a new command. Four examples are shown below starting with Figure 8-50.

For geographic mirroring see Figure 8-50. The production copy will show the state of the iASP device description under STATUS. The mirror copy, on the other hand, will show the state of the geographic mirroring environment instead of showing the mirror copy device description status. In this example geographic mirroring was active.

```
                          Display ASP Session                    NODE5
                                                        12/10/07  11:32:19
Session . . . . . . . . . . . . . . . . . . :   XSM
  Type  . . . . . . . . . . . . . . . . . . :     *GEOMIR
  Mode  . . . . . . . . . . . . . . . . . . :     ASYNC
  Suspend timeout . . . . . . . . . . . . . :     0
  Synchronization priority  . . . . . . . . :     *MEDIUM
  Track space . . . . . . . . . . . . . . . :     5

                                                                  Bottom
                             Copy Descriptions

ASP           ASP                                      Data
Device        Copy              Role        State      State    Node
HEAT          HEAT_K       PRODUCTION     AVAILABLE    USABLE   NODE5
HEAT          HEAT_P          MIRROR        ACTIVE    USABLE   NODE7

                                                                  Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel
```

Figure 8-50 Example of the geographic mirroring production copy Display ASP Session panel

The mirror copy side will not know the state of the production copy and will show UNKOWN. This is normal. See Figure 8-51. However, DSPASPSSN should show the geographic mirroring status correctly.

```
                          Display ASP Session                    NODE7
                                                        12/10/07  11:32:47
Session . . . . . . . . . . . . . . . . . . :   XSM
  Type  . . . . . . . . . . . . . . . . . . :     *GEOMIR
  Mode  . . . . . . . . . . . . . . . . . . :     ASYNC
  Suspend timeout . . . . . . . . . . . . . :     0
  Synchronization priority  . . . . . . . . :     *MEDIUM
  Track space . . . . . . . . . . . . . . . :     5

                                                                  Bottom
                            Copy Descriptions

ASP           ASP                                        Data
Device        Copy               Role         State      State    Node
HEAT          HEAT_K          UNKNOWN     AVAILABLE     USABLE    NODE5
HEAT          HEAT_P           MIRROR        ACTIVE     USABLE    NODE7

                                                                    Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel
```

Figure 8-51 Example of the geographic mirroring mirror copy Display ASP Session panel

For metro mirroring see Figure 8-52. The source copy will show the state of the iASP device description under STATUS. The target copy, on the other hand, will show the state of the metro mirroring environment, instead of showing the mirror copy device description status. In this example metro mirroring was active.

```
                          Display ASP Session                    RCHASJAZ
                                                        05/12/08  08:55:07
Session . . . . . . . . . . . . . . . . . . :   METRO
  Type  . . . . . . . . . . . . . . . . . . :     *METROMIR

                                                                  Bottom
                            Copy Descriptions

ASP                                                          Data
device        Name               Role         State        State    Node
METRO         METROS2          SOURCE     AVAILABLE       USABLE    RCHASJAZ
METRO         METROS1          TARGET        ACTIVE     UNUSABLE    ZG2OP1

                                                                    Bottom
Press Enter to continue

F3=Exit   F5=Refresh   F12=Cancel
```

Figure 8-52 Example of the metro mirroring Source copy display ASP Session panel

The target copy side will not know the state of the source copy and will show UNKOWN. This is normal. See Figure 8-53. However, DSPASPSSN should show the Metro mirroring status correctly.

```
                          Display ASP Session                    ZG2OP1
                                                              05/12/08  07:55:28
Session . . . . . . . . . . . . . . . . . . :   METRO
  Type  . . . . . . . . . . . . . . . . . . :     *METROMIR

                                                                        Bottom
                            Copy Descriptions

ASP                                                              Data
device        Name               Role         State           State    Node
METRO         METROS2          SOURCE       UNKNOWN         USABLE    RCHASJAZ
METRO         METROS1          TARGET        ACTIVE       UNUSABLE    ZG2OP1

                                                                        Bottom
Press Enter to continue

F3=Exit    F5=Refresh    F12=Cancel
```

Figure 8-53 Example of the metro mirroring target copy Display ASP Session panel

During a CHGCRGPRI the role and state will shown UNKOWN. When the switchover is done, then they will report information again.

End Auxiliary Storage Pool Session

The End Auxiliary Storage Pool Session (ENDASPSSN) command will end an existing iASP session. This is a new command. Figure 8-54 shows an example.

```
                        End ASP Session (ENDASPSSN)

Type choices, press Enter.

Session  . . . . . . . . . . . . >              Name

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

Figure 8-54 Example of ENDASPSSN command

Start Auxiliary Storage Pool Session

The Start Auxiliary Storage Pool Session (STRASPSSN) command assigns a name to geographic mirroring, metro mirror, global mirror, and FlashCopy sessions and starts High Availability Solutions Manager (HASM) sessions for them. A geographic mirroring session will exist on the node from the time that the geographic mirroring mirror copy of the iASP is created. Metro mirror, global mirror, or FlashCopy sessions exist in TotalStorage® from the time that they are configured in the TotalStorage devices. What this command does is assign a name to that session that will allow the session to be used by other commands.

This is a new command. Figure 8-55 gives an example.

```
                        Start ASP Session (STRASPSSN)

Type choices, press Enter.

Session  . . . . . .                    Name
Session type . . . . . .                *GEOMIR, *METROMIR...
ASP copy:
  Preferred source . . . . . .          Name
  Preferred target . . . . . .          Name
  Consistency source . . . . . .  *NONE Name, *NONE
  Consistency target . . . . . .  *NONE Name, *NONE
              + for more values

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

Figure 8-55 Example of the STRASPSSN command

HASM sessions allow HASM to manage and monitor the activity of the iASPs. The HASM sessions are named so that users and HASM can identify to one another which session they are referring too.

A session will exist until it is ended even if the session's operation has completed. This is done so that information about the session can be retrieved at a later time.

All FlashCopy operations must be performed from the FlashCopy target. Any iASP can be the source of a FlashCopy operation. The following items cannot be a target of FlashCopy operations:

- A geographic mirroring production or mirror copy, whether or not it is detached.
- A metro mirror target.
- A global mirror target.
- A target of some other FlashCopy cannot double as a target of another FlashCopy.

For geographic mirroring the suspend time-out, mirroring mode, synchronization priority, and tracking space parameters will come up. However, these are ignored through the command and will be taken from the settings used when geographic mirroring was set up through one of the GUI interfaces.

8.4.5 Administrative domain commands

The following commands in the PowerHA for i LPP effect changes on cluster administrative domains and allow information about them to be viewed.

Add Cluster Administrative Domain Monitored Resource Entry

The Add Cluster Administrative Domain MRE (ADDCADMRE) command can be used to add a monitored resource entry (MRE) to an administrative domain. A monitored resource is a system object or set of attributes not associated with a specific system object, such as a set of system environment variables. A resource represented by an MRE is watched for changes by an administrative domain. If the MRE is changed on a non-replicate node in the administrative domain the change will also be made to the other active nodes.

This is a new command. Figure 8-56 gives an example.

```
                    Add Admin Domain MRE (ADDCADMRE)

Type choices, press Enter.

Cluster  . . . . . .                  Name
Cluster administrative domain  .       Name
Monitored resource . . . . . .         Character value
Monitored resource type  . . . .       *ASPDEV, *CLS, *ENVVAR...
Library  . . . . . .                   Name
Monitored attributes . . . . . .   *ALL  Name, *ALL
              + for more values

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

Figure 8-56 Example of the ADDCADMRE command

When you add a MRE for a system object the resource name will be the name of the system object. You can specify one or more attributes to be monitored. This command:

- Creates a monitored resource entry on all nodes in the administrative domain.
- The attributes for the created resource will be set to the value of the attributes from the monitored resource on the node from which the command was called.
- If the administrative domain has been started, then the values of the attributes for the MRE will be synchronized. If the administrative domain has not been started then the values of the MRE will not be synchronized until it is started.

Clustering must be active on the node running the command. Resources cannot be added to an administrative domain when it is partitioned. The MRE must exist on the node from where the command is issued. The user profile of the person running the command must exist on all nodes in the administrative domain. The system value QRETSRSEC must be set to 1 to monitor any secure attribute. The command can only be run from a node in the administrative domain.

All MREs must exist in the system auxiliary storage pool, which is ASP 1. The single exception is for network server storage spaces (*NWSSTG), which must exist in an independent auxiliary storage pool.

Object types that can be monitored resources with i 6.1 are:

- 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)
- 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)

- 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)

You can find more information about this in the Information Center by following the path **Availability → High availability technologies → i5/OS Cluster technology → Cluster concepts → Cluster version**.

Add Cluster Administrative Domain Node Entry

The Add Cluster Administrative Domain Node Entry (ADDCADNODE) command can be used to add a new node into a existing cluster administrative domain. All MREs that have been defined for the administrative domain will be added on the specified node. Any resource that does not already exit will be created.

This is a new command. See Figure 8-57 for an example of this command.

```
                  Add Admin Domain Node Entry (ADDCADNODE)

Type choices, press Enter.

Cluster  . . . . . .                Name
Cluster administrative domain  .     Name
Node identifier  . . . . . .         Name

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

Figure 8-57 Example of the ADDCADNODE command

If the administrative domain is active the monitored resources will be synchronized with the active domain from the node where the command was called from.

The cluster must be active on the node running the command. At least one node in the administrative domain must be active in the cluster. The node being added to the administrative domain must also be active.

Change Cluster Administrative Domain

The Change Cluster Administrative Domain (CHGCAD) command can change the settings of an existing cluster administrative domain.

This is a new command. Figure 8-58 gives an example.

```
                    Change Cluster Admin Domain (CHGCAD)

Type choices, press Enter.

Cluster  . . . . . .                    Name
Cluster administrative domain  .         Name
Synchronization option . . . . .   *SAME  *SAME, *LASTCHG, *ACTDMN

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

Figure 8-58 Example of the CHGCAD command

The cluster must be active on the node running the command. All nodes in the administrative domain must be active in the cluster.

Create Cluster Administrative Domain

The Create Cluster Administrative Domain (CRTCAD) command creates a peer CRG for the cluster administrative domain. The administrative domain will provide synchronization of monitored resources through the nodes that are active.

This is a brand new command at V6R1M0 replacing a similar command at V5R4M0 called CRTADMDMN. Figure 8-59 gives an example of the CRTCAD command.

```
                    Create Cluster Admin Domain (CRTCAD)

Type choices, press Enter.

Cluster  . . . . . .                    Name
Cluster administrative domain  .         Name
Admin domain node list . . . . .         Name
             + for more values
Synchronization option . . . . .   *LASTCHG      *LASTCHG, *ACTDMN

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

Figure 8-59 Example of the CRTCAD command

The CRG name will be the same as the administrative domain. If the CRG creates successfully it will start a system job with the same name as the CRG. The command will:

- Create the cluster administrative domain on all nodes in the defined administrative domain. The administrative domain can be worked with by a CRG command on any node in the cluster.
- When active, any changes made to the monitored resources will be made to all nodes in the administrative domain.
- The CRG will be owned by the QCLUSTER user profile.

The cluster must be active on the node running the command. All nodes in the administrative domain must be active in the cluster.

Delete Cluster Administrative Domain

The Delete Cluster Administrative Domain (DLTCAD) command will delete a CRG associated with a cluster administrative domain from all cluster nodes in the administrative domain.

This is a brand new command at V6R1M0 replacing a similar command at V5R4M0 called DLTADMDMN. Figure 8-60 gives an example of the DLTCAD command.

```
                    Delete Cluster Admin Domain (DLTCAD)

Type choices, press Enter.

Cluster  . . . . . .                Name
Cluster administrative domain  .     Name


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

Figure 8-60 Example of the DLTCAD command

The cluster must be active on the node running the command. The CRG being deleted must not be active.

End Cluster Administrative Domain

The End Cluster Administrative Domain (ENDCAD) command will disable synchronization in a cluster administrative domain. The administrative domain status will change to inactive. When the cluster administrative domain is ended, any changes made to MREs are set to a pending status and will be synchronized on all active nodes in the domain when it is restarted. This is a new command (Figure 8-61).

```
                   End Cluster Admin Domain (ENDCAD)

 Type choices, press Enter.

 Cluster  . . . . . > REDBOOK       Name
 Cluster administrative domain  .   redbookadm    Name



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

Figure 8-61 Example of the ENDCAD command

Remove Cluster Administrative Domain Monitored Resource Entry

The Remove Cluster Administrative Domain Monitored Resource Entry (RMVCADMRE) command will remove a Monitored Resource Entry (MRE) from an administrative domain.

This is a new command (Figure 8-62).

```
                    Remove Admin Domain MRE (RMVCADMRE)

Type choices, press Enter.

Cluster  . . . . . .                    Name
Cluster administrative domain  .         Name
Monitored resource . . . . . .           Character value
Monitored resource type  . . . .         *ASPDEV, *CLS, *ENVVAR...
Library  . . . . . .                     Name

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

Figure 8-62 Example of the RMVCADMRE command

The command:

- Removes the MRE from all nodes in the cluster administrative domain.
- Any system objects that were created or any system environment variables that were added when the monitored resource was added will not be deleted.

The cluster must be active on the node running the command. One node in the administrative domain must have a status of active. Resources cannot be removed from a cluster administrative domain when it is partitioned. The user profile being used must exist on all nodes in the cluster administrative domain. The command must be called from a node in the cluster administrative domain.

Remove Cluster Administrative Domain Node Entry

The Remove Cluster Administrative Domain Node Entry (RMVCADNODE) command can be used to remove a node from an administrative domain. The node being removed does not have to be active in the cluster. The MREs are removed from the node that is being removed.

This is a new command (Figure 8-63).

```
                Remove Admin Domain Node Entry (RMVCADNODE)

Type choices, press Enter.

Cluster  . . . . . .                    Name
Cluster administrative domain  .         Name
Node identifier  . . . . . .             Name

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

Figure 8-63 Example of the RMVCADNODE command

The cluster must be active on the node running the command. One node in the administrative domain must have a status of active. The last node in the domain list cannot be removed if the cluster administrative domain is active.

Start Cluster Administrative Domain

The Start Cluster Administrative Domain (STRCAD) command will start a monitored resource synchronization for the specified domain. This is a new command (Figure 8-64).

```
                     Start Cluster Admin Domain (STRCAD)

Type choices, press Enter.

Cluster  . . . . . .                 Name
Cluster administrative domain  .     Name


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

Figure 8-64 Example of the STRCAD command

One node in the administrative domain must have a status of active.

8.5 Cluster commands in QUSRTOOL

Since most of the cluster commands have moved from QSYS to an LPP at V6R1M0, you might ask whether this is going to require someone who does not need the new function to have to buy the LPP just to keep the command-line interface. If you do not need the new enhancements, you can still get the R540 commands at the new release with all of the old functions. The source for the cluster resource commands is in the QUSRTOOL library.

The member TCSTINFO in the QUSRTOOL/QATTINFO file contains information about these commands. The source for an example CRG exit program is also included.

To unpack the previous release's commands to a CLUR540 Library run:

```
CALL QUSRTOOL/UNPACKAGE ('*ALL ' 1)
CRTLIB CLUR540 TEXT('R540 Cluster Commands')
CRTCLPGM PGM(CLUR540/TCSTCRT) SRCFILE(QUSRTOOL/QATTCL)
CALL CLUR540/TCSTCRT ('CLUR540 ')
```

9

Migration

This chapter provides information about the steps that are necessary to migrate an existing clustering environment from IBM i 5.4 to IBM i 6.1 in a way that interrupts your production process the least, as well as ensures that you can use all the new functions that are available with the new release.

9.1 Migrating a geographic mirroring environment

Let us assume that you are running a geographic mirroring environment with IBM i 5.4 and want to do a release upgrade to IBM i 6.1. You would want to make sure that this process gives the least disruption to your production environment and afterwards provides you with all the new functions that PowerHA for i offers. The following sections outline how this can be achieved.

The first decision to make is how you want to do the release upgrade itself. If your business can afford to power down your system then you could just upgrade both systems to the new version of the operating system at the same time. If you need to run continuos operations you would have to do a so-called rolling upgrade. During this process either the production or the backup system is available to the user. You should be aware though that there are periods of time where you do not have a secondary system that you could fail over to should your current production system fail.

9.1.1 Doing a rolling upgrade in a geographic mirroring 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. Before starting the upgrade of the backup system you would want to perform a suspend with tracking for your iASPs. This ensures that geographic mirroring knows that you want to take down the backup system and uses the source side tracking space to track changed pages in the iASP until the backup system is available again, thus eliminating the need for a full resynchronization.

You could then perform the release upgrade on the backup system and install the new PowerHA for i license program (5761-HAS) and all required PTFs. After starting up the system and the clustering environment again you have to resume geographic mirroring to move changes that have occurred on your production system while the backup system was not available to your backup iASP. Once this partial synchronization is finished you can find a convenient point in time to do a controlled switchover to the secondary system.

You could then start production on that system, again suspend geographic mirroring with tracking, and perform the release upgrade on your primary system. Once this is done you would restart the primary system and clustering and resume geographic mirroring from the secondary system to the primary system. Depending on your system setup you would then switch back to your original primary system after the partial synchronization is finished (if your secondary system provides less performance than you primary system) or keep on working on the secondary system (if both systems are identical).

After these steps are done both systems are now running with the new version of the operating system and the new PowerHA for i license program is installed. There are, however, some additional steps that need to be performed to also migrate your cluster environment to the new functionality.

9.1.2 Upgrading your cluster environment to the new release

After doing the release upgrade on both systems your cluster environment will still run with the old functionality. The cluster version is not automatically upgraded. It will still be on Version 5 (which corresponds to IBM i 5.4). You can check your cluster version by either doing a DSPCLUINF or by showing cluster properties in either the management central part of

iSeries Navigator or in the cluster resource services part of the new IBM System Director Navigator.

The upgrade to the new cluster version can be done either by issuing a CHGCLUVER CLUSTER(clustername) or by using any of the graphical interfaces. This command can be issued during normal operations. You do not have to end clustering to perform it. Afterwards, your cluster version should be Version 6, as shown in Figure 9-1.

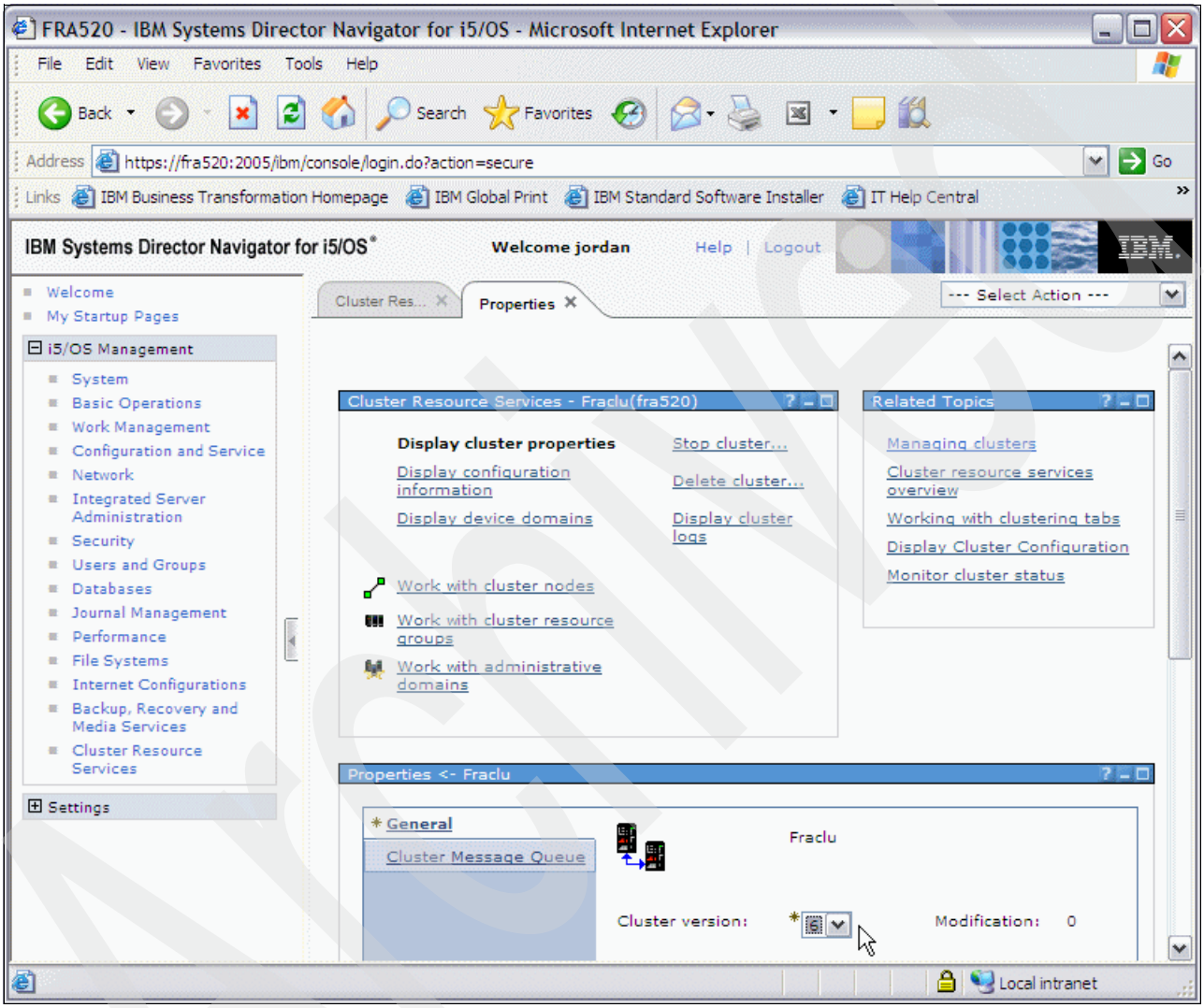


Figure 9-1 Show cluster version

You are now ready to perform the manual steps needed to provide you with the XSM objects that were introduced by PowerHA for i.

9.1.3 Creating new PowerHA for i objects for XSM

The next steps you must perform include the creation of copy descriptions and sessions for your geographic mirroring environment. Again, these steps can be performed using either 5250 commands, iSeries Navigator, or IBM System Director Navigator. The example here assumes that you are using the new graphical interface, for example, IBM System Director Navigator. The implementation that will be migrated here consists of a primary and a secondary iASP that are both controlled by one CRG.

1. In the IBM System Director Navigator for i5/OS window (Figure 9-2), under i5/OS Management select **Configuration and Services**. Then in the right pane click **Disk Pools**.

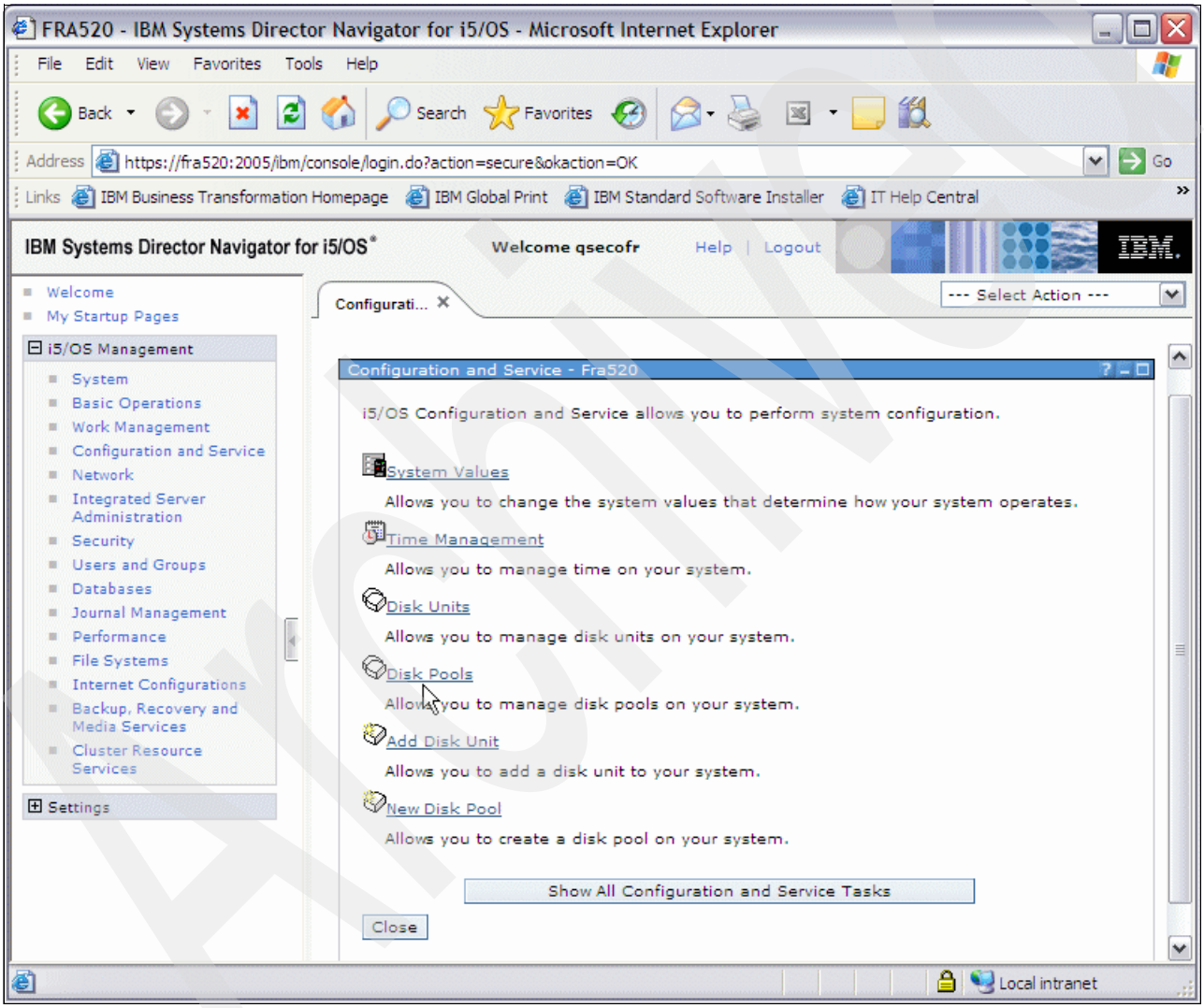


Figure 9-2 Starting the disk GUI

2. If your environment consists of multiple iASPs that are controlled by one single CRG, then you check the boxes next to all your disk pools. Then from the Select Action list (Figure 9-3) select the desired action and click **Go**.

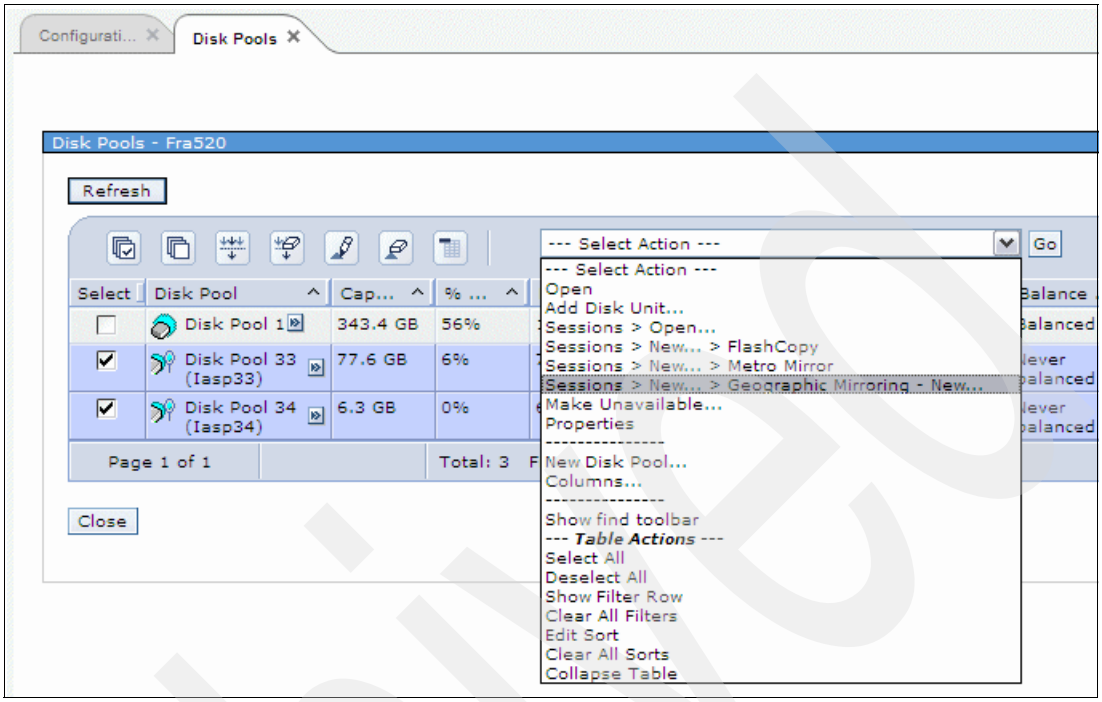


Figure 9-3 Create session for geographic mirroring for multiple iASPs

3. Alternatively, if you have just one iASP, you can also select the double arrow besides the disk pool and then select **Sessions** → **New** → **Geographic Mirroring - New**, as shown in Figure 9-4.

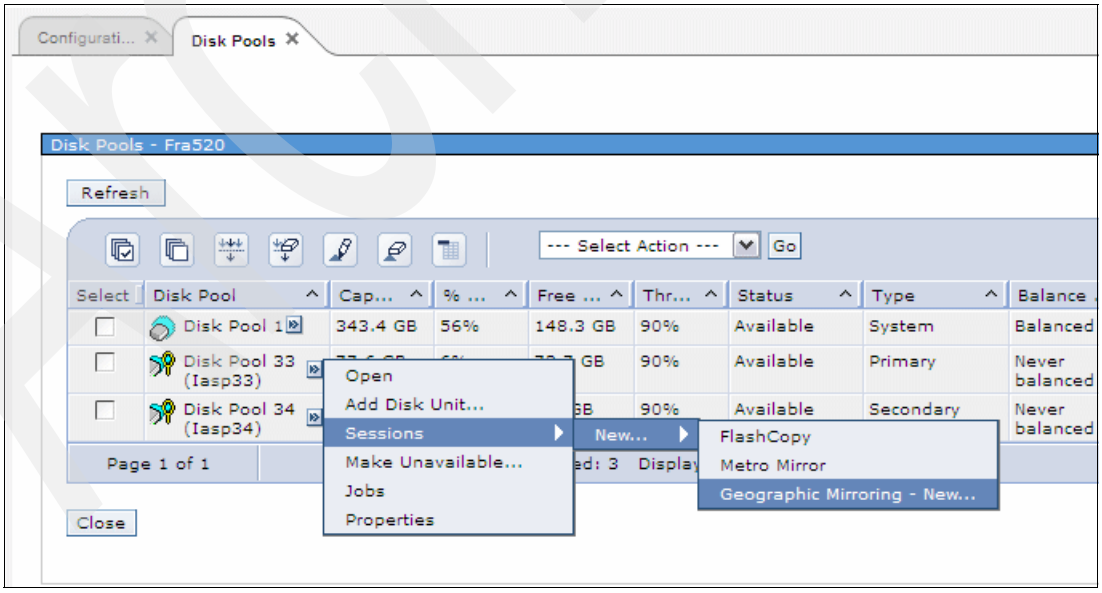


Figure 9-4 Create new session for single iASP

4. Both methods present you with a welcome panel to the new session wizard, as shown in Figure 9-5. This wizard guides you through the steps necessary to create the required copy and session descriptions. You cannot make any selection on this panel. Simply click **Next**.

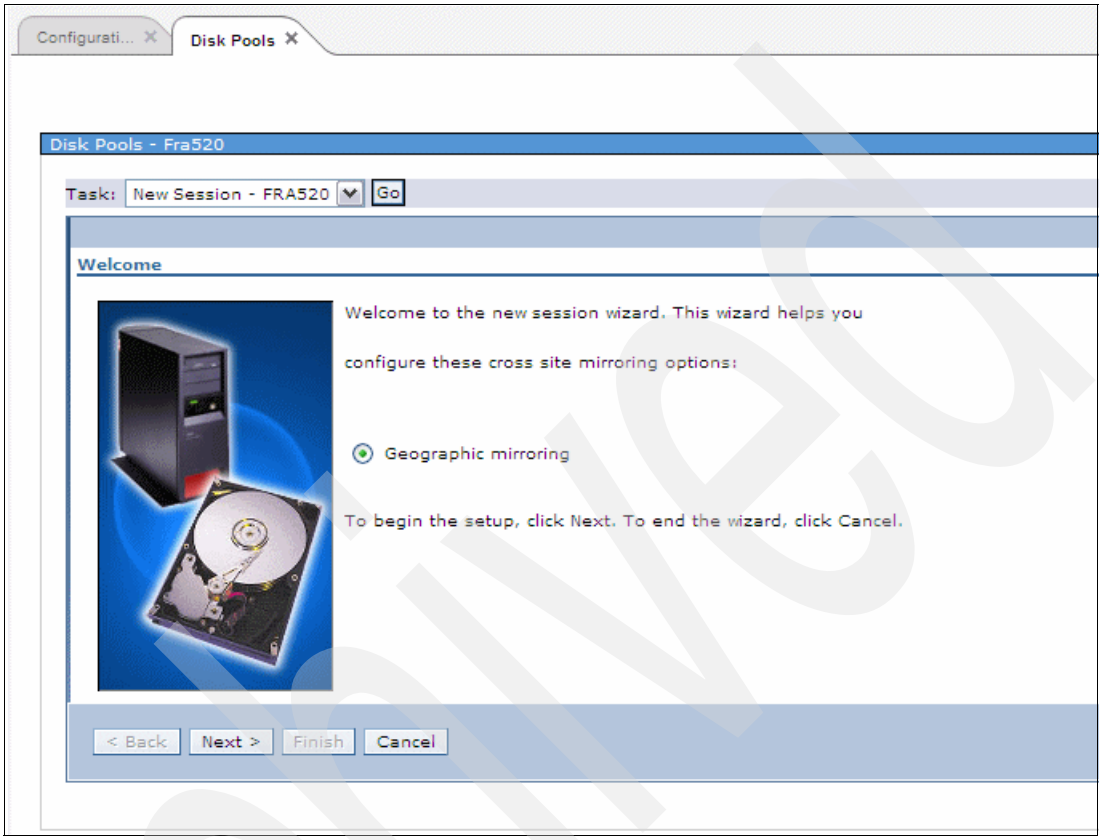


Figure 9-5 Welcome window for session creation wizard

5. On the local copy description page (shown in Figure 9-6), if you already had existing copy descriptions associated with your iASP they are shown in the list. As you are just about to set up your environment, your list should be empty. Click **Add Copy Description**.

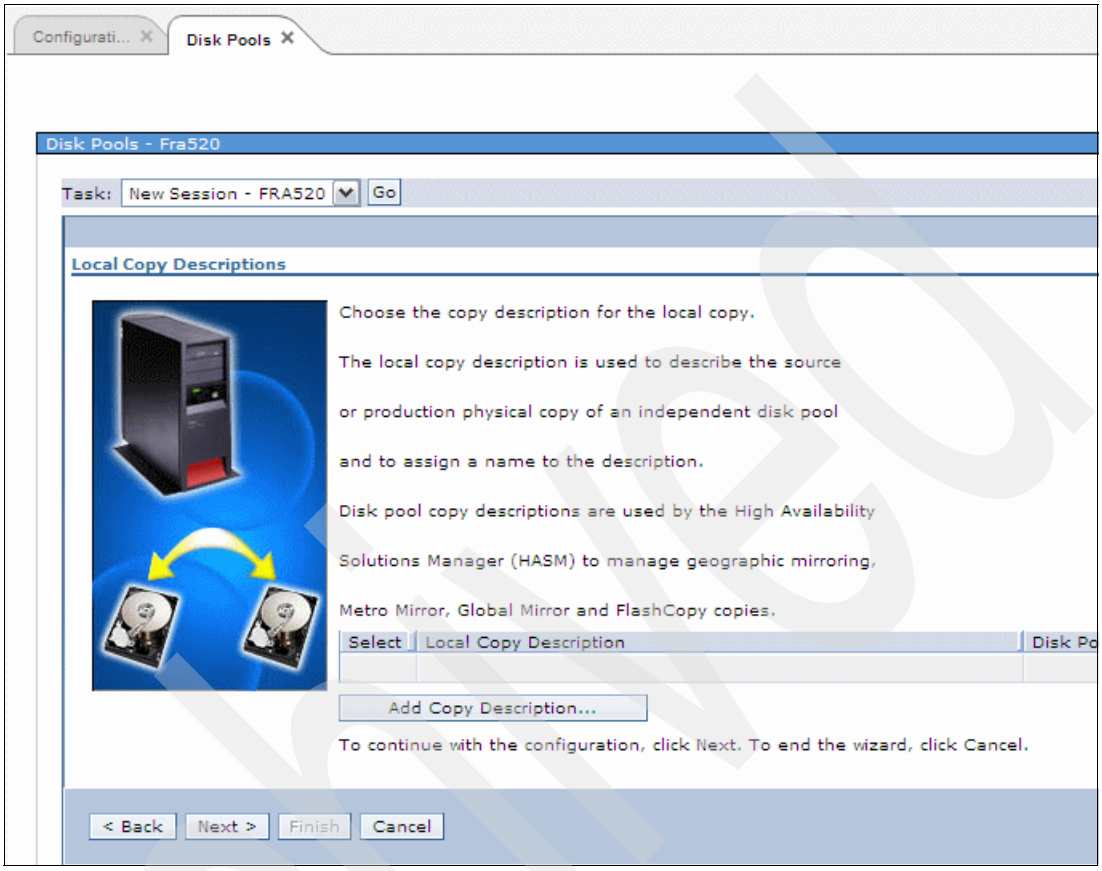


Figure 9-6 Add copy description for primary system

6. On the next panel, fill in the details for the copy description. The copy description requires a name, and you must fill in the name of the CRG that is associated with the iASP that you are currently working on as well as the site name for the current primary node from that CRG, as shown in Figure 9-7.

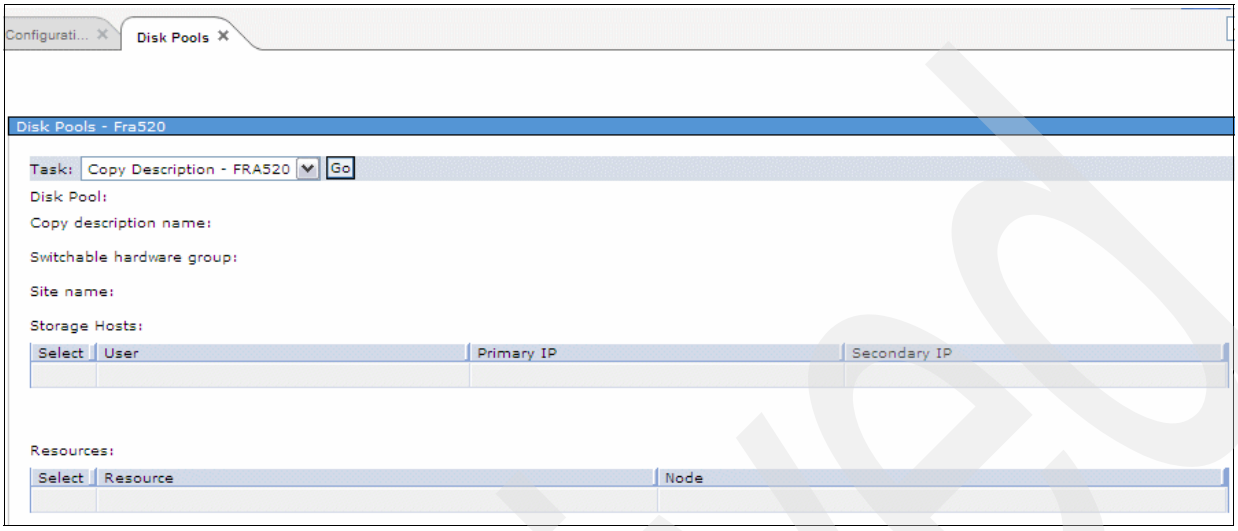


Figure 9-7 Details for copy description

7. Clicking **OK** takes you back to the Local Copy Description panel that we saw in Figure 9-6 on page 295. This time you will see your newly created copy description in the list. Make sure to check the radio button next to the copy description before clicking **Next**. This will present you with basically the same panel, but this now states `remote copy`, as shown in Figure 9-8.

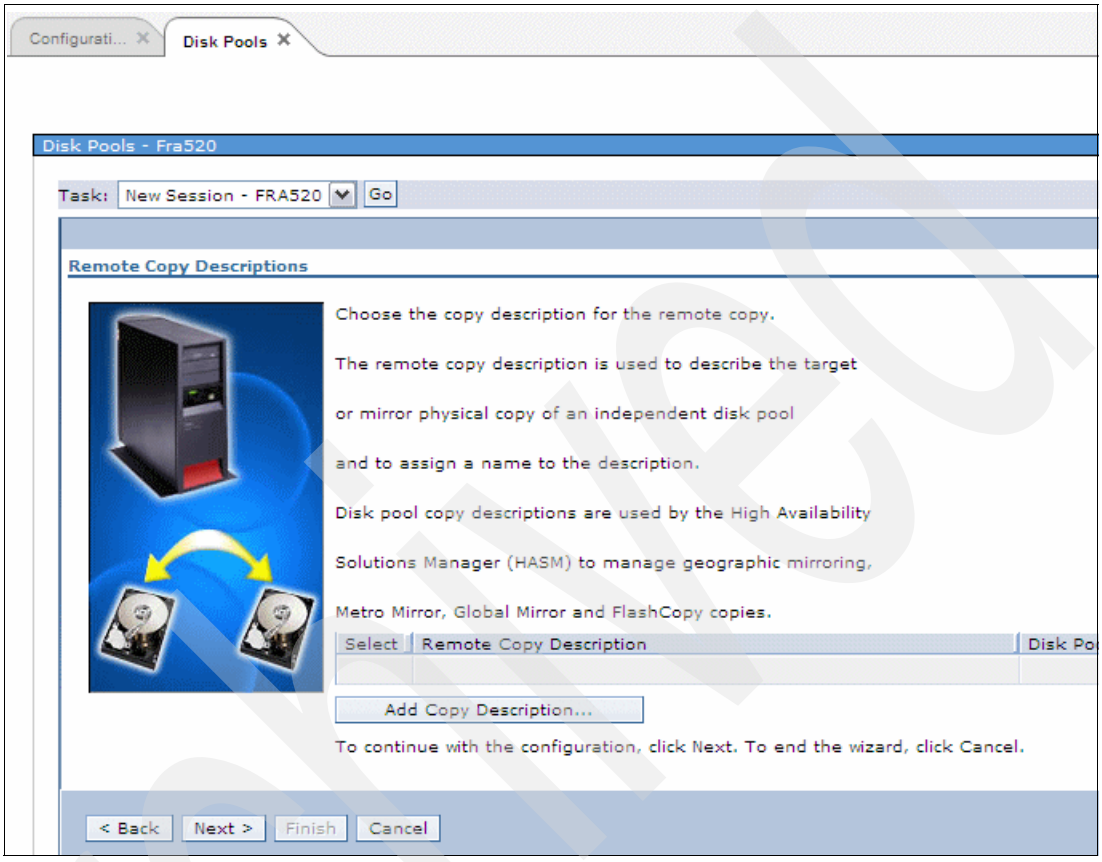


Figure 9-8 Add copy description remote

8. By clicking **Add Copy Description** you are then presented with the panel that allows you to put in the copy description details again. Make sure that you specify the correct site name for your remote system here. Again, you are presented with the panel shown in Figure 9-8. Only now it will show the copy description that you just created. Again make sure to check the radio button next to the newly created copy description before pressing **Next**.

On the panel shown in Figure 9-9 you are asked to provide a session name for the connection between these copy descriptions.

Figure 9-9 Create session description

9. The next panel does provide you with an overview of how your session will be created. It shows you the copy descriptions with their parameters. Recheck that everything is correct to create the session description. Once this is done, your migration to the new release and the new functionality is finished.

Be aware that you must use the approach shown in Figure 9-3 on page 293 if you have an environment with multiple iASPs being controlled by only one CRG. This approach provides you with a way to select all iASPs and then to choose Session → New → Geographic Mirroring → New as the desired action. Make sure to play close attention to the information about the panels that follow. They will prompt you for copy descriptions for all your iASPs on all your nodes. It is your responsibility to make sure that you enter the data correctly.

9.1.4 Doing the upgrade while retaining the old production system intact

If you do not have a test environment available to make sure that all your applications run without any problems after performing a release upgrade you might consider some alternative steps to achieve the upgrade in a cluster environment using geographic mirroring. This method allows you 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. This is not possible with the scenarios described in 9.1.1, “Doing a rolling upgrade in a geographic mirroring environment” on page 290. The reason for this is that once an iASP is varied on to a higher release it cannot be varied on to a lower release again.

The first steps for doing this secure release upgrade are identical to what is described in 9.1.1, “Doing a rolling upgrade in a geographic mirroring environment” on page 290:

1. Suspend geographic mirroring from your production system
2. End the cluster node on the backup system.
3. Perform the operating system upgrade on the backup system.
4. Once this system is started again with the new level of operating system code and clustering is restarted, resume geographic mirroring from the production system and wait for the resynchronization to finish.
5. Next you DETACH geographic mirroring from the production system. This gives you the opportunity to vary on the iASP on the backup system without switching clustering to that system. Make sure that your users can *only* reach one copy of your application environment. Otherwise, you might end up with inconsistent data on both systems.

   If you find any problems with running your application with the new level of the operating system you can simply route your users back to the original production system that is still running the previous version of the operating system. You can use the backup system to fix the problems.

6. Once you are content that your applications run correctly with the new release you have to make sure to change your cluster environment to match the current situation. Remember that from a cluster perspective the primary system is still the system running the old version of the operating system, whereas the backup system runs the new version of the operating system (and due to user activity holds the most recent application data). You therefore must tell clustering that the roles are changed. This is dome using the CHGCRG command, as shown in Figure 9-10. Notice that the recovery domain action is set to *CHGCUR and that the roles of the two nodes are changed.

```
                Change Cluster Resource Group (CHGCRG)

Type choices, press Enter.

Cluster  . . . . . . . . . . . . > REDBOOK       Name
Cluster resource group . . . . . > REDBOOKCRG    Name
Cluster resource group type  . . > *DEV          *DATA, *APP, *DEV, *PEER
CRG exit program . . . . . . . .   *SAME         Name, *SAME, *NONE
  Library  . . . . . . . . . . .                 Name, *CURLIB
Exit program format name . . . .   *SAME         *SAME, EXTP0100, EXTP0200
Exit program data  . . . . . . .   *SAME



User profile . . . . . . . . . .   *SAME         Name, *SAME, *NONE
Text description . . . . . . . .   *SAME

Recovery domain action . . . . . > *CHGCUR       *SAME, *CHGPREFER, *CHGCUR
Recovery domain node list:
 Node identifier  . . . . . . . > NODE3         Name, *SAME
 Node role  . . . . . . . . . .> *BACKUP       *SAME, *BACKUP, *PRIMARY...
 Backup sequence number . . . .   *SAME         1-127, *SAME, *LAST
 Site name  . . . . . . . . . .  *SAME         Name, *SAME, *NONE
 Data port IP address action  .   *SAME         *SAME, *ADD, *REMOVE
 Data port IP address . . . . .   *SAME
           + for more values

 Node identifier  . . . . . . .   NODE4         Name, *SAME
 Node role  . . . . . . . . . .  *PRIMARY      *SAME, *BACKUP, *PRIMARY...
 Backup sequence number . . . .   *SAME         1-127, *SAME, *LAST
 Site name  . . . . . . . . . .  *SAME         Name, *SAME, *NONE
 Data port IP address action  .   *SAME         *SAME, *ADD, *REMOVE
 Data port IP address . . . . .   *SAME
           + for more values
```

Figure 9-10 CHGCRG command to change current role in cluster

7. While still running in detached mode you can now upgrade the operating system of your original production environment. Once this is finished and clustering is restarted you can than perform a REATTACH of geographic mirroring. Note that this starts a full synchronization of your iASP from the backup system to the production system. Once this is finished you can either decide to keep working on the original backup system or switch back to the original production system using the CHGCRGPRI command. You still have to perform the steps outlined in 9.1.2, “Upgrading your cluster environment to the new release” on page 290, and 9.1.3, “Creating new PowerHA for i objects for XSM” on page 292, to complete the upgrade of your cluster environment.

9.2 Migrating a switched disk environment

If you want to migrate a switch disk environment from IBM i 5.4 to IBM i 6.1 there is no need to create copy descriptions or ASP session descriptions. All you would have to do in this environment is to perform the release upgrade on both systems, install additional required license programs and PTFs, and the promote your cluster environment to the current version by using the CHGCLUVER command, iSeries Navigator, or IBM System Director Navigator.

9.3 Migrating from using the Copy Services Toolkit

If you are using the Copy Services Toolkit with IBM i 5.4 and want to migrate to IBM i 6.1 using the new PowerHA for i license program, IBM Systems and Technology Group Lab Services is available to assist you with this. Go to the following Web site for more information and to contact them:

http://www-03.ibm.com/systems/services/labservices/

10

Sizing considerations for geographic mirroring

Geographic mirroring, when used with IBM i cluster technology, provides a high availability solution where your production independent auxiliary storage pool (iASP) data is mirrored to a backup iASP that is attached to another, remote system.

This chapter provides sizing considerations for geographic mirroring.

Among the topics that we will be discussing are:

- Communication requirements
- Network topology
- Backup considerations
- CPU considerations
- Machine pool size considerations
- Disk unit considerations

10.1 How geographic mirroring works

XSM provides the ability to replicate changes made to the production copy of an iASP to a mirror copy of that iASP. As data is written to the production copy of an iASP, the operating system mirrors that data to a second copy of the iASP on another system. This process keeps multiple identical copies of the data.

Changes written to the production copy on the source system are guaranteed to be made in the same order to the mirror copy on the target system. If the production copy of the iASP fails or is shut down, you have a hot backup, in which case the mirror copy becomes the production copy.

In synchronous mode:

1. The system pages out (writes) changed information to the production copy of the iASP.
2. While that write is occurring, the synchronous send mode task transfers the information to the system that owns the mirror copy of the iASP.
3. The information is written to disk.
4. When the data is written to disk on the mirror copy of the iASP, an acknowledgement is sent to the system owning the production copy of the iASP.
5. Assuming that the write is also complete on the production copy of the iASP, control is returned to the user or application of the original write request.

In asynchronous mode:

1. The system pages out (writes) changed information to the production copy of the iASP.
2. While that write is occurring, the asynchronous send mode task transfers the information to the main memory of the system that owns the mirror copy of the iASP.
3. That system then sends an acknowledgement back to the system owning the production copy of the iASP. In asynchronous send mode, the data on the system owning the mirror copy of the iASP is not written to disk before the acknowledgement is sent back.
4. Assuming that the write is also complete on the production copy of the iASP, control is returned to the user or application of the original write request.

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.

10.2 Communication requirements for geographic mirroring

When you are implementing an IBM i high availability solution that uses geographic mirroring, you should plan for adequate communication bandwidth so that the communications bandwidth does not become a performance bottleneck in addition to system resources.

Network topology

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, and the distance between nodes can all affect the communications latency. As a result, these become additional factors that may impact geographic mirroring performance. In correctly configured customer production environments, when the distances

between the nodes is less than 50 KM (30 miles), acceptable geographic mirroring performance has been obtained.

We recommended:

- In order to provide consistent response time, geographic mirroring should have its own redundant communications lines. Without dedicated communication lines, there may be contention with other services or applications that utilize the same communication line. Geographic mirroring supports up to four communications lines and cluster heartbeat can be configured for up to two lines. However, we recommend utilizing Virtual IP addresses to provide redundancy to the cluster heartbeat.

  If configured with multiple lines, geographic mirroring distributes the load over multiple lines for optimal performance. The data is sent on each of the configured communication lines in turn, from 1 to 4, over and over again. Four communication lines allow for the highest performance assuming an optimal configuration. Relatively good performance improvements can be obtained with two lines.

  If you use more than one communication line between the nodes for geographic mirroring, it is best to separate those lines into different subnets so that the usage of those lines is balanced on both systems and further enhances redundancy of the solution.
- If your configuration is such that multiple applications or services require the use of the same communication line, some of these problems can be alleviated by implementing quality of service (QoS) through the TCP/IP functions of IBM i. The IBM i QoS solution enables the policies to request network priority and bandwidth for TCP/IP applications throughout the network.
- Ensure that throughput for each connection matches. Also, the speed and connection type should be the same for all connections between system pairs. If throughput is different, performance will be gated by the slowest connection. For example, a customer can have 1 G Ethernet speed from their servers to the attached switches. However, if the connection is using a DS-3, then from site to site, they are utilizing a 44.736 mbits/sec connection.
- Physical capacity is not throughput capacity. For older 10 M, 100 M Ethernet connections use earlier implementations of Carrier-Sense Media Access/Collision Detection (CSMA/CD). You should plan on no more than 30–35% throughput. As the network becomes more saturated, there are more collisions, causing more retransmissions. This becomes a limitation on the data throughput, as opposed to the speed of the actual line. With newer implementations of 10 M and 100 M, the data throughput can vary from 20% to 70%, and it is again dependent on network saturation.
- Ensure that your connections are taking an appropriate route. You want to understand wither it is a circuit-switching protocol (like a T-1), and whether the connection goes directly from point A to point B or is it routed through other switching offices. For example, we have seen systems 30 miles physically apart, but found that the transmission was routed through a middle office along the way, which made the transmission distance over 100 miles, and not the 30 miles that the customer had assumed.
- Size the communications bandwidth for both resynch and normal production in parallel. In a disaster situation you may end up with a scenario where you have switched over to your disaster system and you are running production. The *production* comes online and now you need to send all of the changes that have occurred since you switched over from the original production and also send normal production changes. The apply of the tracked changes is a high-priority job and may cause application performance degradation if the communications pipe is saturated.

10.3 Backup planning for geographic mirroring

Before implementing high availability based on geographic mirroring, you should understand and plan a backup strategy within this environment.

Before configuring any high availability solution, assess your current backup strategy and make appropriate changes if necessary. Geographic mirroring does not allow concurrent access to the mirror copy of the iASP, which has implications to perform remote backups. If you want to back up to tape from the geographically mirrored copy, you must quiesce mirroring on the production system and detach the mirrored copy with tracking enabled. (Tracking allows for changes on the production to be tracked so that they can be synchronized when the mirrored copy comes back online.) Then vary on the detached copy of the iASP to an available status, perform the backup procedure, vary off, and re-attach the iASP to the original production host. Following this process only requires a partial data resynchronization between the production and mirrored copies. A detach without tracking would require a full synchronization to occur.

While the iASP is detached on the backup system your changes on the production system are being tracked and are not transmitted to the backup system until it becomes available after the backup is complete. Your target system and production system will not be in synch until all tracked changes have been transmitted. To minimize synchronization, which in turn limits your exposure, we recommend always using tracking when you suspend or detach the mirror. Loss of communication between the production and the mirror copy might also require a partial synchronization. It is considered a best practice to use redundant communication paths to help eliminate some of the risks associated with a communication failure.

10.4 CPU considerations

Geographic mirroring increases the CPU load, so there must be sufficient excess CPU capacity. 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. 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.

If you are a customer, consider using IBM i Benchmark Center to help you evaluate and build sizing guides based on load testing of your solution on IBM products and technology. If you have never used the facilities of IBM i Benchmark Center before, you can start exploring center-based offerings by visiting this Web site at:

http://www-03.ibm.com/servers/eserver/iseries/benchmark/cbc/

If you are an independent software vendor (ISV), consider using an IBM Innovation Center to help you evaluate and build sizing guides based on load testing of your solution on IBM products and technology. If you have never used the facilities of IBM Innovation Centers before, you can start exploring center-based offerings by visiting this Web site at:

http://www-1.ibm.com/partnerworld/pwhome.nsf/weblook/mkt_innovation.html

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.

10.5 Machine pool size considerations

For optimal performance of geographic mirroring, particularly during synchronization, increase your machine pool size by at least the amount given by the following formula. The amount of extra machine pool storage is:

```
300 MB + .3 MB x the number of disk ARMs in the IASP
```

The following examples show the additional machine pool storage needed for IASP with 90 disk ARMs and a 180 disk ARMs, respectively:

```
300 + (.3 x 90 ARMs) = 327 MB of additional machine pool storage
300 + (.3 x 180 ARMs) = 354 MB of additional machine pool storage
```

The extra machine pool storage is required on all nodes in the cluster resource group (CRG) so that the target nodes have sufficient storage in case of switchover or failover. To prevent the performance adjuster function from reducing the machine pool size:

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 Work with Shared Storage Pools (WRKSHRPOOL) command or Change Shared Storage Pool (CHGSHRPOOL) command.

   Note: We recommend using this option with the Work with Shared Storage Pools (WRKSHRPOOL) option.

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.

10.6 Disk unit 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 a synchronous mode, all write operations on the production copy are gated by the mirrored copy writes to disk cache. Therefore, a slow target disk subsystem can affect the source-side performance. You can potentially 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. Note that this still is in essence a synchronous transaction because the production system waits for an acknowledgment that the target system has received the transaction.

10.7 Journal planning for geographic mirroring

When implementing high availability based on IBM i geographic mirroring, you should plan for journal management.

When you journal an object, the system keeps a record of the changes that you make to that object. Regardless of the high availability solution that you implement, journaling is considered a best practice to prevent data loss during abnormal system outages, as it forces pages out of memory quickly. If you are not journaling today then visit the Journaling

Performance Utilities Web page. That site has tools to help you determine the effect that journaling will have on your system. You can find more information at:

http://www-03.ibm.com/systems/i/software/db2/journalperfutilities.html

At a minimum you should be journaling minimal data for your most important data files to ensure that they are written out to disk.

You can use the Power Systems Workload Estimator to help determine the effect that the addition of journaling will have on your environment. See the Using Measured Data tab at the following Web site for more information:

http://www-912.ibm.com/wle/EstimatorServlet

10.8 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 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 shown in Table 10-1.

Table 10-1 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 SYSBASE should have at least 3. As another example, if iASP contains 50 drives, then SYSBASE should have at least 10.

Note: This is very application dependent and the results of the above calculation are only given as a starting point. You may find that fewer or more arms may be required for your application environment. Performance monitoring is critical until your application environment is known.

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.

10.9 Topology environments

There are several different topology environments on which geographic mirroring can be implemented. Each environment has its own characteristics, and the expectations for its operations are different. Table 10-2 describes the different topology environments.

Table 10-2 Topology environments

Geographic mirror environment	Characteristics of the environment
Campus environment	► Systems are very close together. ► Low cost to provide separate comm for application and geo mirror and no network infrastructure changes are required. ► Minimum 2 x 1 Gbps Ethernet cards dedicated to XSM. ► Very low latency on the remote writes. ► Very low chance of full resync due to comm failure. ► Customer has end-to-end network control.
High speed remote environment - dark fiber connected (DWDM)	► Systems geographically separated over a short distance. ► Usually bandwidth that is sufficient for high I/O, but single path only. ► Can use multiple Ethernet cards in the system effectively to maintain separation for geo mirror traffic and applications. ► Low latency. ► Low chance of full resync due to comm failure. ► Customer sometimes has end-to-end network control.
Remote environment - comm line telco connection (DS3, OC3, and so on)	► Systems geographically separated. ► High cost of communication bandwidth. ► Long lead time to upgrade bandwidth. ► Usually minimum comm required for run time purchased due to expense. ► Medium to high latency maybe added to disk writes. ► Medium chance that comm failure could cause a full resync. ► Customer not in control of network.

Expectations for operations in the various environments

In this section we discuss expectations for operations in the various environments.

Campus

A campus environment is very favorable for geographic mirroring. There are no communications performance issues since the customer is in control of the network. Simply add another 1 GB Ethernet switch if communications does not have enough capacity.

Normal performance analysis applies in this environment. Configuration changes may be needed whenever full resynch is required (memory moved to machine pool and Sysbase I/O activity increased due the quantity of data being applied to the target system). You have 1 GB Ethernet, configured so you have maximum throughput so you have the least resynch time.

High speed remote environment: Dark fiber

This environment is similar to the campus environment. Communications stability is high. Ensure that you have adequate communications throughput during a resynch. If you are sharing the dark fiber with other systems, they may need to hold their communications until the resynch is completed to not saturate the communications link.

Remote environment: Communications lines leased from local Telco

In this environment, planning is critical. Communications costs are expensive and there are usually long lead times to implement changes to bandwidth capacity. Customers tend to size for minimum requirements due to high cost of communications. You are not in control of the network and vendor changes in implementation could lead to latency changes, which can affect application response time. If communications mechanisms are unstable then a full resynch of databases could be necessary due to IBM i not being able to suspend clustering gracefully.

10.10 Network configuration considerations

In this section we discuss how to size the amount of data that will be sent from one system to the other.

When Geographic Mirror writes the changed data in an iASP to the backup systems it sends the changed data across in 4 K memory blocks. You can use performance tools to give you estimates of the amount of data that is currently being written in your current environment. You can use that information as a basis for planning the amount of communications bandwidth needed for day-to-day operations and also to determine how long the initial synchronization would take as well as any subsequent time should a full resynch be required.

The process to size the amount of data that will be sent from one system to the other is straight forward:

1. Use system performance tools in your current environment to determine how many MBps of changes would be sent across the communications mechanism during a representative normal peak period.
2. Divide that number by 2 since we recommend at least two lines for redundancy. Refer to 10.11, “Examples and scenarios” on page 310, to determine the appropriate communications line capacity needed on a day-to-day basis.
3. Divide the size of the iASP used in the implementation by the communications bandwidth to determine the approximate time that maybe e necessary for the initial synchronization and any full synchronization in a disaster scenario.

10.11 Examples and scenarios

In this section we illustrate three scenarios:

- An existing IBM i application environment
- A Windows-hosted environment
- An IBM i hosting another IBM i

10.11.1 Scenario 1: An existing IBM i application environment

To determine the Megabytes of writes per second for each interval, run IBM i performance tools during a representative 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 do the following 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 block 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 since the disk writes are doubled since this system is using mirrored disk.
5. Plot by time interval for each storage pool.

If you use DB2 Web Query to create the plot, the formula for the Define field of the graph is:

```
((QAPMDISK.QAPMDISK.DSBLKW / QAPMDISK.QAPMDISK.INTSEC)
```

This gives us the amount of traffic assuming that we were sending all write information across the communications mechanism.

Our plot graph of MB per second per time interval looks like the one shown in Figure 10-1.

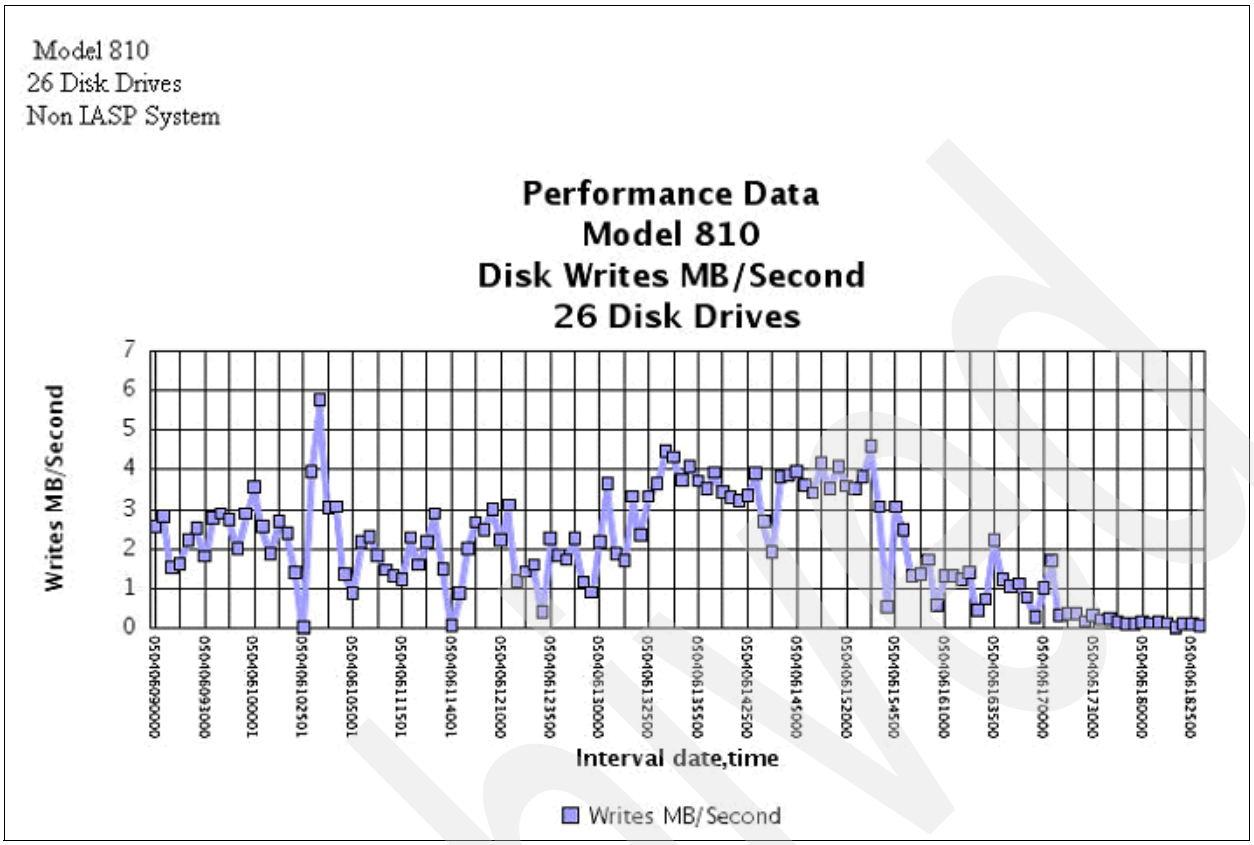


Figure 10-1 How much data is sent from one system to the other

In this environment the customer has not implemented iASP yet so the performance tools are showing all writes for the system (that is, system, database, and non-database writes). That gives us a high water mark for the current environment. However, once we implement independent auxiliary storage pools, we will just send over the changed database pages. To estimate how many of the writes are database writes then you can use the QAPMJSUM file from the performance tools. From the QAPMJSUM file calculate the percentage of database writes:

- JSPWRT number of writes: Total number of physical database and non-database write operations. PD (11,0).
- JSDBW number of synchronous database writes: Total number of synchronous physical database write operations for database functions. PD (11,0).
- JSADBW number of asynchronous database writes: Total number of asynchronous physical database write operations for database functions. PD (11,0).

The percentage of DB writes is obtained by the following formula:

```
% of DB Writes = (JSDBW + JSADBW) / JSPWRT * 100
```

The DB2 Web query formula for the Compute field is:

```
(JSDBW + JSADBW) / JSPWRT * 100
```

An example would look like Figure 10-2.

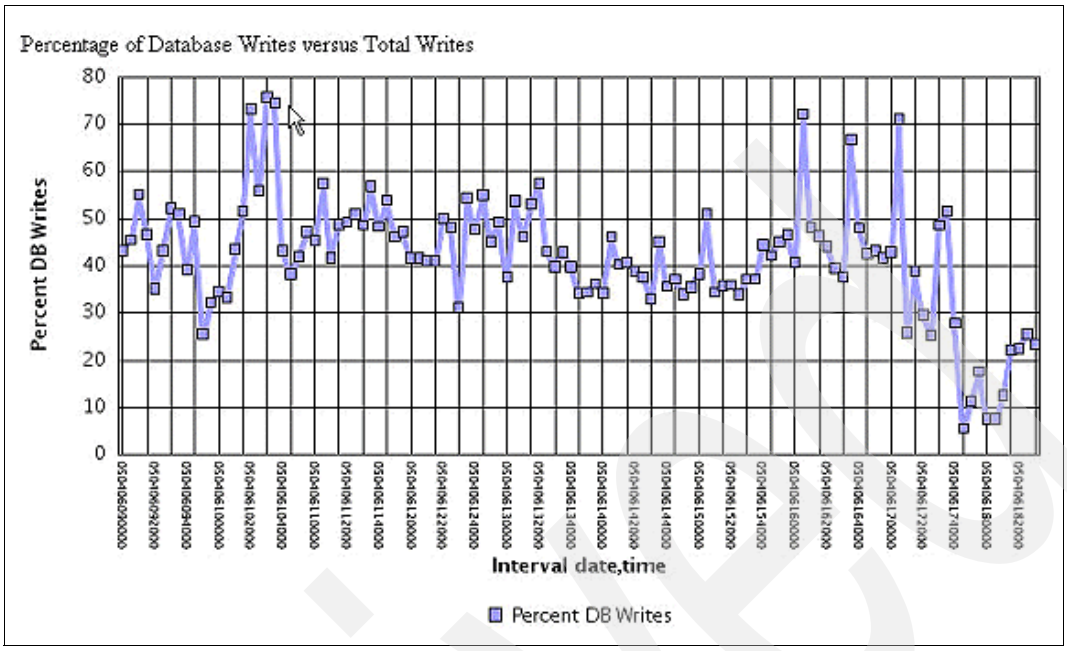


Figure 10-2 Percentage of database writes versus total writes

If we want to just transmit database writes then we take the maximum DB writes per second and multiply it by the percentage of DB writes that we want to use. If we look at the graph, above our average database writes is about 60%. So our minimum throughput needed would be 4 MBps multiplied by 60% = 2.5 MBps.

Using this customer as an example, if they want to use geographic mirroring for disaster recovery then they need a pipe that can accommodate somewhere between 2.5 MBps and 4 MBps of data being transferred.

Since we are configuring with two lines we need between 1.2 MBps and 2 MBps per line.

From 10.11.4, “Communications transports speeds” on page 316, we can see that:

- A DS3/T3 allows 5.6 MBps theoretical throughput with 2 MBps with best practices 30% utilization. Two DS3/T3 lines are about right.
- An OC-3 (optical carrier) line allows 19.44 MBps theoretical throughput with 6 MBps with best practices 30% utilization.

For this customer we recommend two DS3/T3 lines for normal day-to-day operations with the knowledge that if their transactions grow or they share the DS3/T3 with other applications it may not have enough capability.

To determine the time needed for initial synchronization divide the total space utilized in the iASP by the effective communications capability of our communications lines. Speed of the lines makes a big difference. In the above customer the iASP size was 900 GB. If we were in the same machine room using 1 Gb Ethernet switches then the initial synchronization time might be less than one hour. However, using our two T3/DS3 lines in the above example, each having an effective throughput of 7.2 GB/hour 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.

The above solution does meet the customers recovery point objective (RPO) and recovery time objective (RTO).

10.11.2 Geographic mirroring for hosted Windows servers

To determine the number of bytes written you need to use Windows performance tools. You would need to start performance tools for the Windows server that you want to put under geographic mirroring. To do this:

1. Select **Administrative Tool** → **Performance**.
2. Then with the System Monitor selected click the plus sign (+) on the toolbar.
3. On the Add Counters dialog, choose the **Physical Disk** performance object.
4. Select the **Disk Bytes/sec** counter and **Total** instance.
5. Click **Add**.

To end the counter:

1. Start **Administrative Tools** → **Performance**.
2. Expand **Performance Logs and Alerts®**.
3. Select **Counter Logs**.
4. On right pane, right-click the log file name and select **Stop**. The counter is now ended.

From the resulting reports determine the peak writes. Divide the peaks by two since you are going to implement over two communications lines. Determine what line speed is needed for day-to-day operations.

For our example customer the MBps of writes is 12 MBps. So for each line we would need a peak of 6 MB sustained performance, so we recommend 2-OC-3 lines.

To compute the initial synchronization time divide the size of the network storage space of the IBM i hosted partition by the effective speed of the communications mechanism.

Our iASP that contains the network storage space is 2.4 TB. If we divide 2.4 TB by 12 MBps our install synchronization and any subsequent full resynch is approximately 56 hours. The cost of upgrading the communications capability to achieve a shorter synchronization time was too great and the 56-hour synchronization time was too long for the client.

The client decided that high availability was more important than disaster recovery so they brought a new Power 520 Capacity Backup Server into another building so that they could get some physical separation for disaster recovery. Fiber between their own buildings was more cost effective and the synchronization time was under 2 hours. Every night they suspense geographic mirror with tracking so production stays running. They perform a save to virtual tape on the target system. Then they resume geographic mirroring, which sends the changed pages from the primary to the target. The next morning they copy the virtual tape to physical tape and send it offsite for disaster recovery.

Note: These sync times are approximations only. Workload types, network topology, server and storage design, and operations can greatly effect sync times.

10.11.3 Geographic mirroring: IBM i operating systems hosted by another IBM i System

It is fairly simple to determine the bandwidth needed for day-to-day operations and initial synchronization in this case. All write I/O will need to be sent from the primary system to the secondary system. The steps are the same as for scenario 1 except that you do not need to

determine the percentage of database writes to non-database writes because IBM i will be sending all of the writes.

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 would have to:

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 since the disk writes are doubled since this system is using mirrored disk.
5. Plot by time interval for each storage pool.

We used DB2 Web query to create the plot. The formula for the Define field of the graph is:

```
((QAPMDISK.QAPMDISK.DSBLKW / QAPMDISK.QAPMDISK.INTSEC)  * 520) / 1000000 / 2
```

This gives us the amount of traffic assuming that we were sending all write information across. The plot graph of MB per second is shown in Figure 10-3.

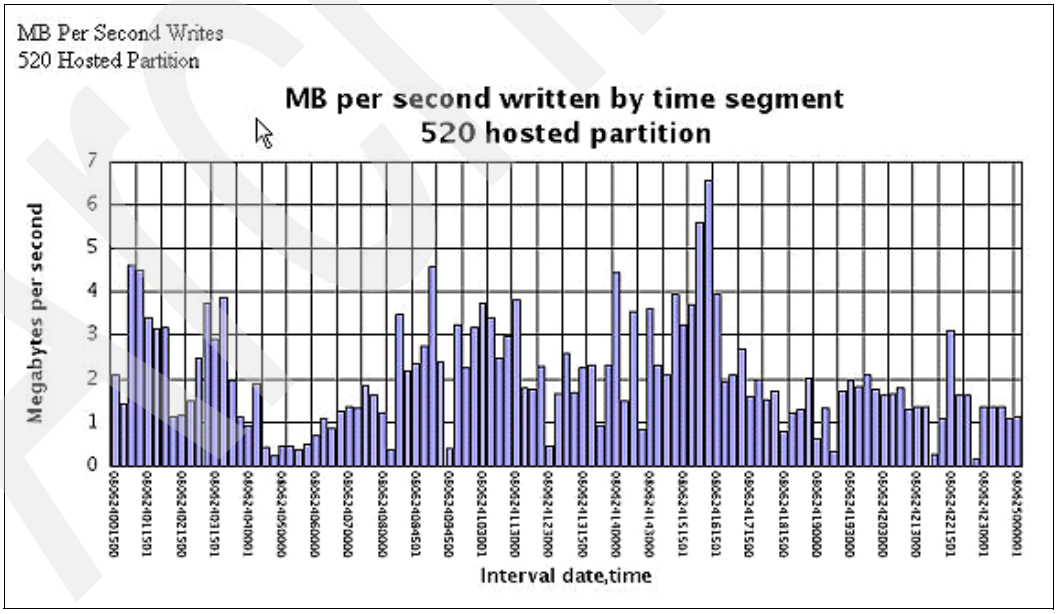


Figure 10-3 Hosted partition MBps

Using this customer as an example if they want to use Geographic Mirroring for disaster recovery they need a pipe that can accommodate 5 MBps of data being transferred. Since we are configuring with two lines we need between 2.5 MB per second per second per line.

From 10.11.4, “Communications transports speeds” on page 316, we can see that:

- 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. That is too large.

For this customer we would initially have them look at two DS3 lines, but we will let them know that with growth they might have to go to two OC-3 lines. To compute the initial synchronization time, divide the size of the network storage space of the IBM i hosted partition by the effective speed of the communications mechanism.

In our example the network storage space hosting IBM i was set up as 600 GB. In this case it would take 42 hours to do the initial synchronization for a disaster recovery scenario using two DS3 lines. This meets the customer RTO and RPO requirements.

By contrast, in a high availability campus environment this could take less than two hours with appropriately configured machines.

10.11.4 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 plan for 30%. Therefore, the best that a T1 line can transfer is .135 MBps. If I 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 will need more than a T1 line in order to achieve effective geographic mirroring transaction throughput.

T3 lines are a common aggregation of 28 T1 circuits that yields 44.736 Mbps total network bandwidth or 5.5 MBps with a best effective throughput of 70% = 3.9 MBps and planning number of 2 MBps.

The OC (optical carrier fiber optic based broadband network) speeds help you grow.

Other communications line speeds are shown in Table 10-3.

Table 10-3 Communication line speed

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

Type	Raw speed MBps	Raw speed MBps	30% planning MBps	GB/hour during synch
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

Part 3

Implementation examples using PowerHA for i

In this part we explain some implementation examples using PowerHA for i to implement a high availability solution in the following environments:

- Chapter 11, “Implementing Oracle JD Edwards EnterpriseOne high availability using PowerHA for i” on page 321
- Chapter 12, “Implementing Lawson M3 Business Engine high availability using PowerHA for i” on page 327
- Chapter 13, “Implementing SAP application high availability using PowerHA for i” on page 335

11

Implementing Oracle JD Edwards EnterpriseOne high availability using PowerHA for i

This chapter addresses the considerations and process for using PowerHA for i to migrate a JD Edwards EnterpriseOne environment into a high availability (HA) environment. It reflects an implementation of geographic mirroring using the PowerHA for i HASM GUI, but the considerations are the same for any JD Edwards EnterpriseOne environment, which makes use of independent auxiliary storage pool (iASP) support.

11.1 Background and history

IBM i and the Oracle JD Edwards products reflect a long-standing and dynamic relationship that continues to provide outstanding value to our joint customers. Both the JDE World (which runs only on IBM i) and JD Edwards EnterpriseOne products are actively being marketed. Both IBM and Oracle continue to provide enhancements that provide increased functionality and improved price/performance.

In keeping with this relationship, both product lines have supported iASP-based implementations for many years. Specifically, the JD Edwards World has been supported in an Independent ASP environment since release A 7.3. JD Edwards EnterpriseOne has been supported since the release of EnterpriseOne Tools release SP21 or 8.94. This chapter reflects an extension of the support already in place, which was described in previously published IBM Redbooks publications and whitepapers.

11.2 Application architecture

The architecture for JD Edwards EnterpriseOne on i contains three major components:

- Oracle JD Edwards EnterpriseOne Application Server
- IBM DB2 for IBM i Database Server
- Oracle JD Edwards EnterpriseOne HTML Web server

Because the IBM i operating system is characterized by its integration, IBM often recommends *All on i* configurations in which the three major components of the JD Edwards EnterpriseOne environment are run in the same IBM i partition. The Web server component is often run in a separate IBM i partition or on an operating system other than i.

Another key component of any JD Edwards EnterpriseOne environment is the deployment server, which is used to initially install the application, to apply fixes, and to deploy the application. The deployment server is Windows based and thus can be implemented using the Windows integration support in which IBM i hosts the disk and other resources for the Windows server.

This chapter reflects the migration of the application, database, and deployment servers into an iASP environment for use with geographic mirroring. The technique is the same for any other iASP-based solution, including storage-based solutions such as metro mirroring and global mirroring. It also reflects some important enhancements to the previous support:

- This project used a newer version of JD Edwards EnterpriseOne: EnterpriseOne Application 8.12 and EnterpriseOne Tools 8.96.
- It includes migrating the Windows-integration-based deployment server into the HA environment.
- Although it is not described in this chapter, follow-on work will be done to also migrate the Oracle JD Edwards EnterpriseOne HTML Web server into the HA environment.

Items to be migrated

This section describes the items that need to be migrated for the JD Edwards EnterpriseOne environment.

User profiles to be migrated

By default, JD Edwards EnterpriseOne uses two user profiles named JDE and ONEWORLD, which need to be migrated into the Independent ASP. In addition, any user profiles associated

with EnterpriseOne application users also need to be migrated into the Independent ASP. These are often called proxy users. As part of this migration, these user profiles will be associated with a job description that specifies an Initial ASP Group (INLASPGRP) consistent with the migration of the application into an Independent ASP. Note that the user profiles are not actually migrated into the Independent ASP but are instead added to the administrative domain so that they can be managed across the cluster nodes.

Libraries to be migrated

The following libraries must be migrated for the production environment:

- COPD812, OWJRNL, PRODDTA
- DD812, PD812, SVM812
- JDEOW, PD812FA, SY812
- OL812, PRODCTL

In previous whitepapers and Redbooks publications, the foundation library could not be moved because of restrictions related to the use of subsystem descriptions and other work management objects such as job queues, classes, and job descriptions. IBM i 6.1 resolved this issue with the addition of the ASP Group (ASPGRP) parameter to the Change Subsystem Description (CHGSBSD) command, which can be used to include an Independent ASP in the namespace of the subsystem monitor job in order to locate work management objects in an Independent ASP. There are still, however, naming dependencies within EnterpriseOne that preclude placing the E812SYS Foundation library into the Independent ASP.

For this project, only the production environment was migrated, but the same approach should be applicable to the other standard environments of development, pristine, and prototype. The libraries for those environments are:

- CODV812, DV812, PS812FA
- COPS812, DV812FA, PY812
- COPY812, PS812, PY812FA
- CRPCTL, PS812CTL, TESTCTL
- CRPDTA, PS812DTA, TESTDTA

This list of libraries is based on the Oracle installation documentation for JD Edwards EnterpriseOne and is subject to change. For more information about these libraries and their contents, see the appropriate installation document.

Directories to be migrated

The following directories need to be migrated for any EnterpriseOne production environment:

- E812SYS
- JDE812
- PD812
- OneWorld

The following directories are required for the standard environments of development, pristine, and prototype:

- DV812
- PS812
- PY812

If a hosted Windows server is being migrated, it is also necessary to move any of the virtualized disk drives that are implemented with NWS storage spaces. The names will look like the following:

```
QFPNWSSTG/nws-storage-space-name
```

These names can be determined using the Work with NWS Storage Spaces (WRKNWSSTG) command.

11.3 Process

The following section provides an overview of the process for migrating a JD Edwards EnterpriseOne into a Geographic Mirroring environment using the PowerHA Solution-Based GUI. It assumes that JD Edwards EnterpriseOne is already installed on a server that will become the primary node of the cluster.

This process is documented in more detail in the IBM whitepaper “Implementing a High Availability solution for Oracle JD Edwards EnterpriseOne using an Independent Auxiliary Storage Pool.” This whitepaper will also be used to document any updates to support described in this chapter. When the document is published, it will be available for download from the IBM TechDocs Web site and can be found using the search options at the following URL:

http://www.ibm.com/support/techdocs/atsmastr.nsf/Web/Techdocs

11.3.1 Actions to take before migration

In this section we provide the list of actions to take before starting the migration process. Note that these actions are specific to JD Edwards EnterpriseOne and do not include generally recommended actions such as backing up the system environment.

1. Quiesce the entire application environment and vary off the Windows-based deployment server.
2. Do any cleanup that will improve the performance of the migration. Some possibilities include:
   - Delete the older unneeded journal receivers in library OWJRNL.
   - Delete unneeded logs in the JDE812 directory.
   - Archive older application data that will not be needed after the migration. This is especially useful as the migration of the application data is one of the longer steps in the process.
3. Apply any updates recommended by IBM or Oracle for running JDE EnterpriseOne in an Independent ASP.

11.3.2 Migration process

The general process for using PowerHA for i to implement geographic mirroring has already been described previously in this document. Only three steps have specific actions with regard to JD Edwards EnterpriseOne:

1. Migrate user profiles.

   Select **JDE**, **ONEWORLD**, and any other proxy users, as described above.

2. Migrate libraries.

   Select the libraries indicated above depending on the environments that you wish to migrate. Note that there may be library dependencies indicated. For example, PRODDTA and OWJRNL must be migrated together because tables in library PRODDTA are journalled to journals in library OWJRNL.

3. Migrate directories.

   Select the directories indicated above depending on the environments that you wish to migrate. Also select the directory entries corresponding to the NWS storage spaces if a hosted Windows server is being migrated.

11.3.3 After migration

At a high level, most of the tasks to be performed after the migration are needed because the application was installed to a specific server, but several components need to be updated to reference the new name associated with the takeover IP address rather than the original server name.

The following tasks must be performed after successful completion of the migration using PowerHA for i:

1. Copy the E812SYS library.

   Use save/restore to save the contents of the E812SYS library on the primary node and to restore them on the secondary node. This can be done by using physical media or a combination of ftp and a save file. Note that this action must be repeated any time updates are made to EnterpriseOne, including the installation of fixes or upgrades.

2. Update the JD Edwards EnterpriseOne database files.

   The JD Edwards EnterpriseOne application makes use of metadata files that contain configuration information about where objects are located. These files must be updated to reflect that the name of the Enterprise server has changed.

3. Update the configuration files.

   Several parameters in the jde.ini file contain server names that must be updated to reflect that the Enterprise server has effectively been renamed. Also, a new parameter has been added to enable Independent ASP functionality in EnterpriseOne. There are also changes to the jas.ini and jdbj.ini files.

4. Update any ODBC-based clients.

   Any data source that was configured to access the old server name must be updated to reference the name associated with the takeover IP address.

5. Update the Web servers.

   The configuration of any Web servers needs to be updated to reflect the name associated with the takeover IP address rather than the name of the primary node.

6. Update the deployment server.

   The deployment server must also be updated to reference the name associated with the takeover IP address. This is again implemented usings scripts developed in collaboration with Oracle.

7. Create the Windows integrations objects on the backup node.

   The general approach for implementing HA for Windows integration is to place the network storage spaces in the Independent ASP. It is also necessary to create configuration objects on the secondary system in the cluster.

11.4 Validation

In order to validate the successful migration of EnterpriseOne into an Independent ASP, use the PORTTEST command. This process should be repeated after a switchover to ensure that it also works on both the primary and secondary nodes in the cluster. We also recommend that you verify that the integrated Windows server will vary on successfully on the backup system.

11.5 Conclusions

PowerHA for i significantly simplifies the process of implementing EnterpriseOne in an Independent ASP environment. Specifically, its easy-to-use graphical interfaces support the initial creation of the cluster and Independent ASP, the migration of the necessary objects, and the management of the clustered environment.

12

Implementing Lawson M3 Business Engine high availability using PowerHA for i

In this chapter we show how the PowerHA for i solutions based graphical user interface (GUI) can be used to migrate the Lawson M3 Business Engine (M3 BE) into the high availability environment that we are creating. Cross-site mirroring with geographic mirror was the solution that chosen for our high availability solution.

12.1 Background

In today's world having a high availability solution is becoming more and more of a requirement for customers. Even smaller to mid-sized customers need to minimize down time, while still accounting for planned outages from operations like backups or upgrades. A high availability solution that accounts for both planned and unplanned outages can help minimize down time. In this chapter we show how the core components of the Lawson M3 ERP application can be moved into a high availability environment that we created.

We used the High Availability Solutions Manager GUI or solution-based GUI to set up our high availability solution and migrate the Lawson M3 BE application into that solution. We selected cross-site-mirroring with geographic mirroring as our high availability solution of choice. Cross-site mirroring with geographic mirroring provides more than just a backup system to fail over to in the event of an unplanned outage. This also gives us a second copy of the data that for planned outages can be detached in order to do operations such as a back up, apply fixes, or upgrade. While this is offline, the primary system continues to run without interruption. Thus, you can minimize downtimes using this solution. When the operation is complete the backup system can then be reattached and mirror is resumed after resynching target and source copies. Thus, our high availability environment is once again active.

Note: While the mirroring is detached there is no recovery in case the primary fails.

12.2 Lawson M3 architecture

Lawson M3 is a supplier of collaboration software that focuses on the manufacturing, maintenance, and distribution industries and serves many customers around the world. The Lawson M3 7.1 ERP solution used here is a Java-based application that runs on IBM i servers.

The Lawson M3 7.1 introduces several new key features:

- A new foundation that now is decoupled from the business logic
- Support for server-side pooling of application programming interface (API) connections
- Support for data compression in API connections
- Support for the new IBM Technology for Java Virtual Machine
- Support JDK™ 1.5
- Support for WebSphere 6.1 with M3 WorkPlace

12.3 Entities to be migrated

The following Lawson M3 components were moved into the independent auxiliary storage pool:

- M3 Business Engine database
- M3 Business Engine IFS directory
- Life Cycle Manager IFS directory

12.4 Entities to be configured

Configuration needs to be done prior to implementing a high availability solution. A few key items that must be configured are:

1. For geographic mirroring there are several recommendations for an optimal configuration. See 4.3.4, “Recommendations when using geographic mirroring” on page 49, for details.
2. IP address available to be the server takeover IP address.
3. Set SST to work on IBM System Director Navigator. See “Setting up the SST user ID connection” on page 75.

12.5 Process

This section provides the steps to set up the high availability environment and move the necessary directories, databases, and user profiles to this environment.

12.5.1 How IBM high availability solutions work

The first step ensures that you understand how each of the potential high availability solutions works. The solution-based GUI provides a Flash demo with details on each of the potential solutions. We highly recommended that you watch this demo. See 6.2, “HASM GUI” on page 90, for details on how to view this.

12.5.2 Select your high availability solution

The next step is to choose the high availability solution that you want. We chose cross-site-mirroring with geographic mirroring. See 6.2, “HASM GUI” on page 90, for details on how to chose the high availability solution that you want.

12.5.3 Verify requirements before setting up your high availability solution

The next step is to verify requirements that need to be met in order to set up the high availability solution that you selected. During this step you will also need to provide information about the backup system, data ports to use, and takeover IP address to use. To do this:

1. See 6.5, “Verifying requirements for your high availability solution” on page 101, for steps to gather a list of requirements that need to be met. Once you have this list continue on with the steps below.

2. Ensure that the following are entered on the Verify Requirements panel (Figure 12-1):
   - Backup node name
   - Enter any additional data port name.
   - Enter server takeover IP address.

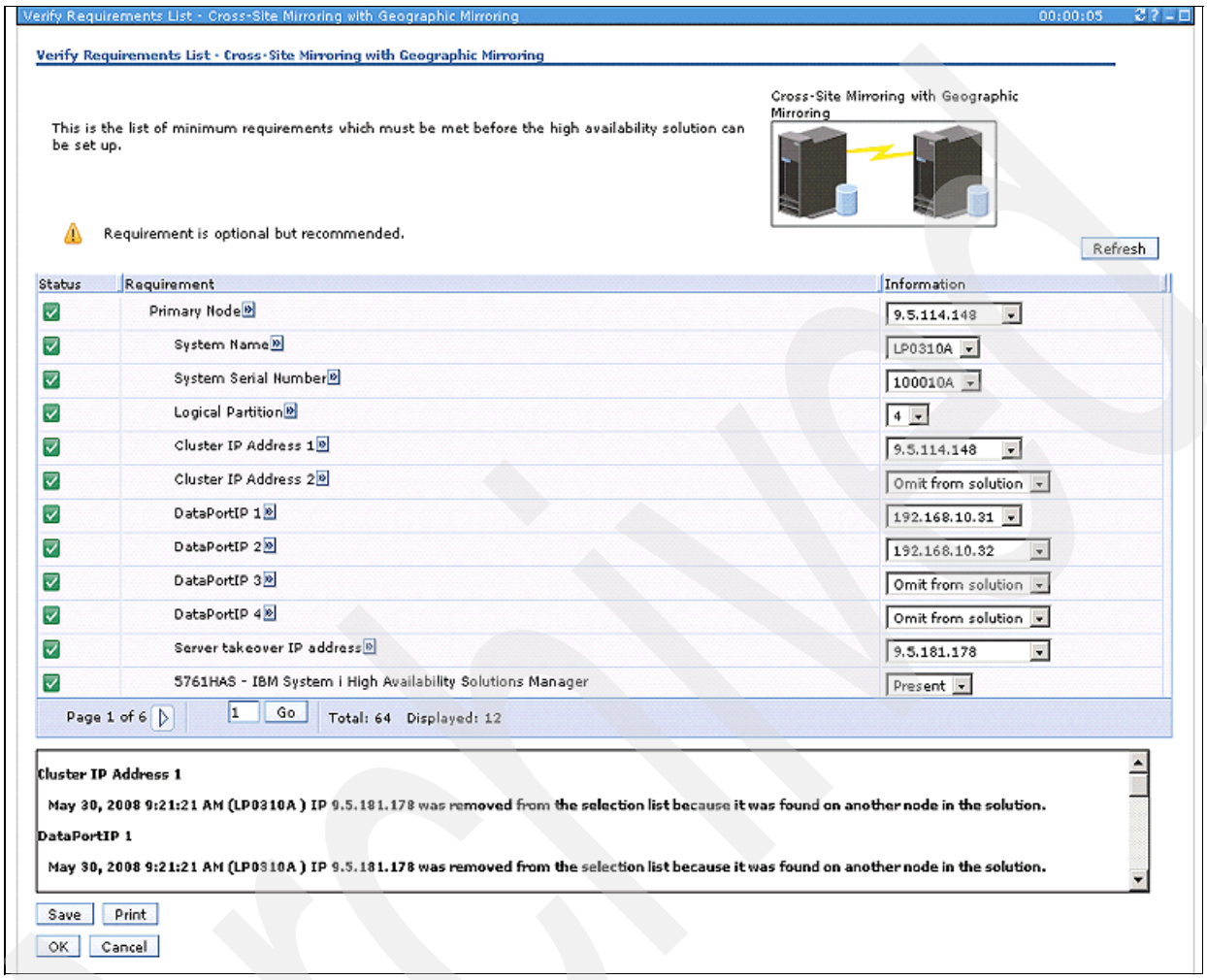


Figure 12-1 Add IP resources to the shopping list

3. Perform any other actions needed to address requirements that still need to be met.

12.5.4 Set up your high availability solution

Now that all the requirements have been met, the next step is to set up the high availability solution. You will also migrate the Lawson M3 BE databases, directories, and user profiles during this process. The steps below guide you through this.

Note: This step will be the longest and will likely take significant time to complete. Up to this point no changes to the system have occurred. However, starting with this step changes to the system will be done.

1. Follow the steps in 6.6.2, “Setting up your high availability solution” on page 108. Stop when you get to 6.6.3, “Migrating user profile” on page 122.

2. Follow the steps in 6.6.3, “Migrating user profile” on page 122, selecting all of the Lawson M3 BE user profiles. The profiles we migrated were:
   – M3DBREADS
   – M3DBUSR
   – M3SRVADM
   – M3SRVADMS
3. Follow the steps in 6.6.4, “Migrating libraries” on page 125, selecting all of the Lawson M3 BE databases that you require. The database that we migrated was MVXJDTA.
4. Follow the steps in 6.6.5, “Migrating directories” on page 129, selecting all of the Lawson M3 BE directories that you require. The directories that we migrated were:
   – /M3B3: Default M3 Business Engine root dir
   – /LifeCycle: Default Life Cycle Manager root dir
5. Follow the steps in 6.6.6, “Switching” on page 132.

Once you have completed all the steps above and have verified that you can do a switchover, the setup of the high availability solution is now complete. However, there are additional steps required for the Lawson M3 BE application to run in this environment.

12.5.5 Manage your high availability solution

The solution-based GUI provides a management function to monitor your high availability solution and perform tasks such as switchover. See 6.7, “Managing your high availability solution” on page 141, for more details on the manage function.

12.6 Post tasks

Below are the steps required to configure the Lawson M3 Business Engine to work in the high availability environment that we created in the previous section.

Ensure M3DJVA library is installed on backup system

The M3DJVA library contains job descriptions and subsystem descriptions that are objects that cannot reside on an independent auxiliary storage pool. It also contains service programs that were not put in a separate library for this exercise and hence the library needs to be duplicated on the backup system. To do this:

1. Save M3DJVA on the primary system:

```
SAVOBJ OBJ(*ALL) LIB(M3DJVA) DEV(*SAVF) SAVF(DLTME/M3DJVA)
```

2. FTP and restore M3DJVA on the backup:

```
RSTOBJ OBJ(*ALL) SAVLIB(M3DJVA) DEV(*SAVF) SAVF(DLTME/M3DJVA) OPTION(*NEW)
```

Modify MOVEX.properties

The following lines in MOVEX.properties must be changed:

- license.servers.addr=10.10.5.118

  Change to take over IP address.

- db.con.source=My_System_Name

  Change to db.con.source=IASP_NAME.

- db.con.url=jdbc:as400://10.10.5.118;errors=full

  Change to jdbc:as400://take.over.ip;errors=full;database name=IASP_Name.
- auth.dname=cn=MySystemName.my.comany.com_BM,ou=BM,o=lawson,c=US

  Change to takeover IP.

Modify Set Database Authority Program

The Set Database Authority Program needs to be modified in order to work with the independent auxiliary storage pool (iASP) using the steps below:

1. Open M3CJVA/JUSRC, Member: JSETAUTDCL.
2. Change the following line from:

   ```
   CHGVAR     &UpdPath      VALUE('/QSYS.LIB/' *TCAT &DataLib    *TCAT '.LIB')
   ```

   To:

   ```
   CHGVAR    &UpdPath      VALUE('/IASP/QSYS.LIB/' *TCAT &DataLib   *TCAT '.LIB')
   ```
3. Change the following line from:

   ```
   SBMJOB   CMD(CALL PGM(JSETAUTDCL) +  PARM(&File &DataLib &Owner &PrimGroup
   &DBUser &DBRead                  &Batch))   JOB(SETAUTD)
   ```

   To:

   ```
   SBMJOB   CMD(CALL PGM(JSETAUTDCL) PARM(&FILE &DATALIB &OWNER &PRIMGROUP &DBUSER
   &DBREAD + &BATCH)) JOB(SETAUTD) CPYENVVAR(*YES)
   ```
4. Save the changes then compile and overwrite existing member.
5. Run the Set Database Authority Program:
   a. SETASPGRP ASPGRP(IRD_IASP)
   b. ADDLIBLE M3DJVA
   c. GO MVXSTART -> Option 82 -> Option 40

12.7 Validation

When properly set up and configured the Lawson M3 BE application will be able to switch back and forth between production and backup hosts. After switching to the backup host and restarting Lawson M3 BE, it should be transparent to the user which host the Law-son M3 BE is running on. In addition, to further validate the functionality of Lawson M3 BE in our high availability environment we used the Lawson M3 Order Entry Benchmark kit to stress the system. The result of that test showed no errors, and behavior was identical to a standalone system.

12.8 Conclusions

As this chapter has shown, the solution-based GUI greatly simplifies the process of setting up a high availability environment. In addition, it provides a management function where you can quickly see the state of your high availability environment and easily perform tasks such as switchovers. In the past setting up a high availability environment, such as geographic mirroring, as we used here, and migrating an application into it, would have been a complex and lengthy task with numerous steps. Now the solution-based GUI has provided a simple and straightforward way to implement this and reduce the complexity of managing this environment.

Running your Lawson M3 BE application in a geographic mirror environment provides more than just having a backup system to fail over to in the event of an unplanned outage. This also gives us a second copy of the data that for planned outages can be detached in order to do operations such as a backup, apply fixes, or an upgrade. Thus, this solution gives you both the data resiliency and the flexibility to quickly recover from both expected and unexpected outages.

12.9 References

For more information about Lawson M3 on PowerHA see the Lawson M3 on PowerHA whitepaper on IBM tech docs:

http://www.ibm.com/support/techdocs

13

Implementing SAP application high availability using PowerHA for i

This chapter discusses the use of the PowerHA for i High Availability Solution Manager to configure a high availability environment for SAP NetWeaver® running on IBM i V6.1. The high availability environment will be a geographic mirror configuration utilizing internal disk for storage. The SAP application is a NetWeaver 7.0 double-stack system. The SAP system is installed in a two-tier configuration.

13.1 Background

SAP is one of the world's leading providers of business software. SAP products provide a wide range of solutions to empower nearly every aspect of business operations. SAP provides collaborative business solutions for many industries with customers throughout the world.

SAP has supported running its applications in an independent auxiliary storage pool (iASP)since 2002. This has allowed customers to implement various data resiliency solutions to better meet their availability needs. Since the initial support of iASPs, SAP has continued to enhance support of iASPs by making architecture changes, code enhancements, and providing tooling to assist in configuring applications to run in iASPs. Many SAP on IBM i customers are now successfully running production environments in high availability solutions based on iASPs.

13.2 High-level application architecture

The core SAP applications is based on one of two different runtime technology layers:

- Advanced Business Application Programming (ABAP™)
- Java technology

Though the underlying implementations of SAP applications differ significantly between ABAP and Java, the high-level architecture is very similar. For purposes of running SAP applications in an iASP, knowledge of the high-level architecture is sufficient.

SAP applications, regardless of whether they are ABAP or Java, consist of three main components:

- Kernel
- Database
- Integrated file system directories
- Files

13.2.1 Kernel

The kernel is a collection of executable files that provide the infrastructure for an SAP system. The kernel contains executables that are responsible for running SAP applications, managing the work flow, maintaining locks, and maintaining the SAP system to name a few. The kernel provides the interface between the SAP system and the platform-specific operating system and the database. The kernel contains both Integrated Language Environment® (ILE) objects and Integrated File System (IFS) objects. Even though the kernel consists of both ILE and IFS objects, all objects are stored in a library, commonly referred to as a kernel library. In order to use a specific kernel with an SAP system the kernel must be applied to that system. This process extracts the IFS objects to the proper IFS directory, as well as creates necessary symbolic links to enable the kernel for use.

For a complete description of the SAP kernel on IBM i refer to Chapter 12, "The SAP kernel on the System i server," in *Implementing SAP Applications on the IBM System i Platform with IBM i5/OS*, SG24-7166.

13.2.2 Database

SAP applications utilize a relational database to store customer data and application data. On the IBM i, the database is often referred to as the database library. Each SAP system will have its own database and in the case of a double-stack SAP system, where an ABAP and Java system are combined, the SAP system will have two databases, one for ABAP and one for Java.

In addition to the database, SAP systems are configured to use database journaling. With database journaling each transaction is recorded by an object called a journal and written to a repository called a journal receiver. In case of data loss the database can be rebuilt using journal entries. The journal receivers are located in a library separate from the database library.

SAP data libraries have the following naming conventions, where <sid> is the SAP system ID:

- ABAP data library: R3<sid>DATA
- ABAP journal library: R3<sid>JRN
- Java database: SAP<sid>DB
- Java journal library: SAP<sid>JRN

For a complete description of IBM i integrated database as it relates to SAP applications refer to section 9.5, "Basic principles of SAP database and SAP systems," in *Implementing SAP Applications on the IBM System i Platform with IBM i5/OS*, SG24-7166.

13.2.3 IFS directories and stream files

The file system structure for SAP applications consists of two main directory trees:

- /sapmnt
- /usr/sap

SAP applications utilize directories and stream files for storing many different types of objects including:

- Log files
- Configuration data
- Executables
- Standalone tools

Each individual SAP system maintains its own branch in the directory trees with the following naming convention, where <sid> is the SAP system ID:

- /usr/sap/<sid>
- /sapmnt/<sid>

In addition to the individual SAP system directories a directory tree called the transport directory exists. The transport directory is /sapmnt/trans and is global for all SAP systems in an SAP landscape.

For a complete description of the IBM i integrated file system and the SAP directory structure on IBM i refer to section 8.2.4, "The Integrated File System," in*Implementing SAP Applications on the IBM System i Platform with IBM i5/OS*, SG24-7166.

13.3 Application objects in the high availability solution

The key to a high availability solution is determining how to deal with the different objects within the application with respect to the high availability solution. The objects can be grouped into three categories:

- Objects that need to be migrated to the iASP
- Objects that are not migrated to the iASP, but will be managed by an admin domain
- Objects that are neither migrated nor part of an admin domain

13.3.1 Objects migrated to the iASP

SAP application objects that need to be migrated to the iASP include:

- Database library and all of the contents
- Journal library and all of the contents
- Kernel library and all of the contents
- Directories and files under /sapmnt/<sid>
- Directories and files under /usr/sap/<sid>
- Home directory for each SAP user profile

Currently, SAP recommends that all directories under /usr/sap/<sid> and /sapmnt/<sid> be moved to the iASP for any iASP-based high availability solution, including geographic mirror. The transport directory is moved to the iASP only if the SAP system designated as the domain controller is also located in the iASP. For more information about migrating the transport directory to an iASP see SAP note 568820 "iSeries: Implementing an Independent ASP (iASP) System."

To eliminate dual maintenance of the different kernel components in a high availability environment, the kernel library should be located on the iASP for any iASP-based high availability solution, including geographic mirror.

13.3.2 Objects managed by the admin domain

Certain objects utilized by SAP applications cannot be migrated to the iASP. These objects will exist on both the production and the backup system. In order to simplify maintenance and management, some of these objects will be added to an admin domain. Objects added to an admin domain are monitored and then kept synchronized between the production host and the backup host. Some of the objects that can be monitored and managed by an admin domain include system values, system environment variables, subsystem descriptions, TCP/IP attributes, and user profiles. Except for user profiles, the High Availability Solution Manager will automatically add objects to the admin domain to be monitored. The user is responsible for selecting which profiles to add to the admin domain. The following SAP user profiles should be selected and added to the admin domain (where <sid> represents the SAP system ID and <nn> is the SAP instance number):

- <sid>ADM
- <sid>GROUP
- <sid>OFR
- <sid>OPR
- <sid>OPRGRP
- <sid>OWNER
- <sid><nn> (An SAP system may have multiple <sid><nn> user profiles.)
- <sid> R3OWNER
- <sid> R3GROUP

13.3.3 Objects remaining in SYSBAS

Any objects that will be neither migrated to the iASP nor added to an admin domain remain in SYSBAS. These objects mostly consist of temporary objects, static objects, and objects unsupported in an iASP. All SAP application objects that will not be migrated to an iASP or added to the admin domain are contained in the following libraries and will remain in SYSBAS (<sid> is the SAP system ID and <nn> is the SAP system instance number):

- R3<sid>400
- R3<sid><generated_name> (Libraries containing SQL packages have this form.)
- R3SYS
- R3WRK<nn> (An SAP system may have multiple R3WRK<nn> libraries.)
- R3400

These libraries do not contain any objects that need to be migrated.

13.4 Application configuration

Paramount to the ability to recover data from a failure is the ability to resume business operations. It is imperative that the SAP application retains the ability to start and run with minimal interruption if a switchover or failover occurs. Because the host name of the backup system will differ from the production system it is necessary to utilize a virtual host name. A virtual host name is a host name that can switch between two hosts and still retain the same TCP/IP address. The TCP/IP address used for the virtual host name is the server takeover IP address specified as part of the geographic mirror configuration. In addition to the server takeover IP address, a new host name will need to be assigned to this TCP/IP address. This is the virtual host name. To utilize virtual host names, modifications to the SAP application configuration will need to be made in the following locations:

- Default profile
- Instance profile
- Secure store (Java only)
- Instance properties (Java only)

13.5 Implementing the solution

The PowerHA for i High Availability Solution Manager can be used to simplify the implementation of geographic mirror in a simple SAP landscape. Use of the High Availability Solution Manager automates many of the steps required to implement a geographic mirror high availability solution and to migrate SAP applications to an iASP. Consult your business partner to determine whether a geographic mirror solution is the best fit for your situation and availability requirements.

Instructions for manually migrating SAP applications to an iASP can be found in the document *Configuring SAP for use with an Independent ASP*. This document is attached to SAP note 568820 "iSeries: Implementing an Independent ASP (iASP) System." This document is the definitive source for implementing SAP applications in an iASP and was used as the main reference for the sections discussing pre-processing and post-processing tasks. If necessary, refer to this document for further details on implementing SAP in an iASP.

13.5.1 Pre-processing tasks

Before starting the actual geographic mirror configuration there are some tasks that must be completed. It is assumed, at this point, that the SAP system is installed and running on the production host and that all hardware requirements have been met including the configuration of the necessary TCP/IP interfaces necessary for clustering. At a minimum, two additional TCP/IP interfaces are required, one for the server takeover IP and one to four interfaces for the dataport.

Verify health of the SAP system

We highly recommend that you verify the health of the SAP system. It is important to identify any existing problems in order to validate that problems do not arise due to the migration of the SAP system to the iASP and configuration changes that will be made. A list of tests that can be run to verify your SAP system can be found in section 7 of the document *Configuring SAP for Use with an Independent ASP*.

Back up SAP system

Once the health of the system is verified and any existing problems are fixed, it is good practice to completely back up the SAP system to tape.

Prepare backup host

While the production host is being backed up, the backup host can be prepared. To prepare the backup host it is necessary to perform a partial install of an SAP system. The partial install will create the necessary infrastructure to enable switching your SAP system from the production host to the backup host. To prepare the backup host perform the following steps (all commands are run from the IBM i command line unless otherwise noted):

1. Save the kernel library from the production host to a save file using the SAVLIB command.
2. Transfer the save file to the backup host using FTP.
3. Restore the kernel library using the RSTLIB command. The restored kernel library should have a different library name than the original kernel library on the production host. This can be done by using the RSTLIB parameter on the RSTLIB command or by using the RNMOBJ command after the library has been restored. Changing the kernel library name on the backup host will avoid a potential conflict when switching the SAP system to the backup host.
4. Add the restored kernel library to the library list by calling ADDLIBLE.
5. Fix the owners and authorities of the objects in the kernel library by calling FIXR3OWNS.
6. Set the system value QRETVRSEC to 1 to allow user profiles to be migrated to the admin domain. This can be done by running the command:

   ```
   CHGSYSVAL SYSVAL(QRETSVRSEC) VALUE('1')
   ```

7. Create a new SAP system using the CRTR3SYS command. The SAP system ID (SID) should be the same as the SID of the SAP system on the production host.
8. Create an SAP instance using the CRTR3INST command. The instance numbers and roles should be the same instance numbers and roles of the SAP system on the production host.
9. Create an SAP Java user using the CRTSAPUSR command for standalone Java and double-stack systems.

10. Delete the following directories and their contents from IFS using the RMVDIR command:
   - /usr/sap/<sid>
   - /sapmnt/<sid>
   - /sapmnt/trans

The backup host is now prepared and ready for implementing the high availability solution. Further details of the steps outlined above can be found in section 5.2 of the document *Configuring SAP for Use with an Independent ASP*.

13.5.2 Implementation process

The PowerHA for i High Availability Solution Manager is used to easily configure an iASP and a high availability cluster environment. It can also be used to migrate and enable applications to operate in a high availability environment. This section discusses the steps that are specific to implementing a geographic mirror for an SAP system. For a general and more complete reference for using High Availability Solution Manager refer to Chapter 6, "High Availability Solutions Manager GUI" on page 89.

The High Availability Solution Manager is integrated with IBM Systems Director Navigator for i5/OS and can be found by expanding i5/OS Management and clicking the High Availability Solutions Manager menu option. You are now in the High Availability Solution Manager and will see the five action items necessary for setting up your high availability solution (Figure 13-1).

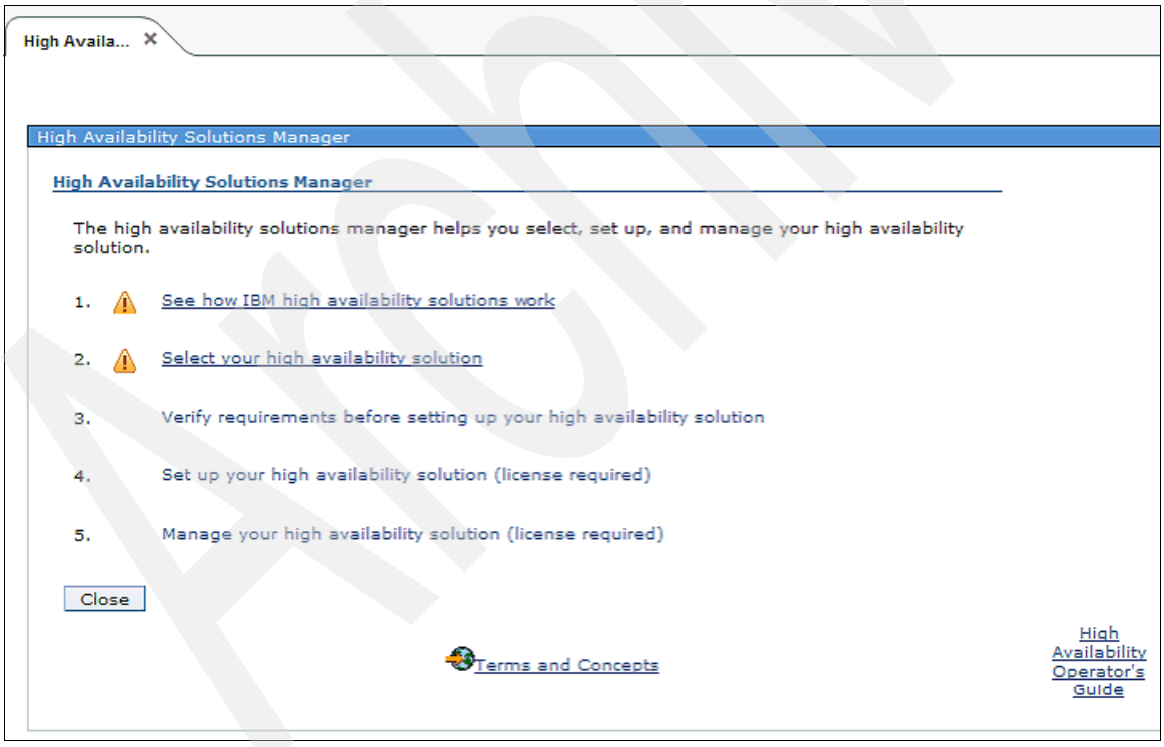


Figure 13-1 High Availability Solution Manager tasks

The action items are meant to be executed serially and will take the user through all of the necessary steps for setting up and managing a high availability environment. Since many of the steps are generic, I will only focus on the steps that are specific to implementing SAP applications in a high availability environment.

Select your high availability solution

After reviewing the high availability solutions and how they work you are now ready to begin the process of implementing your solution.

1. When asked to select your high availability solution, select the option **Cross-Site Mirroring with Geographic Mirroring** (Figure 13-2).

Figure 13-2 Select Cross-Site Mirroring with Geographic Mirroring

2. After making the selection, select **View customized shopping list** and you will be presented with a list of all of the components required for that solution. The High Availability Solution Manager will collect an inventory of your production system and if a TCP/IP host table entry for the backup host exists on the production system it will collect inventory on the backup host as well. The inventory will be merged with the shopping list to verify that the correct and necessary resources are included in the solution. Make sure that a server takeover IP address (this IP address should have already been configured) and an adequate number of dataport IP addresses are included in your solution (Figure 13-3).

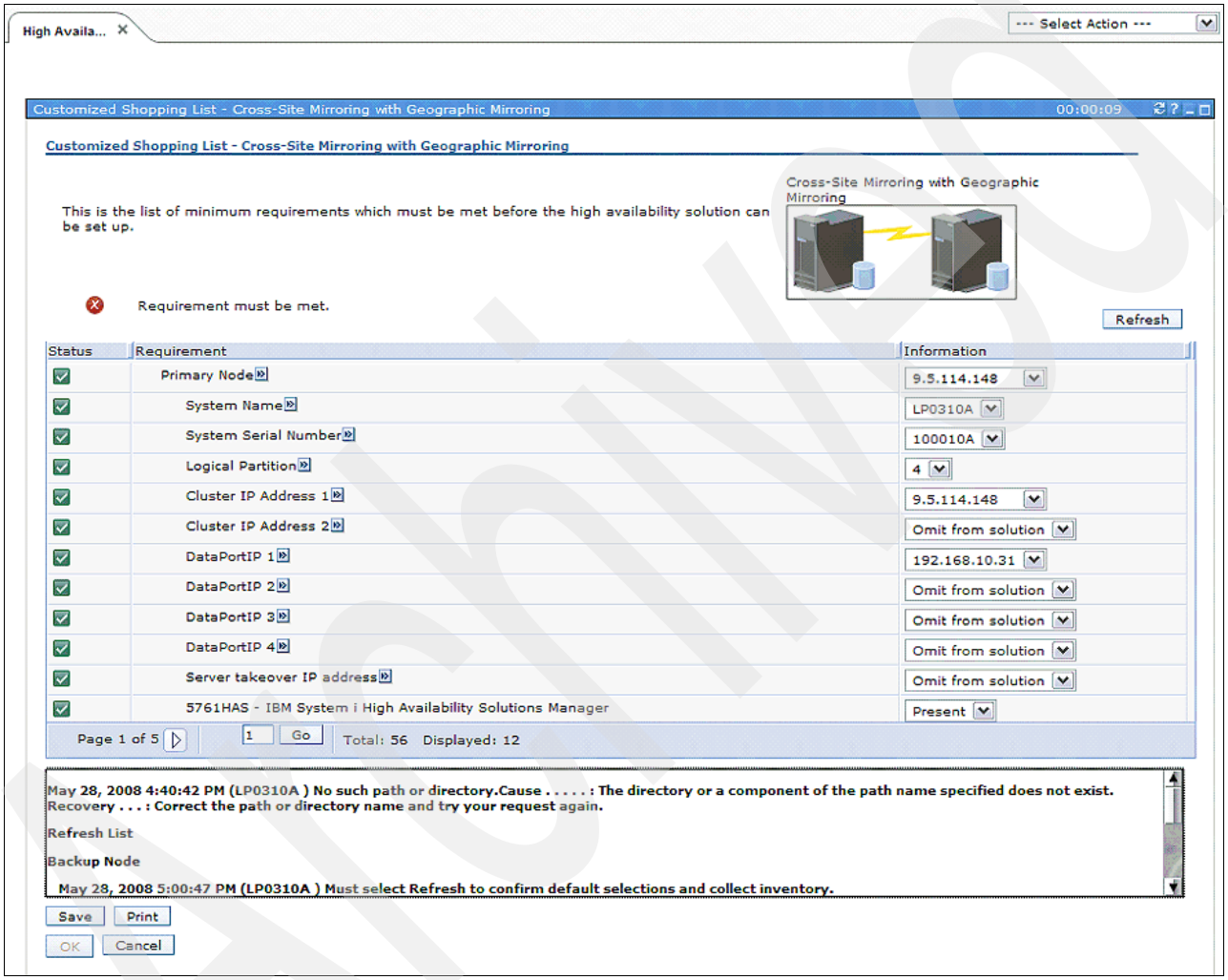


Figure 13-3 Add IP resources to the shopping list

After selecting your high availability solution and ensuring that your shopping list is complete, you can now continue with the next action item.

Set up your high availability solution

After verifying that all requirements have been met, it is now time to start the actual implementation and set up your high availability solution. This action item consists of multiple steps (Figure 13-4). Of these steps only migrate user profiles, migrate libraries, and migrate directories are specific to SAP. All other steps are generic. Refer to Chapter 6, “High Availability Solutions Manager GUI” on page 89, for more details on these steps.

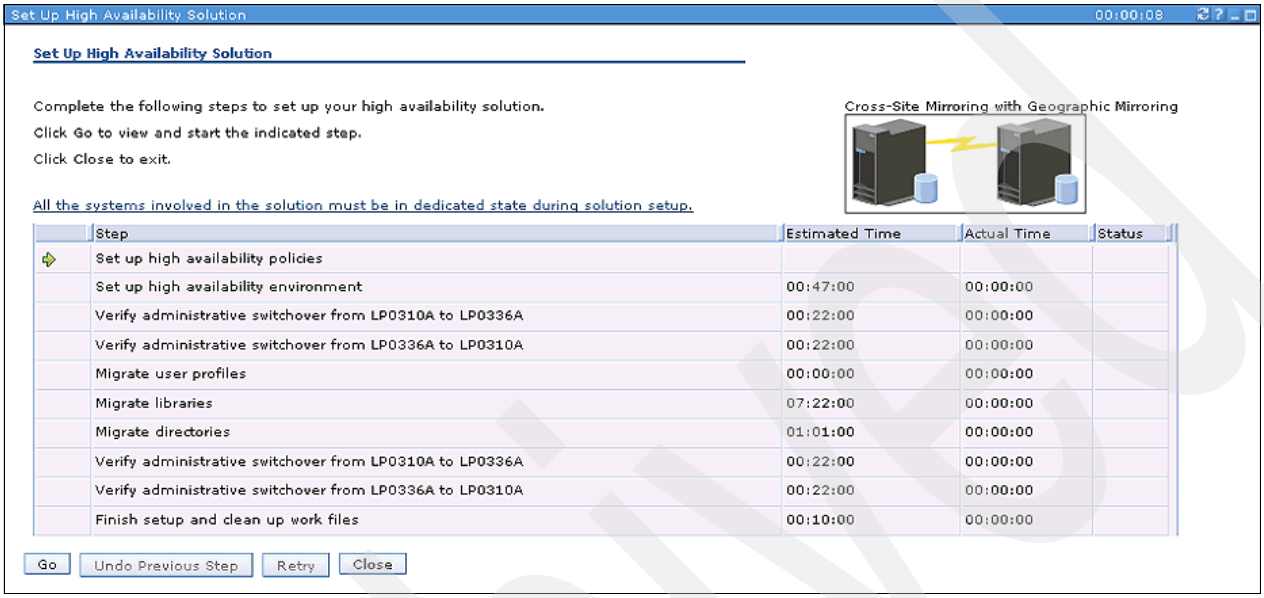


Figure 13-4 Steps to setup your high availability solution

When asked to migrate user profiles, you will need to select the SAP user profiles. Migrating user profiles is part of configuring the admin domain. Due to dependencies the group profiles <sid>GROUP, <sid>OPRGRP, and R3GROUP must be migrated first. After migrating the group profiles the remaining SAP profiles can now be selected and migrated (Figure 13-5).

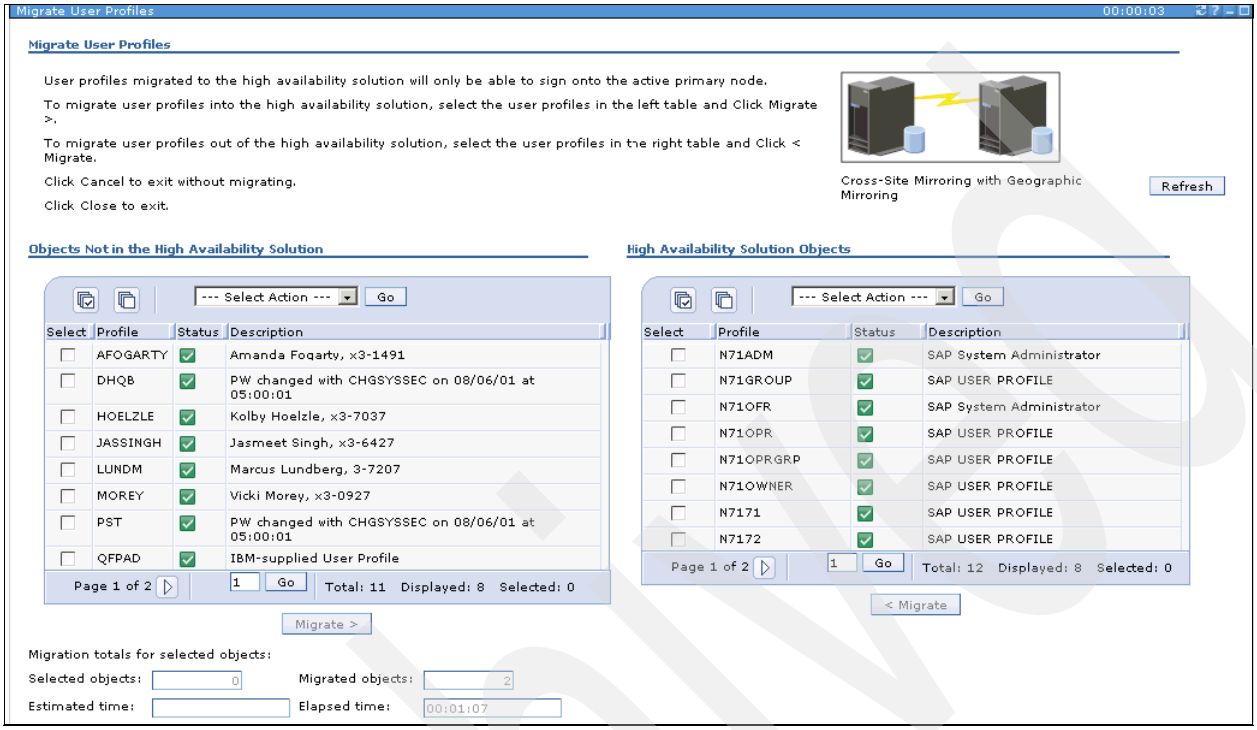


Figure 13-5 Migrate SAP user profiles

If the system value QRETSVRSEC was not set to 1 during the initial SAP application install it will be necessary to sign- on and sign off each user before migrating. For user profiles where the initial menu attribute is set to *SIGNOFF, the sign-on attempt is sufficient (as long as the password is correct) and it is not necessary to change the user profile attributes.

When asked to select libraries to migrate, you will select the kernel library, database library, and journal library (Figure 13-6). Due to dependencies, both the database library and the journal library must be migrated at the same time. For double-stack systems there will be two database libraries and two journal libraries.

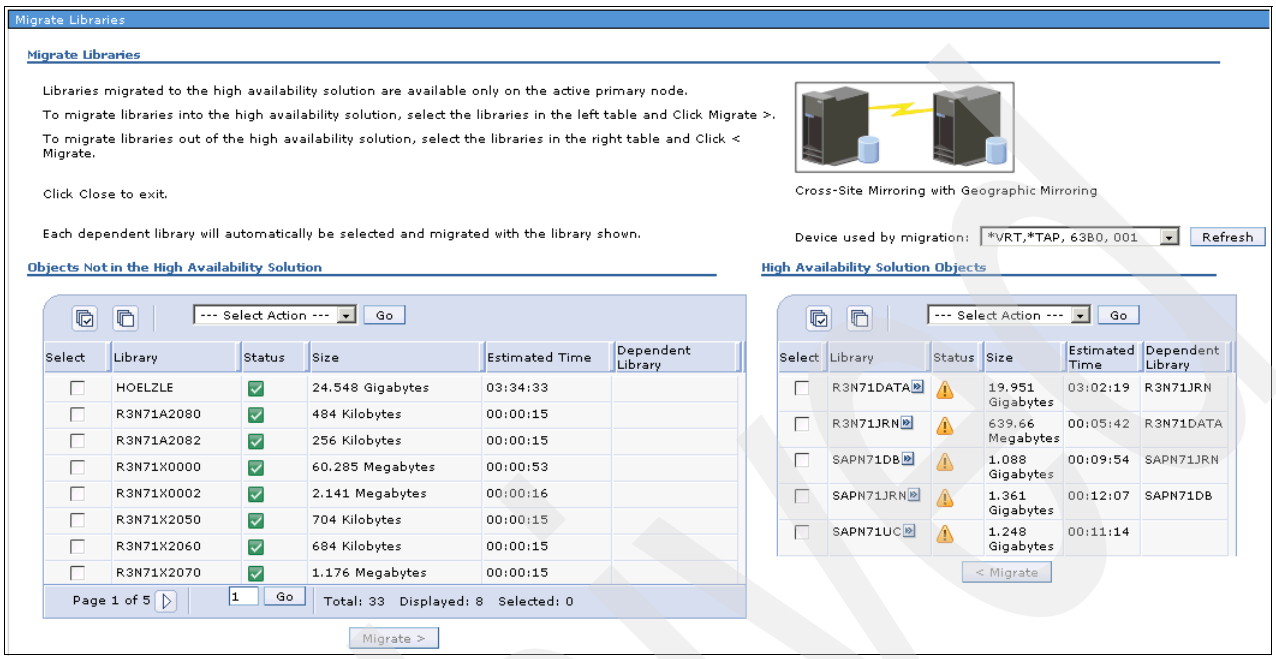


Figure 13-6 Migrate SAP libraries

When asked to select IFS directories to migrate you will need to select the directory trees /sapmnt/<sid> and /usr/sap/<sid>, where <sid> is the name of the SAP system to be migrated.

It is also necessary to migrate the home directories of each SAP user profile. The home directories are found in /home and will have the same name as the user profile. For example, the home directory for <sid>ADM will be /home/<sid>ADM (Figure 13-7).

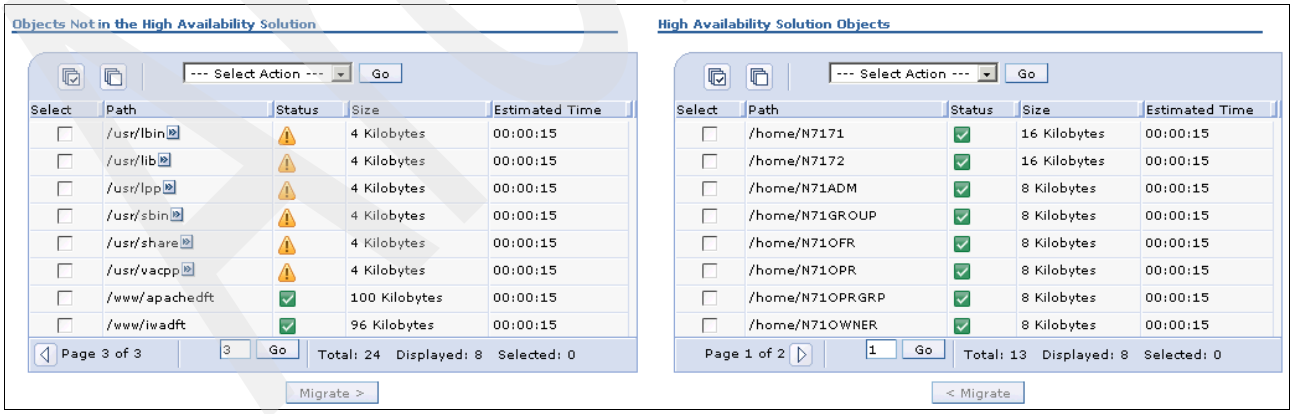


Figure 13-7 Migrate SAP IPS directories

After migrating the IFS directories the remaining steps are generic. Once you have completed the remaining steps your SAP system should now be fully migrated to the iASP. However, to complete the process, you will need to perform some final post-processing tasks.

13.5.3 Post-processing tasks

To enable SAP to run in the iASP and switch over to the backup host it is necessary to complete a few final application configuration tasks.

Reapply the SAP kernel

Now that the kernel library is located in the iASP, it is necessary to remove and reapply the kernel library. Add the SAP kernel to the library list using the ADDLIBLE command. Remember that since the kernel is now in the iASP it is necessary to make the iASP visible by running the command SETASPGRP. Run the command RMVSAP to remove the kernel. Once the remove is complete the kernel can be reapplied by running APYSAP. Finally, run FIXR3OWNS to fix the owners and authorities of the objects in the kernel library.

The SAP system should now start and run on the production system. However, because it has not yet been configured to use virtual host names, it will not start on the backup system. In order to complete the implementation, the SAP system must be configured to use virtual host names.

Configure the SAP system to use a virtual host name

Configuring SAP to use virtual host names is done by modifying the default profile and instance profile for both ABAP and Java systems. For Java standalone or double-stack systems the secure store and the instance properties must also be modified. Detailed instructions for configuring the SAP system to use virtual host names can be found in section 6 of the document *Configuring SAP for Use with an Independent ASP*.

For seamless access to the SAP system regardless of whether it is running on the production host or backup host, SAP logon for each user should also be configured to use the virtual host name.

Verify the SAP system on the production host and the backup host

After configuring the SAP system to use virtual host names start the SAP system on the production host and verify the health of the system. After verifying that the SAP system still works on the production host, switch over to the backup host. After the switchover is complete, start the SAP system as you normally would and verify the health of the system on the backup host.

The SAP system is now completely migrated and configured for a geographic mirrored environment. The SAP system can easily be switched from the production to the backup host with minimal effort.

13.6 Validation and results

When properly migrated and configured the SAP system will be able to switch back and forth between production and backup hosts. After switching to the backup host and restarting SAP, the host that the SAP system is running on should be transparent to the user. Running your SAP application in a geographic mirror environment provides the data resiliency and the flexibility to quickly recover from both expected and unexpected downtime. Using the PowerHA for i High Availability Solution Manager to implement and manage your SAP geographic mirror environment provides a simple solution to achieving your availability goals.

For more information about using PowerHA for i in an SAP environment download the Whitepaper "Leveraging PowerHA for i in an SAP Environment" from:

http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101325

Part 4

Other IBM i 6.1 high availability enhancements

In this part we introduce other important IBM i 6.1 high availability enhancements.

This part includes the following chapters:

- Chapter 14, “Environment resilience” on page 351
- Chapter 15, “Journal-driven data resilience: What is new” on page 357

14

Environment resilience

The bulk of your critical data may well lie within the kinds of objects that can reside within independent auxiliary storage pools. Relying upon a high availability solution that offers replication protection only for independent auxiliary storage pool (iASP) contents may well be shortsighted. It is for this reason that assuring replication of surrounding environmental settings is equally important. This includes objects and values that generally reside within the system ASP including system values, network configuration attributes, work management configuration values, and user profiles. Each of these must be duplicated in order for the application environment to operate the same way on the backup system as it does on the primary production system.

One popular way to achieve this objective is through use of the feature known as clustering administrative domain support. By identifying the critical environmental settings that you want replicated, you can assure that the operating system will execute a matching remote command on the target system each time that you modify a designated object on the production system.

Although the current release still does not support replication of all system auxiliary storage pool (ASP) resident environmental data, and probably never will, the universe of supported values has been substantially broadened for 6.1 to include those identified as the most popular and most critical.

It should be noted that the remote replay is performed as an asynchronous operation. Hence, there is bound to be some latency between the time when the operation is performed on the production machine and the time when the matching action reaches and is replayed on the target machine. As a consequence, fast communication lines with plenty of excess bandwidth help assure that the quantity of environmental changes in-flight and hence subject to loss in the event of a disaster is minimized. Like any asynchronous approach, some loss of recent changes has to be anticipated and thus some manual re-entry of recent commands may be required on the target side if a role swap ensues.

The cluster administrative domain synchronizes information regarding switchable devices. A cluster administrator creates a cluster administrative domain and generates a list of resources to be synchronized by adding them as monitored resource entries (MREs). i5/OS cluster management then maintains the consistency of these resources across the cluster nodes defined by the cluster administrative domain.

14.1 Cluster administrative domain support

The cluster administrative domain now supports monitoring of additional resource types, and it now enables single and compound attributes as monitored resources types. In addition, the cluster administrative domain provides more detailed status messages for both monitored resources and the cluster administrative domain, and it synchronizes information regarding switchable devices.

A cluster administrator creates a cluster administrative domain and generates a list of resources to be synchronized by adding them as monitored resource entries. IBM i cluster management then maintains the consistency of these resources across the cluster nodes defined by the cluster administrative domain.

14.1.1 Administrative domain overview

In order to ensure that an application will run consistently on any node in a high availability cluster, all resources that affect the behavior of the application need to be identified, as well as the cluster nodes where the application may run or where application data might reside.

The primary means of ensuring that system objects and attributes remain synchronized across the nodes in a cluster is the cluster administrative domain. The configuration parameters and information associated with applications and application data are known collectively as the operational environment for the application. Examples of this type of data include user profiles that control access to the application or its data or system environment variables that control the behavior of the application. In 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 the 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 nodes in the active domain.

The set up and daily management of Administrative Domains is not complicated and can be accomplished in one of several ways:

- Using the Cluster Resource Services GUI (See 7.6.4, “Administrative domains” on page 210.)
- Implicitly via the HASM GUI interface
- Via the WRKCLU command
- With the admin domain commands

Before creating an administrative domain you must:

- Identify the nodes to be included in the domain.
- Identify resource elements to be synchronized (monitored resource entries).

During normal cluster operations:

- Changes to resources are monitored on all admin domain nodes.
- Changes on one node are propagated to the others.
- Changes are preserved across node level events (for example, node ending and restarting).

14.1.2 New IBM i 6.1 cluster administrative domain commands and interfaces

In this section we review the new cluster administrative domain commands and interfaces.

- Adding and removing nodes to the cluster admin domain

  The cluster resource services (CRS) GUI provides options to add and remove nodes. Or the following commands can be used:

  - Add Cluster Admin Domain Node (ADDCADNODE)
  - Remove Cluster Admin Domain Node Entry (RMVCADNODE)

- Starting and ending the cluster admin domain

  Both the HASM GUI and the CRS GUI provide interfaces to start and end admin domains. Or the following commands can be used to manage admin domain operations:

  - Start Cluster Admin Domain command (STRCAD)
  - End Cluster Admin Domain command (ENDCAD)
  - Change Cluster Admin Domain (CHGCAD)

- Adding and removing administrative domain monitored resource entries

  Both the CRS GUI and the HASM GUI provide functions to add and remove MREs. Or the following commands can be used:

  - Add Cluster Admin Domain Monitored Resource Entry (ADDCADMRE)
  - Remove Cluster Admin Domain Monitored Resource Entry (RMVCADMRE)

- Displaying and working with administrative domain monitored resource entries can also be performed with the WRKCLU command

14.1.3 Monitored resources

In this section we list the enhancements in IBM i 5.4 and IBM i 6.1 regarding monitored resources.

- IBM i 5.4 monitored resources (with i 6.1 enhancements)
  - User profiles (*USRPRF)

    Added home directory and locale
  - System Values (*SYSVAL)

    Added approximately 65 new system values
  - Text attribute was added to the following monitored resources
    - Class (*CLS)
    - Job description (*JOBD)
    - ASP device description (*ASPDEV)
    - Network attributes (*NETA)
    - Environment variables (*ENVVAR)
    - TCP/IP attributes (*TCPA)
- New IBM i 6.1 Monitored Resources
  - Subsystem Descriptions (*SBSD)
  - Network Server Descriptions (*NWSD) of types *WINDOWSNT, *IXSVR, and *ISCI
  - NWS configurations (*NWSCFG)
  - NWSH Device Descriptions (*NWSHDEV)
  - NWS Storage Spaces (*NWSSTG)
  - Tape Device Descriptions (*TAPDEV)
  - Optical Device Description (*OPTDEV)

- Ethernet Line Descriptions (*ETHLIN
- Token-ring Line Descriptions (*TRNLIN)

14.1.4 Resource synchronization

Synchronization of resources within an administrative domain can occur at different times under varying circumstances, such as:

- When a change is made to an object and there are active nodes in the administrative domain
- When a node rejoins the administrative domain
- When a partitioned environment is repaired or corrected

With IBM i 6.1 you now have more control over how you want synchronization to occur when a new node enters an already existing administrative domain or after a partitioned cluster situation was corrected. The different synchronization options are part of the properties of the cluster administrative domain.

Synchronization options

The synchronization options are:

- Last change (default)
  - The last change (looking at timestamps) that was made before the node joined the cluster administrative domain is processed by all nodes in the active domain.
  - The last change could have been made in the active domain or on the joining node while it was inactive.
- Active domain
  - Only changes made on active nodes in an active cluster administrative domain are processed.
  - Changes made on a node while it was inactive are not passed to the active domain.
  - When a node joins the cluster administrative domain, it will be synchronized with the values from the active domain.

14.2 Failover control

IBM i 6.1 has raised the level at which the option to failover is controlled. It can now be managed at the cluster level rather than at the CRG level. This may reduce the amount of messages that need to be handled during failover in an environment where multiple CRGs are used within one cluster.

- A cluster-wide message queue (CLUMSGQ), the failover wait time (FLVWAITTIM), and default action (FLVDFTACN) can now be specified on the CRTCLU and CHGCLU commands.
- Only one message is necessary now for all CRGs switching to the new primary node.
- If both cluster message and CRG-level failover message queues are defined then only the cluster-wide message queue and its wait time and default action are taken into consideration.

14.3 Device switching

In this section we describe the new switchable devices that were made available on IBM i 6.1.

14.3.1 Device cluster resource group switchover changes

Prior to IBM i 6.1, during a switchover if a vary-on operation on the new primary node fails, then a switchback to the original primary node occurs.

In IBM i 6.1, a minor change to the device cluster resource group (CRG) switchover behavior simplifies user actions if a failure occurs during a vary-on operation for a configuration object. If all the vary-on operations are successful, the switchover behavior is still the same. Most users will benefit from the change and require no additional action. You can still obtain the old behavior with a programming change.

With the new behavior, a switchback to the original primary does not occur. Instead, a new exit program action code dependent the data value of VaryFailed is passed into the exit program, indicating that any vary-on operation failed. Additionally, the device CRG is ended.

To preserve the old behavior, the exit program should return `Failure` if the exit program action code dependent data is VaryFailed. This causes a switchback to the old primary node.

14.3.2 New switchable devices for IBM i 6.1

Prior to IBM i 6.1, i5/OS supported switching-only independent disk pool devices. When a switchover or failover occurs, the device cluster resource group switches the independent disk pools devices from the primary node to the backup node. Starting in IBM i 6.1, other hardware devices can also be marked as switchable. On the backup node, clustering ensures that the independent disk pools report in with the same resource names. Other non-independent disk pool devices, however, may report in with different resource names. Clustering now ensures that the resource names and underlying physical devices for non-independent disk pool devices controlled by a device CRG are the same on all nodes in the device domain.

Information regarding the physical device, such as resource name and type, is saved from the node that owns the hardware and restored to the other nodes in the recovery domain. This is done when the configuration object for the device is included in a device CRG or when a node is added to the recovery domain. When you add a device entry or a node to the recovery domain, the resource name for the physical device must match on every node in the recovery domain or these operations will fail. You can either do this manually or automatically by using cluster administrative domain to keep resource names of these devices consistent across nodes in the recovery domain.

Making devices switchable can be done via the CFGOBJ parameter of the following four commands:

- CRTCRG
- CHGCRG
- ADDCRGDEVE
- CHGCRGDEVE

The new types of switchable devices are categorized as follows:

- *DEVD: Device Description
  - Independent auxiliary storage pool (iASP)
  - Cryptographic device (*CRP)

  - Tape device (*TAP)
  - Optical device (OPT)
  - Network server host adapter (*NWSH)
- *CTLD: Controller Descriptions
  - Local work station controller (*LWS)
  - Tape controller (*TAP)
- *NWSD: Network Server Description
- *LIND: Line Description
  - Asynchronous line (*ASC)
  - BSC line (*BSC)
  - DDI line (*DDI)
  - Ethernet line (*ETH)
  - Fax line (*FAX)
  - PPP line (*PPP)
  - SDLC line (*SDLC)
  - Token ring line (*TRN)
  - Wireless line (*WLS)
  - X.25 line (X25)

14.4 Job queue creation allowed in iASPs

In IBM i 6.1, job queue objects can be created in independent ASPs. This allows applications to run in iASP with fewer changes. iASP job queues behave like job queues in SYSBAS. However, iASP job queues are not completely in the iASP. This causes some important differences.

Similarities of iASP job queues versus job queues in SYSBAS

Normal operations on the iASP job queues behave like they do in SYSBAS. The user is allowed to manipulate jobs on a job queue (submit, hold, release, and so on). The user is also allowed to manipulate the job queues themselves (clear, hold release, and so on). Functionally, they are the same.

Differences

There are some important differences. Since the iASP job queues are not completely in the iASP, they cannot persist across an IPL. Therefore, job entries in the job queues will disappear following an IPL. Also, they will not persist across a vary-off/vary-on cycle of the iASP. Consequently, jobs on a iASP job queue will not be available on the backup system after a switchover or failover.

15

Journal-driven data resilience: What is new

Release i 6 provides several new features that are expected to help enhance software-based, high availability solutions that employ *journal-driven* logical replication. Many of these new features are explained in this chapter.

15.1 Library journaling

One of the challenges faced by logical replication approaches has been *reaction time* when a *new* object such as a database file gets created. Say, for example, that your application creates new files on the fly. You start up the application on the production system, it creates file88, and then promptly opens that new file and begins to add records. What do you suppose needs to happen next so as to assure that this new file and its contents get properly replicated to the target machine?

Traditionally, it has been the practice of most *logical replication* approaches to monitor the *audit* journal. Notice that the new file has been created and attempt to create a matching file on the *target* machine. In addition, they journaling protection must rapidly be enabled for this new file back on the *production* machine. Worse yet, by the time the need is sensed for such actions, your application might very well have the new file locked. Hence, there could be additional delays.

Unfortunately, by the time these steps have been accomplished it is likely that the version of the new file residing on the production side has long since begun to fill up. That is, the *reaction time* of the logical replication products was rarely fast enough and could be thwarted by factors (such as locks held) that the high availability (HA) solution could not control.

What was needed was a technique to ensure that the creation of the new instance of file88 on the target side was performed in lock-step with the matching actions on the production side and that journaling protection was enabled on the source side, all *before* your application made another step. Why? So that no records entered file88 *before* journaling protection had been enabled.

Journaling at birth

IBM i release 6.1 makes both of these critical operations possible and quite easy. It provides a new feature (library journaling) by which you can designate, on the production side, that all newly created journal-eligible objects (database files, data areas, data queues) *inherit* journal protection as soon as they are created. You specify this new behavior as a library attribute. Thereafter, all journal-eligible objects added to this library are given consideration for journaling protection, starting from the instant they enter the library.

In the case of our file88 this would mean that the newly created file would be journaled at birth. Better yet, the birthing process itself would even be journaled. As a consequence, all the information (file name, file description, file properties) needed in order to create a duplicate file on the target side is embedded in the resulting journal entry.

This same *birthing* journal entry is sent via the *remote journal transport mechanism*, and hence can be replayed on the target side. And what happens when this *birthing* journal entry is replayed? A replica file (a twin) is created on the target machine. The presence of this new entry coupled with the fact that the instance of the file on the source side begins life in a journal-protected state represents a substantial improvement over the techniques provided prior to 6.1 and helps ensure that reaction time is no longer an issue.

The fact that our new file, on the production side, had journal protection enabled at birth also ensures that none of the new records added to file88 are missed. They too all show up in the journal receiver on the source side and hence are all replayed on the target side. The result is that the target instance of the new file is an exact replica of the one produced on the production machine. It has the same name, the same attributes, and the same contents.

The trigger mechanism

What triggers this new behavior? Starting in release 6.1, you now have the capability of associating a journal with an entire library. In order to start journal protection for library (*LIB) objects you would use the Start Journal Library (STRJRNLIB) command (Figure 15-1).

By using this command, all subsequently created journal-eligible objects (like file88 above) that are added to this library will have journal protection enabled and will begin to route their journal entries to the same journal that is associated with the surrounding library in which they reside.

```
                  Start Journal Library (STRJRNLIB)

Type choices, press Enter.

Library  . . . . . .   LIB1          Name, generic*
             + for more values
Journal  . . . . . .   LIB1JRN       Name
  Library  . . . . . .   LIB1        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
             + 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 Start Journal Library command example

You will probably recall that SQL schemas (also sometimes collected collections) have had this kind of behavior in the past. That is, the surrounding schema generally had a default journal (QSQJRN) and all tables created within the schema were automatically journaled at birth to the same journal (QSQJRN) associated with the schema. Now even native files (PFs) can enjoy the same behavior.

You can read more about this new feature in the technote *Journaling at object creation* found at:

http://www.redbooks.ibm.com/abstracts/tips0662.html

Note: What is called journaling on IBM i is often called logging on other platforms.

Priming the pump

Note, however, that the STRJRNLIB operation is *not* a priming step. That is, granting a library this status affects objects moved into, restored into, or created within the library *hereafter*. It does not grant journal protection to objects (such as files) *already* residing within the library. Such objects get no free ride.

In order to prime-the-pump, by granting journal protection to a set of files *already* residing within a library, you would need to execute a Start Journal Physical File (STRJRNPF) command. That command has also been enhanced for release 6.1. In particular, the priming step got a lot easier because STRJRNPF now supports an *ALL option so that all physical files or tables residing within a production library can have journal protection enabled by executing *one* command rather than hundreds.

15.2 Logical file journaling

Newly created *physical* files (PF) are not the only variety of file that caused logical replication providers difficulty the past. Many of the same reaction-time frustrations that accompanied physical files were magnified for *logical* files (LFs).

Note: On IBM i what the native file system often calls keyed logical files (LFs) SQL would call indexes. The native system also provides non-keyed LFs. These might be called views by SQL. In order for your applications to work properly, all of the varieties, regardless of whether they are keyed, need to be replicated by the HA software.

Say that an application created a new logical file (CoolView) over an existing physical file. For a number of releases, the logical replication provider had to monitor the *audit journal* to detect this occurrence and once again was faced with the challenge of determining the logical file attributes, determining which physical file it was built over, determining the logical file's properties, and then had to scurry to create a matching view on the target side.

Worse yet, by the time the logical replication third-party software reacted, the file might well be locked on the production side. As you can imagine, orchestrating all of the correct steps was challenging, error-prone, and usually not a very good performer. In order to assist them, beginning in earlier releases and continuing in release 6.1, there are now enhanced journal entries produced when a LF is created as well as when it experiences attribute changes (new name, revised authorization, and so on).

Background

A keyed logical file can be thought of as consisting of two parts:

- The surrounding object that houses the attributes (the LF)
- The sub-object, an access path (AP), which in turn houses an index, It is this index within the AP that houses the keys, as shown in Figure 15-2.

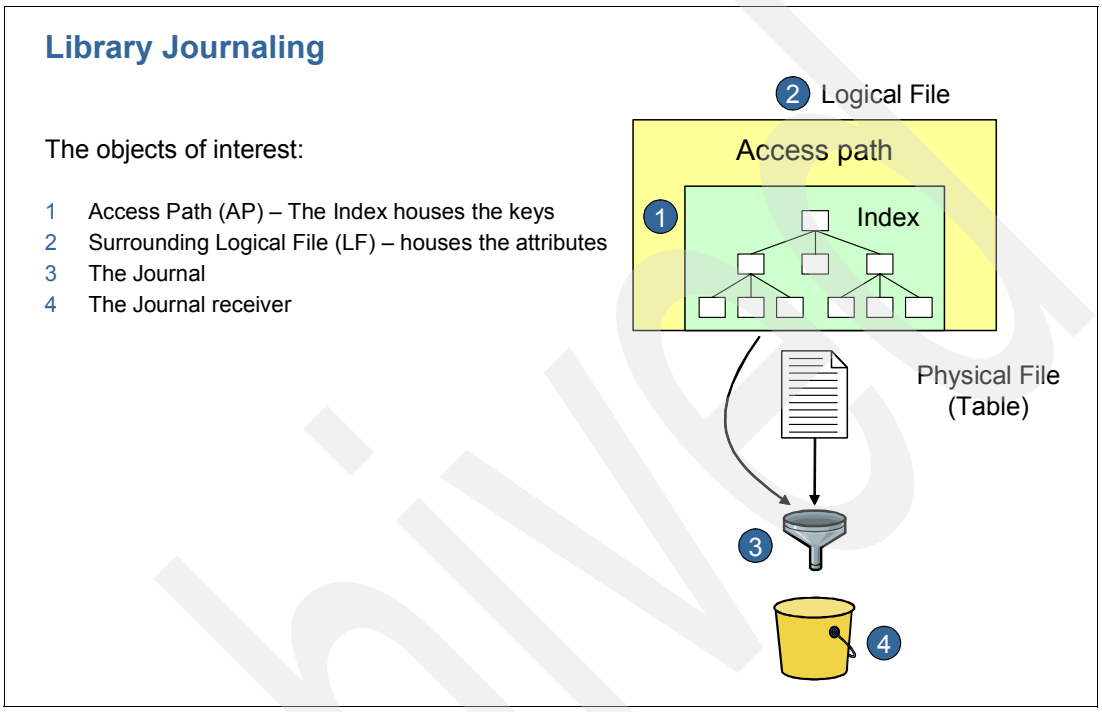


Figure 15-2 Journaling protection for both the LF and AP

Although the internal object (the AP) housing the key values has been eligible for explicit journaling for many releases, users were never able to overtly request corresponding journal protection for the surrounding logical file itself. Hence, there has long been a STRJRNAP command to protect the keys but not a STRJRNLF command. This meant that keys were protected but your LF's object-wide attributes were not—at least not until some recent releases began to step up to implicitly (some call it *covertly*) move in this new direction. Lack of assured attribute logging in earlier releases was troubling for logical replication HA providers.

Journal protection for the surrounding logical file is now consistently provided implicitly/covertly by the operating system. Hence, all you need to do in order to ensure that journaling protection is in place for the LF is to initiate journal protection for the underlying physical file.

Logical files will be journaled to the same journal as the associated physical file. The journaling is started when the operating system detects a need for it, such as logical file creation or attribute/authority modifications experienced by the logical file itself (since these are the operations that need to be replayed on the target side by the third-party logical replication software).

What is different for 6.1 is that these logical files now show up as listed/protected objects associated with the journal (they are not quite as covert as they used to be) and the corresponding journal entries now house the *name* of the LF—another step in the direction of making them feel a little less covert.

It is easy to overlook the need to journal your access paths

It should be noted that journal protection for the *logical file* itself (the shell that surrounds the access path housing the keys) is not the same as initiating journaling for the *access path*. Having the system start journal protection covertly for the LF does not guarantee that it will elect to do the same for the AP. Hence, if you wanted journal protection for the access path (index) itself (and that is a wise choice if you want to limit recovery time following a crash) you would be well advised to separately issue the STRJRNAP command so as to achieve your recovery time objective (RTO).

When journaling is implicitly started on a logical file, only the file (the LF) is journaled. The members and access paths are not journaled. As Figure 15-3 shows, on release 6.1, when you display the *journaled file* information for a journal you will see evidence of this behavior. The logical files are listed as being journaled but the members are not counted.

```
                    Display Journaled Files

Journal  . . . . . :   TESTJRN        Library  . . . . . :   JRNLIB

Number of journaled files  . . . . . . . . . . . . . . . . :   6
Number of journaled members  . . . . . . . . . . . . . . . :   4

                        Type of                                 Type of
File        Library      file          File        Library      file
FILE1       LIB1          PF
FILE2       LIB1          PF
FILE3       LIB1          PF
FILE4       LIB1          PF
VIEW1       LIB1          LF
VIEW2       LIB1          LF




                                                                  Bottom
Press Enter to continue.

F3=Exit   F5=Refresh   F12=Cancel   F16=Repeat position to   F17=Position to
```

Figure 15-3 Example of logical files being journaled

Note: Notice the presence of logical files, which show up as views in Figure 15-3.

Why is access path journalng wise in a high availability environment? Because it helps you achieve your RTO (especially in an independent auxiliary storage pool (iASP) environment) by reducing the risk that a large access path will need to be rebuilt when your iASP is varied on. Of course, AP journaling is wise in other scenarios as well. It helps speed up abnormal IPL duration if the library housing your AP resides in the system ASP (*SYSBAS).

15.3 Remote journal enhancements

The *remote* journal technology, as illustrated in Figure 15-4, is a basic building block used by nearly all third-party logical replication high availability solutions. There have been a number of attractive enhancements added for remote journaling in 6.1.

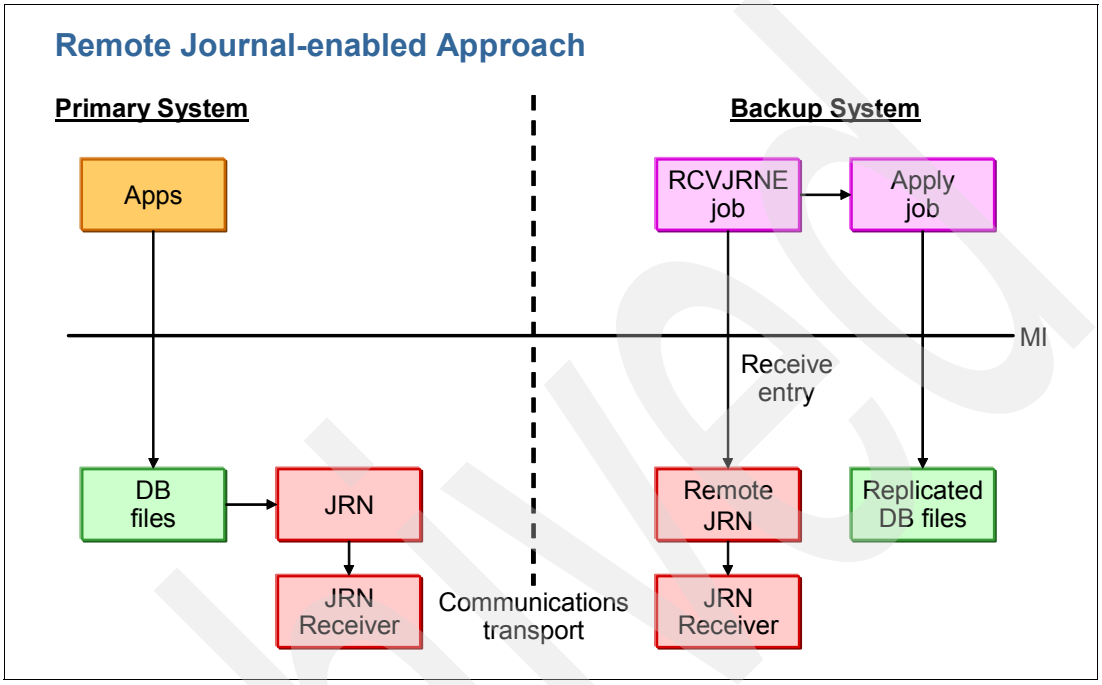


Figure 15-4 Basic remote journaling

15.3.1 Change Remote Journal (CHGRMTJRN) command enhancements

The Change Remote Journal (CHGRMTJRN) command is indispensable to remote journal configuration. Many of the new features that arrive with i 6 can be controlled with this command. Figure 15-5 shows an example of the new command structure.

```
                  Change Remote Journal (CHGRMTJRN)

Type choices, press Enter.

Relational database  . . . . . > NODE4         Name
Source journal . . . . . . . . . > LCLJRN       Name
  Library  . . . . . . . . . . .   LCLJRNLIB    Name, *LIBL, *CURLIB
Target journal . . . . . . . . . > TGTJRN       Name, *SRCJRN
  Library  . . . . . . . . . . .   TGTJRNLIB    Name
Journal state  . . . . . . . . . > *ACTIVE      *SAME, *ACTIVE, *INACTIVE
Delivery . . . . . . . . . . . .   *ASYNC       *SAME, *ASYNC, *SYNC
Starting journal receiver  . . .   *ATTACHED    Name, *ATTACHED, *SRCSYS
  Library  . . . . . . . . . . .                Name, *LIBL, *CURLIB
Data port services:
  Node identifier  . . . . . .   NODE4         Name, *SAME, *NONE
  Data port IP address . . . . .   '192.168.154.34'

              + for more values   '192.168.154.44'

Sending task priority  . . . . .   *SAME        1-99, *SAME, *SYSDFT
                                                                  More...
F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
F24=More keys
```

Figure 15-5 Change remote journal example

Note: Notice the data port services section near the bottom of the panel above. That support is new for 6.1. It is here that you can confirm that the remote journal connection has linked up with the *data port service* feature. Read more about this new feature below.

When you page forward you will the panel shown in Figure 15-6.

```
                  Change Remote Journal (CHGRMTJRN)

Type choices, press Enter.

Synchronous sending time-out . .   *SAME         1-3600, *SAME, *SYSDFT
Validity checking  . . . . . .   *DISABLED     *SAME, *DISABLED, *ENABLED

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

Figure 15-6 Change remote journal example (continued)

Note: Notice the two new 6.1 features (*Synchronous remote journal time out* and *Validity checking*) whose status can be viewed and revised on the panel shown in Figure 15-6. Those features are described in more detail below.

A lot of new choices

Among the new 6.1 features that you can select with the CHGRMTJRN command are those described below.

Data port services

Data port services is a layer of internal software used to provide transmission support over TCP/IP lines between machines. Formerly, it was used exclusively for cluster traffic, while the remote journal traffic traditionally flowed over a different kind of connection and was *not* allowed to use the data port services style of data flow. That all changes for 6.1.

While most of the *cluster* communication traffic continues to be handled by this layer of software (data port services), it has been opened up to allow *remote journal* traffic. A matching keyword also has been introduced on the CHGRMTJRN command. Use of the data port services software for a remote journal is accomplished by specifying the data port services (DTAPORTSRV) parameter.

Doing this may be attractive if you want to allow the remote journal traffic to flow over *more than one line* at the same time (that is, if you want multiple *parallel* paths all servicing the same journal). The data port services style of connection allows you to define up to four IP addresses associated with the target system instead of just the single address that was allowed with the prior relational-database-entry connection style.

As a consequence, use of the *data port services* style of connection along with use of additional parallel communication lines may lead to enhanced resiliency regarding the communications data flow to the target system in a remote journal environment. This means that if the path associated with one of the IP addresses on the target system becomes inoperable, you still have up to three additional addresses for the journal entries to be transmitted over.

The resulting remote journal traffic may then be limited to only three lines instead of four, but the remote journal connection stays up. Hence, a data port services flavored remote journal connection may be more robust than a single-line relational database entry connection.

Note: In order to use data port services *both* the source and target systems must be configured as cluster nodes and both must be active in the same cluster.

Synchronous sending time out

Remote journal connections can be set up either to provide *synchronous* transmission from the source to target machine or to employ asynchronous transmission. Obviously, those environments where no risk of loss of in-flight data is tolerable will elect to use the synchronous variety. That style, however, comes at a cost, in that the application performing database operations on the source side will *wait* both for the transmission to complete and for it to be acknowledged. It is this idle time spent waiting for these actions (confirmation that the new journal entries have arrived on the target side) that is the focus of the 6.1 enhancement described here.

Waiting a modest amount of time might be tolerable and a proper trade-off for gaining the confidence that your changes are safely on the target side before proceeding. Waiting a substantial amount of time can be onerous. Waiting in vain can be frustrating. The concern addressed by the new parameter is how long your application should wait before it concludes that the communication line must be down and hence waiting any longer would be counterproductive?

In the past, the quantity of time that your application might wait was a fixed value shipped with the operating system. You could not change it. This meant that some synchronous-flavored remote journal environments ended up waiting in vain far longer than was appropriate for the response time objectives, while others gave up much too easily when a particularly large binary large object (BLOB) field was modified and the matching journal entry had to be transported down the communication line. Neither condition (waiting in vain nor throwing in the towel) was desirable. Instead, what was needed was a knob by which each sync-flavored remote journal environment could customize the wait time, selecting a value that made sense for their needs.

In order to place an upper limit on the wait time, a new parameter was added in 6.1. It applies only to synchronous flavored remote journal environments. The *synchronous sending time-out* (SYNCTIMO) parameter allows you to define a maximum amount of time the source system will wait for a response from the target system. You can see the range of allowable values in Figure 15-6 on page 365.

Once the time-out value has been exceeded, the remote journal connection is deactivated. By doing so, the application that had been waiting is freed up and can continue.

Once freed up, the corresponding database change ensues on the source side (even if the remote journal wait timed out) and a matching journal entry on the source side is produced. The remote journal connection will be deactivated at that point. It is this journal entry that, as a result of the time out, may not have reached the target side.

Once the communication line problems are addressed and the remote journal connection is reactivated the two sides will compare journal receivers, and any entries present on the source but missing from the target will be resent.

The benefit derived by having this new knob is that remote journal users need no longer rely on a hard-coded time-out value provided by the operating system itself but can, instead, customize the time-out value to be a better fit for their needs. Those shops that occasionally expect to encounter busy/overloaded communication lines (maybe during a batch job) or are likely to modify journaled files housing huge BLOB fields (housing images so large that they will take a substantial amount of time to be transported) may want to select a larger time-out value if they have configured the synchronous variety of remote journal. On the other hand, shops concerned that interactive production work could occasionally stall for an unacceptable time duration if the line goes down may want to cap the total wait time.

It is clearly a trade-off. Give up too quickly and your remote journal connection gets deactivated. Be too patient and you have slowed down the application for a persistent communication problem that will not rapidly resolve itself. The good news is that starting with 6.1 you no longer have to settle for an operating system one-size-fits-all time-out value but can, instead, select a setting that makes sense for your needs and your environment.

Validity checking

Nobody likes garbled data and the remote journal is no exception. What is sent as a packet down the wire from the production system to the target system needs to be a trustworthy representation of the changes that occurred. A few bits dropped or altered along the way is not acceptable and yet, prior to 6.1, there were instances in which garbled data was not consistently detected. That stems in part from the fact that prior to 6.1 the only validity checking software employed was the industry-standard TCP/IP error detection algorithms and, frankly, they simply are not robust enough to catch all transmission errors. Remote journal needed access to a more robust approach.

Release 6.1 addresses that concern head on by providing a new *validity checking* (VLDCHK) parameter that allows you to turn on additional communication validity checking to verify that the data that comes out the far end of the wire is the same that was put into the communication line on the source side. This is accomplished by attaching a *cyclic redundancy check* (CRC) value to each set of remote journal images sent.

If the target system detects a garbling error in the data, the data will not be written to the remote journal on the target side. The remote journal environment will be deactivated and an error message indicating a communications failure will be sent to both the journal message queue and to QHST. Seeing such an error message suggests that your underlying communication hardware is misbehaving.

Obviously, it takes some CPU cycles to produce and verify such a CRC on both ends of the transmission. For that reason, the use of this new feature is optional. Shops with very clean lines may derive very little additional benefit from enabling the new support since there is so little likelihood of garbling in the first place. On the other hand, shops that have recently installed new communication gear may want to gain some confidence that hardware switches are maintaining and transmitting a clean unaltered image by enabling the new support for a few days.

Note: Enabling the validity checking may impact performance for the remote journal environment, adding a modest amount of additional CPU overhead on both the source and target ends of the wire.

15.3.2 Work with Journal Attributes (WRKJRNA) enhancements

There are times when it is helpful to dig deeper into the remote journal environment so as to analyze how the transmission mechanism is behaving. This can be especially important if you are attempting to determine whether your remote journal configuration is optimal. Too many protected objects all associated with the same journal can lead to sluggish behavior as well as transmission bottlenecks. To help assist with such analysis, new options have been added in 6.1 to the Work with Journal Attributes (WRKJRNA) command, shown in Figure 15-7.

```
                        Work with Journal Attributes

Journal  . . . . . :   J1             Library  . . . . . :   LWY

Attached receiver  . :   R1           Library  . . . . . :   LWY

Text . . . . . :   *BLANK

ASP  . . . . . :   1              Journaled objects:
Message queue  . . . :   QSYSOPR          Current  . . . . . :         1
  Library  . . . . . :     *LIBL          Maximum  . . . . . :    250000
Manage receivers . . :   *SYSTEM         Recovery count . . . :   *SYSDFT
Delete receivers . . :   *NO             Receiver size options:   *RMVINTENT
Journal cache  . . . :   *NO                                      *MAXOPT2
Manage delay . . . . :   10             Fixed length data  . :   *JOB
Delete delay . . . . :   10                                      *USR
Journal type . . . . :   *LOCAL                                  *PGM
Journal state  . . . :   *ACTIVE
Minimize entry data  :   *NONE

                                                                      Bottom
F16=Work with remote journal information   F24=More keys
```

Figure 15-7 Sample of WRKJRNA panel

Note: Notice the F16 key on this panel. It allows us to bring up the new remote journal information provided by 6.1.

Provided that you have a remote journal connection configured, when you press F16 you will see the panel shown in Figure 15-8.

```
                    Work with Remote Journal Information

Journal  . . . . . :   J1              Library  . . . . . :   LWY

Journal type . . . . :   *LOCAL         Journal state  . . . :   *ACTIVE
Remote journal type  :                  Delivery mode  . . . :
Local journal  . . . :                  Source journal . . . :
  Library  . . . . . :                    Library  . . . . . :
  ASP group  . . . . :                    ASP group  . . . . :
  System . . . . . :                      System . . . . . :
Redirected receiver library . . . . . . . . . . . . . . . . . :   *NONE
Number of remote journals . . . . . . . . . . . . . . . . . . :   1

Type options, press Enter.
  8=Display relational database detail ...   13=Activate   14=Inactivate ...
                        --------------------Remote--------------------
     Relational                                      Journal     Delivery
Opt  Database            Journal     Library        State       Mode
 5   NODE4               J1          LWY            *ACTIVE     *ASYNC
                                                                        Bottom
===>
F3=Exit       F4=Prompt    F5=Refresh   F6=Work with remote journal list
F9=Retrieve   F12=Cancel   F23=More options
```

Figure 15-8 WRKJRNA panel with a list of remote journal connections on the bottom

Note: Notice the presence of the remote journal connection listed on the bottom of the panel shown in Figure 15-8 on page 369.

If you then use option 5 for the remote journal whose statistics you want to see, the panel shown in Figure 15-9 will appear.

```
                      Display Remote Journal Details

Remote journal . . . :   J1             Library  . . . . . :   LWY

Relational database  . . . . . . . . . . . . . :   NODE4
Remote journal state . . . . . . . . . . . . . :   *ACTIVE

Data port services:
  Node identifier  . . . . . . . . . . . . . . :   *NONE
Remote journal type  . . . . . . . . . . . . . :   *TYPE1
Remote delivery mode . . . . . . . . . . . . . :   *ASYNC
Remote journal last catchup date . . . . . . . :   05/15/08
Remote journal last catchup time . . . . . . . :   10:08:56
Remote journal active state date . . . . . . . :   05/15/08
Remote journal active state time . . . . . . . :   10:08:56
Remote journal receiver library  . . . . . . . :
Sending task priority  . . . . . . . . . . . . :   0
Validity checking  . . . . . . . . . . . . . . :   *DISABLED
Number of entries behind . . . . . . . . . . . :   0
                                                                  More...
Press Enter to continue.

F3=Exit   F11=Display relational database detail   F12=Cancel
```

Figure 15-9 The remote journal statistics panel

This is a two-panel panel. The Number of entries behind field shown on the bottom reveals whether your communication lines are adequately sized and keeping up.

Note: The term *behind* as used on these panels means *not yet sent*.

If you consistently see large values here, or worse yet ever-growing values, your lines are either undersized or dirty.

You can see the rest of the statistics if you page forward, revealing the items shown in Figure 15-10.

```
                    Display Remote Journal Details

Remote journal . . . :   J1             Library  . . . . . :   LWY

Relational database  . . . . . . . . . . . . . :   NODE4
Remote journal state . . . . . . . . . . . . . :   *ACTIVE

Maximum entries behind . . . . . . . . . . . . :   2
Maximum entries behind date  . . . . . . . . . :   05/15/08
Maximum entries behind time  . . . . . . . . . :   10:09:45
Hundredths of seconds behind . . . . . . . . . :   3
Maximum hundredths of seconds behind . . . . . :   5
Maximum hundredths of seconds behind date  . . :   05/15/08
Maximum hundredths of seconds behind time  . . :   10:08:59
Number of bundles sent . . . . . . . . . . . . :   3
Maximum bundle size  . . . . . . . . . . . . . :   558
Maximum bundle size date . . . . . . . . . . . :   05/15/08
Maximum bundle size time . . . . . . . . . . . :   10:09:13
Super bundle count . . . . . . . . . . . . . . :   0
                                                                  Bottom
Press Enter to continue.

F3=Exit   F11=Display relational database detail   F12=Cancel
```

Figure 15-10 The rest of the remote journal statistics

The new fields shown on these panels help reveal what is going on (especially the Maximum entries behind, the Hundredths of seconds behind, and the maximum bundle size).

Among the new items of particular interest, take note of the following:

- The status of data port services connections

  Prior to release 6.1 the l*ogical replication* style of high availability offerings and the *cluster-flavored* approaches used little software in common even though they both performed some similar actions.

  A good illustration of this behavior manifested itself in the use of the underlying TCP/IP communication lines. Most logical replication approaches used remote journal technology, which in turn tied directly into a low-level direct TCP/IP transport layer of software that handled not only the transport itself but also caching behavior to help achieve highly optimized performance, while the varieties built upon clustering tied into a customized variety of support known as data port services.

  It made sense to fold new enhancements into one shared approach that both cluster communications and remote journal users could employ. This stems from the fact that the clustered approaches and the remote journal driven logical replication approaches both need to accomplish a similar goal: Rapidly and efficiently transport large quantities of data from the production system to a target system.

  Starting in 6.1, the remote journal transport can tie into either the traditional low-level remote-journal-specific transport support (which has been highly customized over the years to assure rapid high volume transport for remote journal users, but only employs one TCP/IP line) or the newer but more generic *data port service*s support.

The corresponding area on the display labeled data port reveals whether the newer style of support is being utilized, and if so which TCP/IP lines are in use.

- Last catchup phase date and time

  Each time that a remote journal connection is restarted it passes through a so-called catchup phase. It is during this ramp-up phase that any lingering journal entries resident within the local journal on the production system (but not yet present in the remote journal on the target system) are sent in bulk mode (rather than in smaller packets). Until this priming phase is complete, the remote journal connection is not yet fully active. By taking note of the most recent time this catchup phase was executed you can get a sense of the stability of your remote journal environment and the communication lines servicing it.

- Active state date and time

  This shows when the remote journal environment was last activated.

- Validity checking status

  This area of the display reveals whether the extra overhead associated with producing a remote journal CRC and thereby validating proper transmission is enabled.

- Remote journal statistics

  When the *asynchronous* flavor of remote journal transmission is employed, a separate microcode task, executing on the source side, is assigned the responsibility of assuring that recently produced journal entries are placed on the wire and sent down the communication line, as illustrated in Figure 15-11. In fact, it is the presence of this separate task that truly makes the behavior asynchronous (your application does not do the work, hence it is not forced to wait).

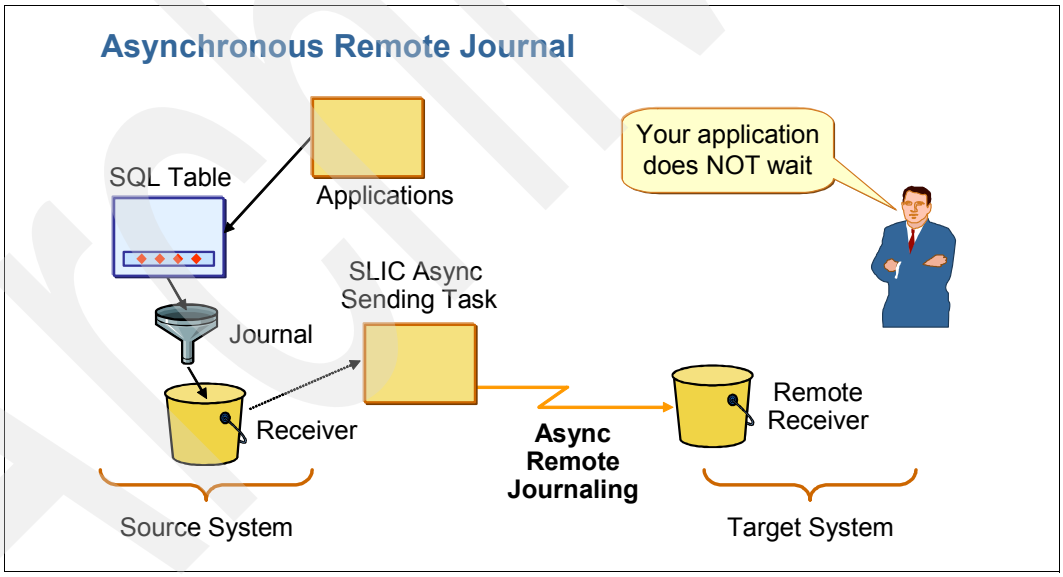


Figure 15-11 Asynchronous remote journal and its use of separate server tasks

However, since there are only a limited number of asynchronous microcode tasks assigned to this mission and since there could be thousands of application jobs enqueuing such work, any sluggishness associated with the communication line, any resends due to the presence of a dirty line, and even any use of suboptimal journal configuration settings (too much chaff sent, bundles too small) can all contribute to a build-up of journal entries enqueued waiting to be sent.

Seeing a large number of unsent entries still queued up (or worse yet an ever-growing quantity of unsent entries) suggests a major traffic jam that deserves attention. Prior to release 6.1 it was difficult to sense that such a backup was forming.

The primary purpose of these new statistics (Maximum entries behind, Hundredths of seconds behind) is to help you detect and monitor the presence of unsent entries (if any). Should you sense that the quantity of unsent bytes is increasing, you know that you have work to do. Perhaps your communication line is undersized. Perhaps your communication network is experiencing a high rate of retransmissions. Release 6.1 now gives you a peek deep within the inner workings of this asynchronous aspect of the operating system so that you can spot trouble before it grows into a serious concern.

How to react

Should you sense a traffic jam there are two primary responses:

- Use the NETSTAT command to discern whether your communication line is experiencing a high degree of re-transmissions. If so, investigate your communication hardware switches and lines and clean them up.
- If no line troubles are detected you may simply have an undersized line. While you can certainly consider adding more lines or stepping up to a faster line, before you do that make sure that you have trimmed as many unneeded bytes (often called chaff) from your journal entries as is practical.

Note: For a more in-depth discussion of ways to minimize the chaff see the technote "Journaling - Journal Receiver Diet Tip 2: Consider using skinny headers" found at:

`http://www.redbooks.ibm.com/abstracts/tips0654.html`

When you execute the NETSTAT command you will see the panel shown in Figure 15-12.

```
                    Work with TCP/IP Network Status
                                                          System:   NODE3
Select one of the following:

     1. Work with IPv4 interface status
     2. Display IPv4 route information
     3. Work with IPv4 connection status
     4. Work with IPv6 interface status
     5. Display IPv6 route information
     6. Work with IPv6 connection status

    10. Display TCP/IP stack status









Selection or command
===>

F3=Exit    F4=Prompt    F9=Retrieve    F12=Cancel
```

Figure 15-12 NETSTAT panel

Select option 3.

You may have to page down a few times until you see the connection for your active remote journal session, as shown in Figure 15-13.

```
                    Work with IPv4 Connection Status
                                                                System:   NODE3
Type options, press Enter.
  3=Enable debug   4=End   5=Display details   6=Disable debug
  8=Display jobs

     Remote           Remote     Local
Opt  Address          Port       Port       Idle Time  State
     *                *          427        022:38:26  *UDP
     *                *          4800       022:38:25  Listen
     *                *          hprip-ctl  022:40:07  *UDP
     *                *          hprip-n >  022:40:07  *UDP
     *                *          hprip-h >  022:40:07  *UDP
     *                *          hprip-med  022:40:07  *UDP
     *                *          hprip-low  022:40:06  *UDP
   5 9.5.168.212      rmtjour >  18975      000:00:17  Established
     9.5.168.212      as-rmtcmd  33127      001:47:17  Established
     9.5.168.212      as-rmtcmd  34287      001:23:16  Established
     9.10.126.242     4279       telnet     000:00:03  Established
     9.65.152.175     2309       telnet     000:00:00  Established
                                                                    More...
F3=Exit   F5=Refresh   F9=Command line   F11=Display byte counts   F12=Cancel
F20=Work with IPv6 connections   F22=Display entire field   F24=More keys
```

Figure 15-13 Illustration of remote journal connection on NETSTAT panel

Note: You can recognize the remote journal connection by the fact that you see the value rmtjour under the remote port column.

If you use option 5 to display the details for this remote journal connection you should see a panel that reveals whether the remote journal communication path is experiencing a high degree of retransmissions. See Figure 15-14. (You may have to page down a few times to find the panel housing the retransmission counts.)

```
                    Display TCP Connection Status
                                                          System:   NODE3
Retransmission information:
  Total retransmissions . . . . . . . . . . . . :   0
  Current retransmissions . . . . . . . . . . . :   0
Send window information:
  Maximum size  . . . . . . . . . . . . . . . . :   262144
  Current size  . . . . . . . . . . . . . . . . :   262144
  Last update . . . . . . . . . . . . . . . . . :   4153769276
  Last update acknowledged  . . . . . . . . . . :   2723047440
  Congestion window . . . . . . . . . . . . . . :   33612
  Slow start threshold  . . . . . . . . . . . . :   32768
  Maximum segment size  . . . . . . . . . . . . :   1440
Precedence and security:
  Precedence  . . . . . . . . . . . . . . . . . :   0
Initialization information:
  Initial send sequence number  . . . . . . . . :   2723025104
  Initial receive sequence number . . . . . . . :   4153767621

                                                                    Bottom
Press Enter to continue.
F3=Exit      F5=Refresh   F6=Print   F9=Command line   F10=Display IP options
F12=Cancel   F14=Display port numbers   F22=Display entire field
```

Figure 15-14 Connection status information for remote journal connection from NETSTAT panel

Note: Notice the retransmission counts at the top of the panel above. If they are large and climbing, you have evidence that your remote journal traffic is being affected by troublesome conditions in the underlying communication hardware.

The options are:

- Super bundle count

  When journal entries are produced in rapid succession on the production machine the underlying microcode of the operating system attempts to improve efficiency by stringing together successive journal entries and forming them into one resulting concatenated burst of bytes. The resulting string of bytes is called a *bundle* and it represents adjacent journal entries that will be written to disk in unison. Such bundling behavior improves efficiency on the production system. Fewer total disk writes need to be scheduled.

  An even stronger degree of bundling may be called for when a *remote journal* environment is present, a variety of behavior that concatenates separate disk write strings into super-strings. Such behavior is called *super bundling* and it represents the kind of behavior that ensues when the remote journal support senses that multiple journal bundles should be placed into the same packet for the trip down the communication line so as to prevent a bottleneck.

Consistently seeing a large count here for *super bundles* indicates that the communication line may be nearing its capacity and that the remote journal traffic may need to be split into separate streams.

Prior to release 6.1 it was difficult to sense that super bundling overdrive mode was being initiated frequently.

- Synchronous sending time out

  This area of the display shows the value selected and currently in force to limit transmission wait time affecting journal packets en route to the target machine. Since such wait time (and the matching time-out behavior) only applies to applications that sit idle waiting for confirmation of packet arrival, this field is only shown for *synchronous* remote journal environments.

15.3.3 Remote journal message enhancements

Staying abreast of the current status of a remote journal connection (whether it is still up) is an important monitoring consideration in a high availability environment that employs a logical replication approach. That connection becomes the weak spot of such an approach.

When the connection is up and operating efficiently all is well, but if the connection has shut down there is no longer a constant flow of information from the production machine to the target. As a consequence, any attempt to role swap would risk using stale down-level data because the target side has not received any refreshed data since the remote journal connection went down. That is, your recovery point objective might not be achieved.

For that reason, it is good to have prompt positive notification whenever any hiccup ensues. You have always been allowed to associate a message queue with a remote journal and then monitor the messages that arrive. But what if that message queue is long gone by the time someone needs to look back and debug? Release 6.1 now makes that more practical. Beginning in 6.1, there will now be status messages sent to the history log (QHST) when a remote journal environment is ended (or restarted).

15.4 Additional enablers for logical replication

Among other enhancements are those discussed in this section.

Allocate object (ALCOBJ) enhancements for data queues

Shops that elect to use logical replication to achieve high availability depend upon purchased software to replay production-side operations on the target side against replicas of the same critical objects that are being monitored on the source side. For example, if your application changes a *data queue* on the source side, the logical replication software attempts to replay the same enqueue or dequeue operation shortly thereafter on the target side.

For all objects managed by a logical replication approach there is also an initial *priming step* required. That is, when you first identify the critical objects to be replicated, the logical replication software must snapshot a copy of the object (and its contents) from the production side and restore an identical copy on the target side so as to ensure that both the original and the replica start out looking identical.

Performing this priming step for most objects has been practical and not particularly challenging. There has, however, been one object in particular that formely behaved in a different way from the others. That object was a data queue. In earlier releases data queues behaved in a much different manner than database files. Whereas the act of saving a copy of

a database file on the production side produced an image of both the surrounding object (the file and the member) and the contents of the member (the rows residing within), data queues did not. That is, saving a data queue produced a copy of only the surrounding (empty) data queue. Strange though it may sound, the resulting image housed no data queue messages.

As you may imagine, this was quite frustrating for vendors who provided logical replication software, and such behavior made priming the target side with both the data queue object and its contents quite difficult. Worse yet, data queues also did not honor normal locking rules. That is, attempts to restrict access to a data queue by locking it did not truly prevent enqueue or dequeue operations from continuing.

The *save the contents* problem was addressed in the release known as IBM i 5.4 (there is now the new parameter, QDTA(*DTAQ), on the SAVOBJ command to ensure that the contents are saved along with the surrounding queue), but the locking issue persisted. Without modifications to both behaviors (locking and saving the queue contents) it was very difficult for logical replication vendors to properly and confidently prime or refresh data queues. Release 6.1 changes all of that with the result that data queues can now be confidently and properly primed or refreshed by logical replication software (just like database files) even if such queues are actively being modified.

In addition, some logical replication software packages deliberately place a lock on the replica objects residing on the *target* side so that they (the logical replication packages) are the only ones that can modify the replicas. Doing so helps ensure that no programmer executing software on the target side errantly modifies the wrong object (the replica being managed by the logical replication software). Until this 6.1 change (honoring locks for data queues), logical replication packages found it difficult to ensure that the target-side replica remained pristine and in-sync.

Hence, the primary data queue locking change for release 6.1 is that data queues now respond to the presence of locks just like database files.

Note: This, however, is not the default behavior. To induce data queues to behave in this new fashion you must modify their locking attribute behavior. Read more about this and the matching API for making such a change in the Info Center at:

`http://publib.boulder.ibm.com/infocenter/systems/index.jsp?topic=/apis/qmhqcdq.htm`

As a consequence, software-based high availability solutions are now able to ensure that the data queue information about the target system cannot be changed by any job other than the replay/apply job that is maintaining the data queue. This helps ensure that the data on the source and target system will remain in-sync.

End-journal-PF restrictions lifted

Another challenge faced by logical replication vendors has been to reconfigure the journal environment on the source side when necessary in order to *balance* the remote journal traffic across multiple remote journal connections.

Imagine a situation where you initially configured one journal and then determine that too much journal traffic is flowing into that journal. In fact, the communication path to the matching remote journal is struggling to keep up. You would want to identify a few files that are producing the bulk of the traffic and try moving them to a separate journal, right?

What does that take? You've got to *stop* journaling those that you want to move, start journaling that subset to a separate *local* journal, and then create a matching second *remote* journal.

This sounds simple, yet prior to i 6 orchestrating such a move was troublesome, especially if you drew this conclusion *after* the affected files were open.

Such attempts to better balance the workload were thwarted by a restriction. Simply stated, prior to i 6, attempts to end journal protection for an open file (at least a file open for update that had truly witnessed some updates) was prohibited. You had to wait until the application *closed* the file, which usually meant that you had to wait until the application was shut down.

That restriction (which was quite onerous) no longer persists. It has been lifted for release 6.1 (at least for files not currently in the midst of any open commitment control managed transaction). As a consequence, achieving balanced remote journal flow (and adjusting that balance during your prime shift) has become an easier task to accomplish.

There may also be times when you want to shift work from journal_1 to journal_2 not just for one singular file but for an entire set of related files, perhaps all of those residing within a designated production library. That too has become easier and less error-prone in release 6.1 since the End Journal Physical File (ENDJRNPF) and End Journal Object (ENDJRNOBJ) commands now support *generic object names*. In addition, a new special value (*ALLLIB) has been added to allow you to end journaling on *all* objects within the specified library.

Balancing your remote journal traffic in this fashion (using two remote journals instead of one) can reap other benefits as well. Not only will you reduce the likelihood that a traffic jam develops on the source side, you will also increase the odds that the target-side HA replay software can keep up. Having two separate remote journals means that you also can configure two separate parallel replay jobs on the target side. This in turn reduces your overall elapsed time required to get caught up following an unplanned role swap and thereby helps you achieve your recovery time objective.

Note: Although the ENDJRNPF enhancements for i 6.1 are a welcome addition, such balancing of traffic across multiple journals can often be avoided if your initial planning step more accurately predicts the total volume of journal traffic likely to ensue. In the past such predictions were a difficult guessing game. The new tool, pseudo journal, can be harnessed to help with this planning step.

15.5 Insuring journal protection

All of the high availability approaches described herein (not just logical replication) will have difficulty achieving both their recovery time objectives and their recovery point objectives unless the user applications being executed are rigorous about promptly flushing recent object changes from main memory to disk. That stems from the fact that nearly all of the approaches replicate only what reaches disk. Hence, images that are still resident in main memory when the machine fails abruptly will be lost.

Loss of this recent data is further compounded by the fact that related changes are not insured to reach the disk surface in the same order in which they were made by the user application.

Say, for example that your application employs a *data area* to keep track of the next customer number to be assigned, uses a *database file* to record the registration information for the new customer, and uses an *IFS file* to capture a scanned image of the customer's signature. In effect, your application is storing related data within three separate objects. At the end of the transaction that enrolls a new customer the *main memory* image shows consistent values for all three of these related values but there is no assurance that the *disk image* has yet received all three changes. In fact, there is a very high probability that the matching images from each

object will *not* reach the disk in the same order in which they were made and that some may linger in main memory minutes longer than others.

For this reason, consistent images across objects, while true in main memory, is certainly unlikely to be true on disk until all files have been closed and all related objects (not just database files) have been flushed from main memory. This means that for a *planned* role swap you are OK (provided that you employ the new 6.1 *quiesce* option prior to the role swap —this quiesce will flush changed pages from main memory). However, such is not the case for an *unplanned* event such as a failover. In fact, for an unplanned event it is a major source of concern. As a consequence, for an unplanned role swap you are likely to be quite surprised regarding the resulting state of your data and its lack of consistency.

Such lack of consistency clearly affects and thwarts your ability to achieve your stated RPO. The fact that some of these object types (database files, for example) must also experience time-consuming internal consistency checks during IPL or iASP vary-on to detect and correct such anomalies will affect your RTO as well.

Journal to the rescue

Such RPO and RTO concerns could have been addressed for our example if each of the objects affected (data area, database file, IFS file) had been enrolled in journal protection and if the matching journal images had been routed to the *same* journal. Hence, the *key step* for both RPO and RTO in this example is to ensure that local journaling is enabled on the production system. This is true whether we are considering a clustered solution or a logical replication solution. Both need journal protection.

This is a consideration easily and often overlooked by people who embrace clustered solutions. They do so at their own risk.

All of this leads us to the conclusion that journal protection should be enabled for *all* critical data areas, data queues, database files, and IFS files that you expect to be in a consistent state in order to achieve your recovery point objectives.

In short, if someone states that they have a clustered HA approach and hence do not need journal protection, they are simply uninformed—a fact that they may not realize until they attempt to recover and realize that they do not have consistent transaction contents across objects.

Generic support when initiating journal protection

As a consequence, the need to enable journaling protection for an entire library worth of objects is often a first step in setting up a robust high availability environment.

Prior to release 6.1, enabling journal protection for a library already housing hundreds of files was challenging, time-consuming, and potentially error-prone since each separate file's name had to be entered individually. What 6.1 provides is a much easier and more efficient way to start journal protection for hundreds of files, data queues, or data areas with a single command, and you no longer need to spell out the name of each object separately. Instead, you can supply generic names on commands such as Start Journal Physical File (STRJRNPF) and Start Journal Object (STRJRNOBJ). Alternatively, you can merely enter the new special value (*ALL), thereby advising the operating system that you want to start journaling on all eligible objects within the specified library. See Figure 15-15.

```
                  Start Journal Physical File (STRJRNPF)

Type choices, press Enter.

Physical file to be journaled  .   T*            Name, generic*, *ALL
  Library  . . . . . . . . . .       LWY         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
```

Figure 15-15 New 6.1 choices on the STRJRNPF command

Note: Notice the first choice on the panel above. We specified a generic name: T*, so that *all* tables (T1, T2, T3, and so on) that begin with the letter T have journaling enabled.

In case you are concerned that issuing such a command against a library that already has some objects journaled might fail, you will be happy to learn that the need to enable journaling on the rest of the objects has also been anticipated and, hence, if you attempt to start journaling a file that is *already* journaled the request will not result in an error (provided that you specify the same journal attributes on the request as are already in use with the journaled file). That is, a no-net-change *start journal* request is treated as a no-op rather than flagged as an error.

In addition, the massive STRJRNPF *ALL request will tolerate any difficulties encountered with one object and keep on progressing. That way it assures that *all* objects that can have journaling enabled are visited and processed. That is, one troubling object encountered along the way does not bring the request to an abrupt halt.

15.6 Less trauma changing journals

With all this journal protection going on, the number of journal receivers consumed is bound to increase. Journal receivers are the objects that fill up. As they near their capacity, new journal receivers need to be created to replace the former ones.

Depending on the rigor with which housekeeping is being practiced, there may well be older instances of journal receivers still lingering on your system when it comes time to create new ones.

15.6.1 The naming convention for journal receivers

The naming convention adopted by most high availability shops is to assign journal receivers a name that starts out with letters and concludes with numbers (for example, JRNRCV9900).

It is similarly common practice to allow the operating system to sense that the former journal receiver is nearing its capacity and to expect the operating system to create a new one. With such a scheme in place the next journal receiver in our example would be JRNRCV9901. As you may anticipate, there comes a time when all numerical suffixes have been exhausted and the system is forced to start over.

Alternatively, some shops elect to schedule their own journal receiver swap operations (creating a new one, detaching the old one). Under these circumstances, a user-provided piece of software (such as a logical replication package) would take on the role of generating a new journal receiver and would often do so by using the Change Journal (CHGJRN) command support with the *GEN option, which means that the system selects the new journal receiver name.

15.6.2 The rules when we wrap

In the past there were two sets of rules regarding what happened when the final name has been consumed. One applied to so-called *user* journals—those employed by products such as logical replication, and the other set of rules for internal system-provided journals. The system-provided journals would not hit a brick wall when journal receiver JRNRCV9999 became full. Instead, the numerical suffix wrapped back to 0000 and the CHGJRN operation succeeded. The user-journals, however, were not as lucky.

As you can imagine, hitting a brick wall and finding that the rapidly filling former journal receiver (JRNRCV9999) could accept no more deposits was not very conducive to achieving high availability.

The good news for 6.1 is that the Change Journal (CHGJRN) command processing has been enhanced to support *receiver name wrapping* for user-managed journals. This means that if your currently attached receiver name is at the maximum value (for example, JRNRCV9999), the system will now generate a new receiver with the receiver number value being wrapped (for example, JRNRCV0000). In effect, the former restriction has been lifted for 6.1, thereby removing the ticking time-bomb that could have brought your production system to an abrupt halt in the past.

And if (often due to poor housekeeping practices) a lingering journal receiver with the name JRNRCV0000 lingers, what then? i 6 has anticipated that, too. It simply increments the numerical suffix a bit to skip over potential lingering receivers from days gone by and continues.

15.7 Finding journal entries by journal identifier

Some applications have a habit of creating an object (like a file) with one name, and then later changing the object's name. Such name changes on the fly can require some careful bookkeeping for logical replication software.

15.7.1 Each object gets a unique and persistent birthmark

At the time that journal protection is first enabled for an object, the object is assigned a unique value called its journal identifier (JID). You can think of this as a birthmark. It is persistent. It does not change when an object experiences a name change or when it is moved from one library to another or even when the object is saved and restored.

In fact, that is one of the reasons that an object created on a production system can be restored on a target system as a priming step and still be recognized by journal commands as the *same* object.

It is also the reason that it can make sense to perform certain housekeeping chores (such as saving off journal receivers) from the *target* side rather than consume CPU cycles on the source side doing so. Similarly, it is the motivation for creating a *vault* housing the remote journal receivers.

All of these instances see the journaled objects through the eyes of the journal identifier—the unique birthmark. Hence, related commands such as Apply Journal Changes (APYJRNCHG) would be able to use either the source-side journal receivers or the target-side remote journal receivers since they agree on the birthmark for each protected object.

15.7.2 Finding what you want by JID rather than by name

The are some challenging recovery scenarios in which the logical replication software needs to look back in a journal receiver, searching for earlier actions regardless of the name the object had at that point in time. In such instances, finding the proper journal entry by JID is a much more attractive and more efficient approach than attempting to search by name. Yet, prior to 6.1 the logical replication products were only allowed to search by name—a name that then had to be mapped into a matching JID before the underlying microcode could perform the search. That restriction has been lifted. For 6.1 they can now do so by JID.

Searching by JID, rather than by name (especially within a *remote* journal), is a much more efficient process. Hence, it is faster and does not consume nearly as many resources on the target machine as the prior practices, which in turn means that the replay processing performed by third-party software should not struggle nearly as much to keep up.

People investigating and debugging logical replication software often want to see the matching journal entries on a panel but may not know what name the matching journaled object had at the time when the journal entry was deposited. In such an instance it is attractive (and less error-prone) if one can merely search/display the journal receiver contents by JID rather than by fully qualified object name. Prior to release 6.1 that was not practical. Now it is.

15.7.3 The DSPJRN command has been enhanced to help with such searches

This is accomplished, once you know the JID of the object that you are interested in, by using the Display Journal (DSPJRN) command. This has been enhanced to allow you to display

journal entries for objects based on their object journal identifier value rather than by their name. In order to accomplish this, a new parameter (OBJJID) has been added for this purpose.

If you prompt on the DSPJRN command, page forward a few panels until you reach the final one. You will see the new OBJJID parameter, as shown near the bottom of the panel shown in Figure 15-16.

```
                          Display Journal (DSPJRN)

Type choices, press Enter.

Job name . . . . . . . . . . . . JOB           *ALL
  User . . . . . . . . . . . . .
  Number . . . . . . . . . . . .
Program  . . . . . . . . . . . . PGM           *ALL
User profile . . . . . . . . . . USRPRF        *ALL
Commit cycle large identifier  . CCIDLRG       *ALL
Dependent entries  . . . . . . . DEPENT        *ALL
Output format  . . . . . . . . . OUTFMT        *CHAR
Include hidden entries . . . . . INCHIDENT     *NO
File identifier  . . . . . . . . OBJFID

                           + for more values

Object journal identifier  . . . OBJJID _________________
                           + for more values
Output . . . . . . . . . . . . . OUTPUT        *
                                                                      Bottom
F3=Exit   F4=Prompt   F5=Refresh   F10=Additional parameters   F12=Cancel
F13=How to use this display        F24=More keys
```

Figure 15-16 Use of OBJJID parameter on DSPJRN to search by JID

Let us consider an example of how this new support might be used. Let us say that you had previously displayed a journal entry of interest (like the creation entry F/JM) for a new member, T1. No matter what its new name is, by the time you decide to look for journaled actions against T1, if you recall the JID you can find *all* the entries you need—and find them faster—by supplying the JID (and be sure that you are looking at the entries for the original T1, even if some *other* object is now named T1).

Figure 15-17 shows the original F/JM journal entry, produced when the member was first journaled. Notice th**e *Journal Identifier* n**ear the bottom of the panel. This is the JID value that serves as a permanent birthmark. It is what we can use for our search.

```
                      Display Journal Entry Details

Journal  . . . . . :   J1             Library  . . . . . :   LWY

Sequence . . . . . :   5
Code . . . . . . . :   F  - Database file member operation
Type . . . . . . . :   JM - Start journaling for member

Ignore APY/RMV . . . :   No
Ref constraint . . . :   No
Trigger  . . . . . :   No
Program  . . . . . :   QCMD
  Library  . . . . . :     *OMITTED
  ASP device . . . . :     *OMITTED
System sequence  . . :   0
Thread identifier  . :   *OMITTED
Receiver . . . . . :   R4
  Library  . . . . . :     LWY
  ASP device . . . . :     *SYSBAS
Journal identifier . :   X'92F000152EC008410005'
                                                                      More...
```

Figure 15-17 DSPJRN of the creation F/JM journal entry revealing the JID for T1

Before we rename the object, let us insert a row and delete that same row. We now display the journal and restrict our view to only journal entries associated with T1. We will see the journal entries from the matching object, as shown in Figure 15-18.

```
                        Display Journal Entries

Journal  . . . . . :   J1             Library  . . . . . :   LWY
Largest sequence number on this screen  . . . . . : 00000000000000000013
Type options, press Enter.
  5=Display entire entry


Opt    Sequence  Code  Type  Object      Library     Job         Time
              5   F     JM   T1          LWY         QPADEV000B   2:50:43
              6   F     OP   T1          LWY         QPADEV000B   3:03:56
              7   R     PT   T1          LWY         QPADEV000B   3:03:56
              8   F     CL   T1          LWY         QPADEV000B   3:03:56
              9   F     OP   T1          LWY         QPADEV000B   3:04:24
             10   F     DE   T1          LWY         *OMITTED     3:04:25
             11   R     DL   T1          LWY         QPADEV000B   3:04:25
             13   F     CL   T1          LWY         QPADEV000B   3:04:25




                                                                      Bottom
F3=Exit   F12=Cancel
```

Figure 15-18 Resulting journal entries when displayed by JID

If we then renamed T1 to T2 using a command such as RNMOBJ OBJ(LWY/T1) OBJTYPE(*FILE) NEWOBJ(T2) and inserted a new row into T2 but wanted to see *all* the activity that had ensued against this object (both when it had originally been named T1 as well as after it became T2), we could search by its JID value:,92F000152EC008410005 (which we had extracted from the F/JM entry when the physical file was still named T1). The command would look something like this:

```
DSPJRN JRN(LWY/J1) INCHIDENT(*YES) OBJJID(92F000152EC008410005)
```

The resulting set of journal entries (all from the same object) would be shown as revealed by the DSPJRN panel shown in Figure 15-19.

```
                          Display Journal Entries

Journal  . . . . . :   J1             Library  . . . . . :   LWY
Largest sequence number on this screen  . . . . . : 00000000000000000018
Type options, press Enter.
  5=Display entire entry


Opt    Sequence  Code  Type  Object      Library     Job         Time
              5   F     JM    T1          LWY         QPADEV000B  2:50:43
              6   F     OP    T1          LWY         QPADEV000B  3:03:56
              7   R     PT    T1          LWY         QPADEV000B  3:03:56
              8   F     CL    T1          LWY         QPADEV000B  3:03:56
              9   F     OP    T1          LWY         QPADEV000B  3:04:24
             10   F     DE    T1          LWY         *OMITTED    3:04:25
             11   R     DL    T1          LWY         QPADEV000B  3:04:25
             13   F     CL    T1          LWY         QPADEV000B  3:04:25
             15   F     MN    T1          LWY         QPADEV000B  3:21:41
             16   F     OP    T2          LWY         QPADEV000B  3:22:37
             17   F     DE    T2          LWY         *OMITTED    3:22:37
             18   R     PX    T2          LWY         QPADEV000B  3:22:37
                                                                      More...
```

Figure 15-19 Results of performing a search by JID for a PF whose name has changed

Note: The F/MN entry is the rename operation. Thereafter, the PF is known as T2. All journal entries shown on this panel have the same JID.

15.8 Assuring efficient operation and low overhead

Local journaling protection on the production system is an essential first step not only for *logical replication* high availability approaches but for *cluster-driven* approaches as well. This logging behavior comes at a price. The presence of journaling protection adds both CPU and elapsed time overhead to each application operation that modifies a journaled object (database file, data area, data queue, IFS file).

15.8.1 The trade-off: To cache or not to cache

In a high availability environment you face a trade-off. On the one hand you probably want to ensure timely transfer of recent application changes from main memory out to disk on the production system (for this is the trigger mechanism that initiates the sector-driven replication of both global and metro mirroring as well as the page-driven replication of geo mirroring). It is also the trigger mechanism if asynchronous-flavored remote journal transport is being employed to drive a logical replication approach.

On the other hand, in an environment with a heavy use of batch processing such journal entries tend to arrive *one-by-one*. If each separate journal entry experienced a separate trip

out to disk, such behavior would put a substantial strain on the underlying set of disks and slow the pace of the batch job noticeably.

As a consequence, many shops elect to settle on a compromise. They grant the underlying operating system permission to bundle up multiple new journal entries all destined to be written to the same journal, accumulate a string of consecutive journal entries in main memory, and then write them in unison to the disk surface. This technique is known as *journal caching* and it can be a powerful technique for ensuring an efficient trade-off between the needs of a timely RPO and the needs of good response times for interactive users. Figure 15-20 provides an illustration of the kind of performance benefit (especially for batch jobs) that often ensues as a result.

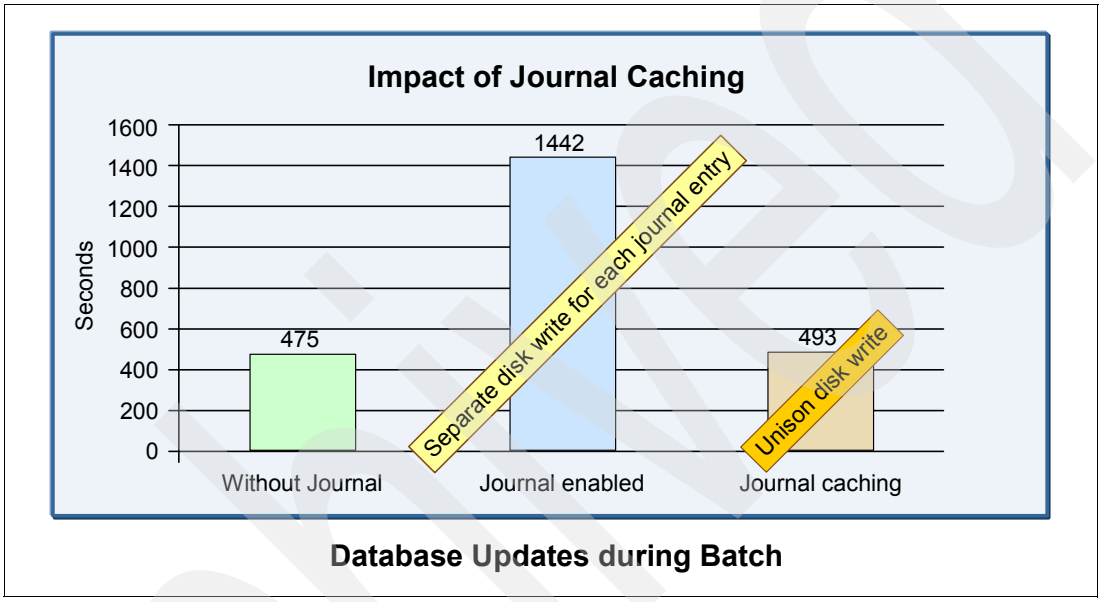


Figure 15-20 Performance benefits often associated with use of journal caching

Such performance gains for batch work can be highly intoxicating. Yes, improved batch performance (shorter runtime duration) can be extremely attractive. However, one must also consider the RPO consequences: The images linger longer in main memory on the source side, so if the source machine were to have an unplanned outage where mainstore is not preserved a few more transactions may be lost.

Note: A good discussion of the factors that you will want to weigh as you ponder this decision (to cache or not to cache in an HA environment) is found in the technote at:

`http://www.redbooks.ibm.com/abstracts/tips0627.html`

In order to select this performance-enhancing option, one must install and configure the journal caching feature (also known as option 42 of the operating system) on the source/production machine followed by enabling caching behavior for the selected journal with the CHGJRN command, specifying the JRNCACHE(*YES) parameter.

15.8.2 Gaining control of journal caching/flushing frequency

Prior to release 6.1, if you elected to enable the journal caching feature the frequency with which the accumulating journal entries left main memory and traveled to disk was driven by a default value selected by the operating system itself.

For some environments the time duration was longer than felt comfortable for achieving your RPO and the resulting recovery point could feel stale. For other environments the time felt so short that full benefits of performance improvement were not realized. What was needed was a knob by which the latency value could be customized for each environment. Release 6.1 supports a new feature that makes that possible. It is adjusted with the Change Journal Attributes (CHGJRNA) command by using the new CACHEWAIT parameter, as shown in Figure 15-21.

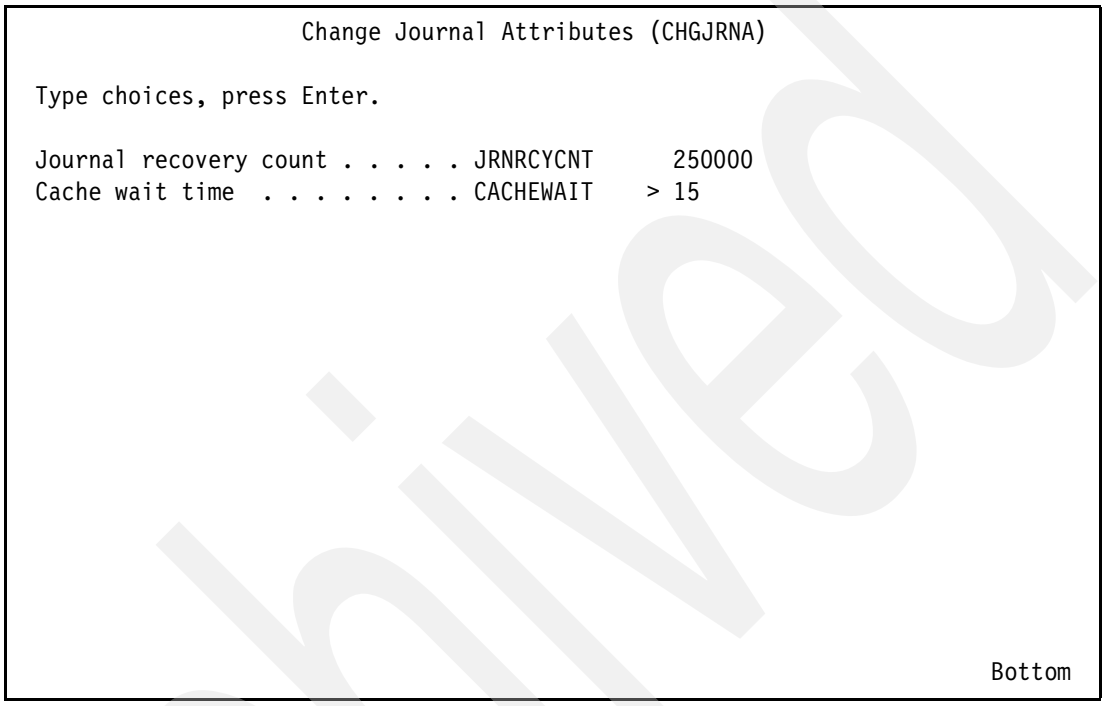


Figure 15-21 CHGJRNA command: Illustrating new 6.1 features

Note: Notice the opportunity to override the system-provided default *cache-wait time* for journal bundling on the panel above.

While tamping down the production-side overhead of journal protection by employing a larger value for the *cache wait time* can be appreciated by users, you will not want to go overboard. The longer that journal entries are allowed to linger in main memory, not yet written to the disk surface on the production side, the longer they similarly linger on the production side not yet *sent* to the target system by means of the remote journal transport mechanism.

Note: For a more in-depth discussion of this trade-off see the technote “Journal caching: Understanding the risk of data loss” at:

http://publib-b.boulder.ibm.com/abstracts/tips0627.html

Whether we are talking about a *logical replication* HA approach or a *cluster-driven* HA approach, images that linger in main memory are images not yet available for HA recovery.

15.9 Pre-planning for journal protection

One of the important choices faced by high availability planners is selecting the quantity of journals to employ. Here too, there is a trade-off to be considered.

If you configure too many journals and assign a separate journal for each physical file, you complicate the configuration process and may simply end up with more journals to manage than is comfortable. You also limit the ability of the underlying journal microcode on the *source* side to achieve the performance efficiency that it could orchestrate if allowed to bundle together journal entries from separate database files. On the other hand, if you swing the pendulum too far in the other direction and configure only *one* journal and route all object changes through the same singular journal you deny yourself (and your third-party HA logical replication package) the opportunity for increased parallelism during the replay actions on the *target* side.

Selecting the quantity of journals is a trade-off. Have too few (when lots of journaled objects are present) and you can overwhelm a journal and the disks below with far more traffic than can be comfortably absorbed. The matching remote journal traffic might similarly experience overload. Yet, estimating the quantity of additional disk traffic and the matching communication traffic associated with journal protection *before* actually enabling such protection for an application has historically been a challenging guessing game.

Fortunately, there is now a new tool (pseudo journaling) that can be used with release 6.1 to analyze and estimate both the quantity of local disk traffic on the production system and the remote journal traffic associated with each database file that you might elect to journal.

Pseudo journaling to the rescue

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 if journal protection is enabled for a set of designated physical files. This can be especially helpful if you have elected to pursue a high availability approach that makes heavy use of journal protection but do not yet have journal support enabled.

In situations like these it can be especially helpful to let the software help produce an estimate *before* you enable journal protection (especially if you are about to enable such protection for lots of files). The questions that ought to be on your mind are:

- How many journals should I configure? Will the total quantity of journal/disk traffic justify use of more than one journal?
- How much corresponding communication traffic will a remote journal generate and hence how much bandwidth will I need?
- 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 placing a high impact on your system as it performs the analysis and, 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 Web site:

`http://www.iseries.ibm.com/db2/journalperfutilities.html`

15.10 New journal entries

With all this new journal function added to i 6 it makes sense that there would be new flavors of journal entries to match. The following journal entries have been added in i 6.

15.10.1 New entries for library journaling

The introduction of library journaling is the primary 6.1 enhancement that mandates the introduction of several new journal entries. The new *library-journaling* entries include the following:

Y LF	Logical file association
Y YA	Change library attribute
Y YB	Start journal for library
Y YD	Library deleted
Y YE	End journal for library
Y YH	Library changes applied
Y YI	Library in use at abnormal end
Y YK	Change journaled object attribute
Y YN	Library renamed
Y YO	Object added to library
Y YS	Library saved
Y YW	Start of save for library
Y YY	APYJRNCHG command started
Y YZ	Library Restored
Y ZA	Change authority
Y ZB	Object attribute change
Y ZO	Change owner
Y ZP	Change primary group
Y ZT	Change audit attribute

15.10.2 Additional new journal entries

Along with the new journal entries for library journaling, the following entries have also been added:

D LF	Logical file association
J MJ	Journal receiver moved
J ZA	Change authority for receiver
J ZB	Object attribute change for receiver
J ZO	Change owner for receiver
J ZP	Change primary group for receiver
J ZT	Change audit attribute for receiver
Q QG	Data queue attribute changed
T XD	Directory services extension

15.11 Best journal practices checklist

The use of journal protection is a common denominator for all of the high availability approaches. Along with commitment control, journal helps assure transaction integrity and a timely recovery point. However, there is a *runtime overhead* associated with achieving such protection.

The runtime overhead is directly proportional to the care with which journal protection is configured. Those shops that follow best practice guidelines have only modest overhead, while those that have given scant consideration to best practices can often experience far more overhead than is necessary. You would obviously like to be among the first group. Listed below is a checklist (in no particular order—what is best for your shop may vary from what is best for another) of practices that may yield performance benefits in a journal-intensive environment.

- ☐ Analyze your IOA write cache performance characteristics, and if you determine that your quantity of fast writes is less than 99%, install more write cache.

  Note: For a more detailed discussion of this topic along with some rules-of-thumb see the matching technote at:

  `http://publib-b.boulder.ibm.com/abstracts/tips0653.html`

- ☐ If you elect to place your journal receivers in a private ASP, do not skimp on the quantity of disk arms available for use by the journal.

  Note: For a more in-depth discussion of use of a private ASP for housing your journal receivers see the matching technote at:

  `http://publib-b.boulder.ibm.com/abstracts/tips0602.html`

- ☐ Customize your journal entries to limit the quantity of descriptive metadata. There is no use capturing or transporting more than you need.

- ☐ Consider employing the MINENTDTA attribute for your journals. Lots of excess bulk (which most logical replication products do not need anyway) can be sent down the remote journal connection in vain unless you put your journal on a diet. In short, do not settle for the defaults. Convince yourself that you truly need a full row image of every change (including those that only modify a single field) before capturing and sending the excess bulk blindly to the target side.

  Note: The tech note regarding *minimized journal entries*, found a the following Web site, provides more insight into this topic:

  `http://www.redbooks.ibm.com/abstracts/tips0626.html`

- ☐ Provide journal protection for all of your critical files but be sure not to go overboard. Journal protection for your work files is probably wasted effort.

- ☐ Analyze your System Managed Access Path (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, say, 50 minutes, you probably ought to give serious consideration to cranking the value down. (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 may still be operating with outdated settings. Continuing to do so can make the vary-on duration for an iASP exceed your RTO.)

- ☐ Take a peek at your *journal threshold*. This is the value that limits the size of your journal receiver. Any value smaller than 1.5 GB probably ought to be increased so that the overhead of creating and attaching new journal receivers does not become too onerous.

Note: For additional discussion of the *journal threshold* and its impacts see the following technotes:

http://www.redbooks.ibm.com/abstracts/tips0603.html
http://www.redbooks.ibm.com/abstracts/tips0652.html

- Be certain that your journal is employing the most modern defaults. Many journals created prior to release IBM i 5.4 may still be locked into ancient settings, and these ancient settings may be thwarting performance (especially in an HA logical replication environment that employs remote journal support). 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 the CHGJRN command.
- Consider installing and enabling the journal caching feature if journal performance (especially during batch jobs) is a major concern in your shop. This feature not only lessens the overhead on the source side (fewer disk writes need to be scheduled), but also increases the efficiency across the remote journal connection (fewer packets need to be sent). It makes sense if your remote journal sending environment is struggling to keep up.

Note: Be sure to read the matching technote found at:

http://www.redbooks.ibm.com/abstracts/tips0627.html

- Give serious consideration to using the Edit Recovery Access Path (EDTRCYAP) command to set your system behavior to INACCPTH(*ELIGIBLE). This probably makes sense on both the production and the target system, but it is especially important to use this setting on the target side if you have also elected to use the (*STANDBY) setting for the local journal beneath the replica database files residing on the target system.

Note: You can read more about standby mode in the technote found at:

http://www.redbooks.ibm.com/abstracts/tips0628.html

- The more actively-being-modified objects (such as open files) that you have associated with a journal, the higher you may want to set your journal recovery count. Failure to do so will slow your runtime performance and increase your housekeeping overhead on your production system without adding much benefit to your RTO. Increasing this value some may make good sense, but do not get carried away.

Note: For a more in-depth discussion of the journal recovery count see the technote “The Journal recovery count, make it count” 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 you have large access paths whose recovery following a crash may take quite a bit of time (due to their size) and your RTO is pretty tight, you might want to elect to journal the access

path via the Start Journal Access Path command (STRJRNAP), or you may want to elect to allow the IPL (or vary-on in the case of an iASP) to proceed without waiting for the APs to finish their recovery actions. This can be accomplished by employing the RECOVER(*AFTIPL) parameter on the CHGLF command. Doing so is probably wise in an HA environment.

- If your applications tend to produce transactions that consist of less than 10 database changes, you may want to give serious consideration to use of *soft* commitment control.

> **Note:** For a more in-depth discussion of *soft commit* and the trade-offs see the technote found at:
>
> http://www.redbooks.ibm.com/abstracts/tips0623.html

More in-depth treatments of a number of these best practices can be found in the IBM Redbooks publication *Striving for Optimal Journal Performance on DB2 Universal Database for iSeries*, SG24-6286, at:

http://www.redbooks.ibm.com/abstracts/sg246286.html

Also, nearly a dozen journal-related tips can be found as technotes on the IBM Redbooks publication Web site:

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

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 publications

For information about ordering these publications, see “How to get Redbooks publications” on page 395. Note that some of the documents referenced here may 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

Online resources

These Web sites 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

How to get Redbooks publications

You can search for, view, or download Redbooks publications, Redpapers publications, technotes, draft publications and Additional materials, as well as order hardcopy Redbooks publications, at this Web site:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

E

F

G

H

I

J

L

M

N

O

P

Q

R

S

T

U

V

W

Implementing PowerHA for IBM i

Implementing PowerHA for IBM i

Embrace PowerHA for i to configure and manage high availability

Leverage the best IBM i 6.1 HA enchancements

Understand pros and cons of different HA alternatives

IBM PowerHA for i (formerly known as HASM) is the IBM high availability disk-based clustering solution for the IBM i 6.1 operating system. PowerHA for i when combined with IBM i clustering technology 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.

This IBM Redbooks publication provides a broad understanding of PowerHA for i. This book is divided into four major parts:

- Part 1, “Introduction and background” on page 1, provides a general introduction to clustering technology and some background.
- Part 2, “PowerHA for i setup and user interfaces” on page 69, describes and explains the different interfaces of PowerHA for i. It also describes the migration process to this product and sizing guidelines.
- Part 3, “Implementation examples using PowerHA for i” on page 319, explains how to use PowerHA for i with three major ERP solutions: SAP, Lawson M3, and Oracle JD Edwards.
- Part 4, “Other IBM i 6.1 high availability enhancements” on page 349, explains additional IBM i 6.1 announced enhancements in high availability.

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

SG24-7405-00 ISBN 0738431982