IBM PowerHA SystemMirror for i: Preparation
(Volume 1 of 4)

David Granum

Edward Grygierczyk

Sabine Jordan

Peter Mayhew

David Painter

Power Systems

IBM

International Technical Support Organization

IBM PowerHA SystemMirror for i: Preparation

June 2016

Note: Before using this information and the product it supports, read the information in "Notices" on page vii.

First Edition (June 2016)

This edition applies to Version 7, Release 2 of IBM PowerHA SystemMirror for i (product number 5770-HAS).

Contents

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

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

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

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 jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

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

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

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

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

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.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

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 actual people or business enterprises 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. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®	IBM®	Redbooks (logo) ®
DB2®	OS/400®	Storwize®
Distributed Relational Database	POWER®	System i®
Architecture™	POWER Hypervisor™	System Storage®
DRDA®	Power Systems™	System Storage DS®
DS8000®	PowerHA®	SystemMirror®
Electronic Service Agent™	PowerVM®	Tivoli®
FlashCopy®	Redbooks®	
HyperSwap®	Redpaper™	

The following terms are trademarks of other companies:

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.

THIS PAGE INTENTIONALLY LEFT BLANK

Preface

IBM® PowerHA® SystemMirror® for i is the IBM high-availability (HA), disk-based clustering solution for the IBM i operating system. When PowerHA for i is combined with IBM i clustering technology, it delivers a complete HA and disaster-recovery (DR) solution for business applications that are running in an IBM i environment. You can use PowerHA for i to support HA capabilities with either native disk storage, IBM DS8000® storage servers, or IBM Storwize® storage servers.

This IBM Redbooks® publication gives a broad understanding of PowerHA for i and provides a general introduction to clustering technology, independent auxiliary storage pools (IASPs), PowerHA SystemMirror products, and the PowerHA architecture.

This book is part of a four-book volume set that gives you a complete understanding of PowerHA for i and its use of native disk storage, IBM DS8000 storage servers, or IBM Storwize storage servers. The following IBM Redbooks publications are part of this PowerHA for i volume set:

► *IBM PowerHA SystemMirror for i: Using DS8000*, SG24-8403
► *IBM PowerHA SystemMirror for i: Using IBM Storwize*, SG24-8402
► *IBM PowerHA SystemMirror for i: Using Geographic Mirroring*, SG24-8401

> **Important:** The information that is presented in this volume set is for technical consultants, technical support staff, IT architects, and IT specialists who are responsible for providing HA and support for IBM i solutions. If you are new to HA, first review the information that is presented in this book to get a general understanding of clustering technology, IASPs, and the PowerHA architecture. You can then select the appropriate follow-on book based on the storage solutions that you are planning to use.

Authors

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

David Granum is an Advisory Software Engineer at IBM Rochester. He has been with the IBM i Global Support Center (iGSC) throughout his 18 years at IBM. His areas of expertise include PowerHA, performance, communications, and TCP/IP applications on the IBM i. He holds a Computer Science degree from South Dakota State University and a Masters degree in Telecommunications from Saint Mary's University of Minnesota.

Edward Grygierczyk is a part of the IBM Lab Services team and is based in Poland. He joined IBM in 2008, working for IBM STG Lab Services as a Consultant in IBM Power Systems™. Since 1991, before Edward joined IBM, he worked in IBM i to build a highly successful IBM Business Partner company in Poland. Edward has a background in systems, availability, integration, and implementation. Edward was also an author of IBM courses in the IBM i area. Edward can be reached at edward.grygierczyk@pl.ibm.com.

Sabine Jordan is a Consulting IT Specialist working in IBM Germany. She has worked as a Technical Specialist in the IBM i area for more than 20 years, specializing in HA. She has worked on IBM PowerHA SystemMirror for i implementations for both SAP and non-SAP environments that use geographic mirroring, IBM DS8000, and IBM SAN Volume Controller (SVC) remote copy services. Among these implementations, she created concepts for the design and implemented the entire project (cluster setup and application changes), in addition to performing client education and testing. In addition, Sabine presents and delivers workshops (internally and externally) about IBM PowerHA SystemMirror for i, HA, and DR.

Peter Mayhew is an IBM i Technical Services Consultant who works for IT Solutions Group, an IBM Business Partner. He has worked with midrange systems for almost 40 years, beginning as an IBM hardware service representative on the 3x systems, AS/400 CISC, and early RISC iSeries models. After leaving IBM after 23 years, he has maintained his skills on the IBM i, specializing in hardware and operating system installations, upgrades and migrations, and HA planning and implementation, supporting clients throughout the United States. Peter can be reached at Pete@itsolutions-group.com.

David Painter is part of the IBM Rochester Lab Services team and based in the United Kingdom. He joined IBM in 2006, working for IBM STG Lab Services as a Senior IT Specialist in IBM Power Systems, HA, and external storage. Before David joined IBM, he set up and built a highly successful IBM Business Partner company in the UK. David has a background in systems, availability, integration, and implementation. Today, David implements HA solutions across the world in clients that range from small-size to medium-size business (SMB) accounts to large Enterprise accounts. David also regularly teaches courses in IBM i HA and external storage for IBM i administrators. He also conducts many availability workshops and holds numerous certifications. David can be reached at davpaint@uk.ibm.com.

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

Philip Albu
Jenny Dervin
Steven Finnes
Scott Helt
Ben Rabe
Curt Schemmel
James Tilbury
IBM Rochester, IBM i Development and Support teams

Christian Aasland
Laural Bauer
Selwyn Dickey
IBM Systems Lab Services

Tracy Eidenschink
Cindy Mestad
Michael K Meyers
IBM i Performance and Scalability Services (PaSS) Center

Debbie Landon
Ann Lund
International Technical Support Organization

Now you can become a published author, too!

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

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

`ibm.com`/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

   ibm.com/redbooks

► Send your comments in an email to:

   redbooks@us.ibm.com

► Mail your comments to:

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

Stay connected to IBM Redbooks

► Find us on Facebook:

   http://www.facebook.com/IBMRedbooks

► Follow us on Twitter:

   http://twitter.com/ibmredbooks

► Look for us on LinkedIn:

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

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

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

► Stay current on recent Redbooks publications with RSS Feeds:

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

1

Introduction to IBM PowerHA SystemMirror for i

This chapter provides an overview of the IBM PowerHA SystemMirror for i business continuity solutions.

The following topics are described in this chapter:

- ► 1.1, "IBM i business continuity solutions" on page 2
- ► 1.2, "PowerHA solution considerations" on page 9

1.1 IBM i business continuity solutions

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

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

Figure 1-1 shows how an HA solution can help you to reduce planned and unplanned application downtimes significantly.
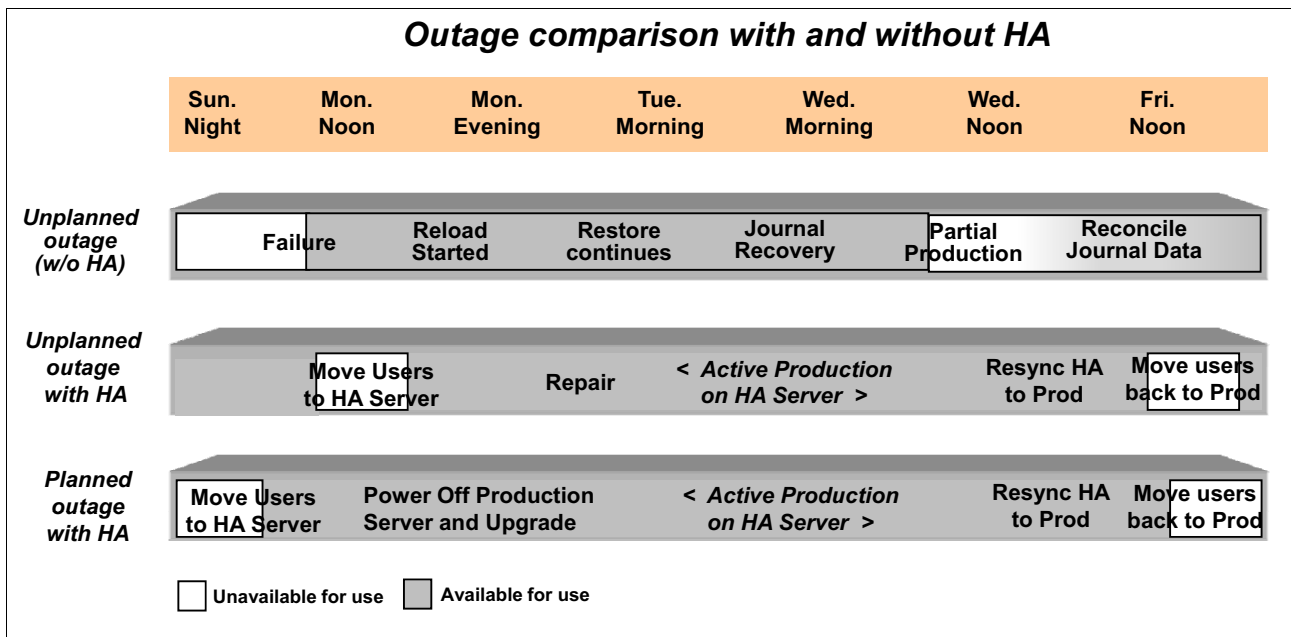
Figure 1-1   Outage comparison with and without HA

To address the client needs for HA for their IBM i environment, IBM provides the IBM PowerHA SystemMirror for i (PowerHA) product.

1.1.1 IBM PowerHA SystemMirror for i solutions

The IBM PowerHA SystemMirror for i solution (Figure 1-2) offers a complete end-to-end integrated clustering solution for HA and disaster recovery (DR). PowerHA provides a data and application resiliency solution that is an integrated extension of IBM i operation system and storage management architecture and has the design objective of providing application HA through both planned and unplanned outages.
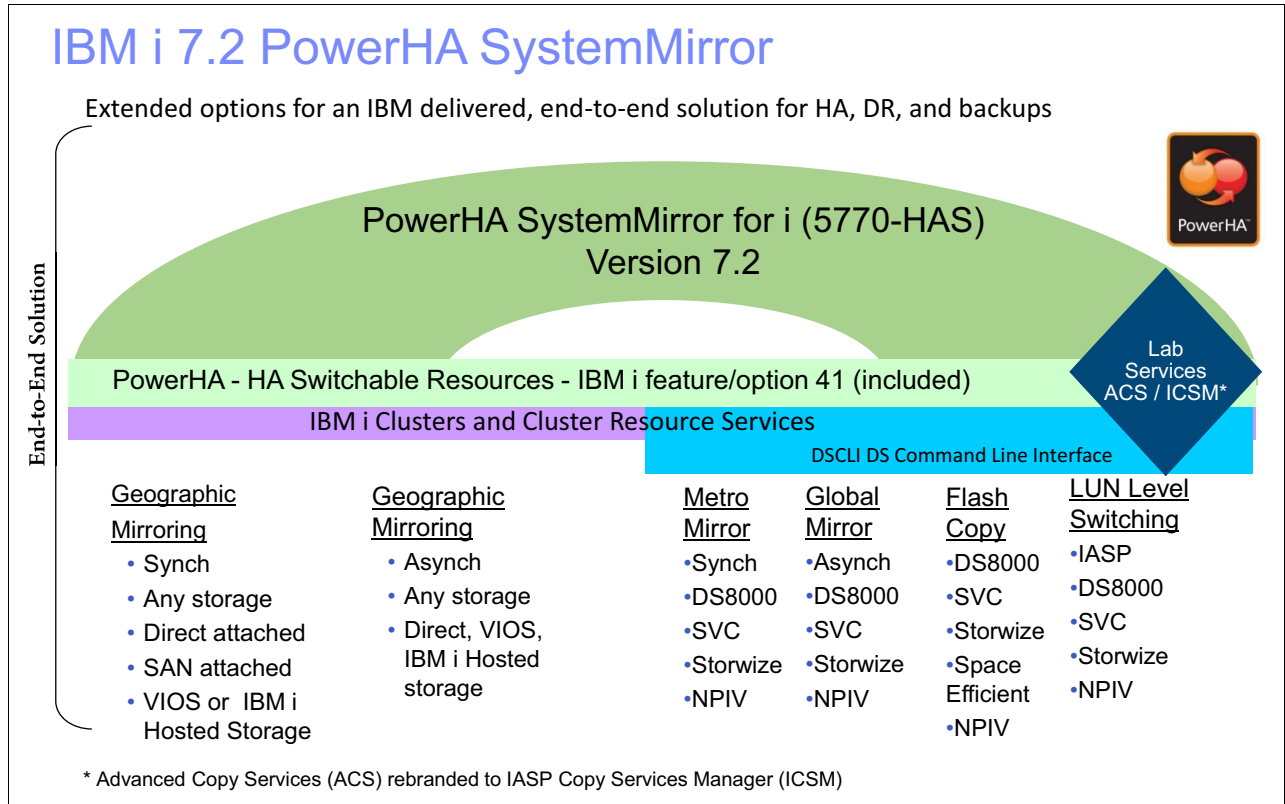
Figure 1-2   IBM PowerHA SystemMirror for i options

A key characteristic of PowerHA is the automation of the solution. Because the data resiliency is managed within the IBM i storage management architecture completely, no operator is involved, just as no operator is involved with RAID 5 or disk mirroring.

Geographic mirroring offers IBM i clients a page-level replication solution for IBM i for implementing HA and DR with any kind of internal or external storage solution that is supported by IBM i.

With IBM System Storage® DS8000 or SAN Volume Controller (SVC)/Storwize V7000 storage servers, clients can use storage-based remote replication functions for HA and DR, logical unit number (LUN)-level switching for local HA, and IBM FlashCopy® for reducing save window outages by enabling the creation of a copy that is attached to a separate partition for offline backup to tape.

PowerHA offers a full range of HA and DR solution choices:

► Geographic mirroring
► LUN-level switching
► Metro Mirror
► Global Mirror
► Combined PowerHA solutions

IASPs: Independent auxiliary storage pools (IASPs) are a fundamental building block for implementing IBM PowerHA SystemMirror for i. For more information about IASPs, see Chapter 2, "Independent auxiliary storage pools (IASPs)" on page 13.

Geographic mirroring

Geographic mirroring is a host-based solution that provides both synchronous and asynchronous options. It is not dependent on any particular storage technology (Figure 1-3).

For more information about geographic mirroring, see 3.4.1, "Host-based replication (geographic mirroring)" on page 53.
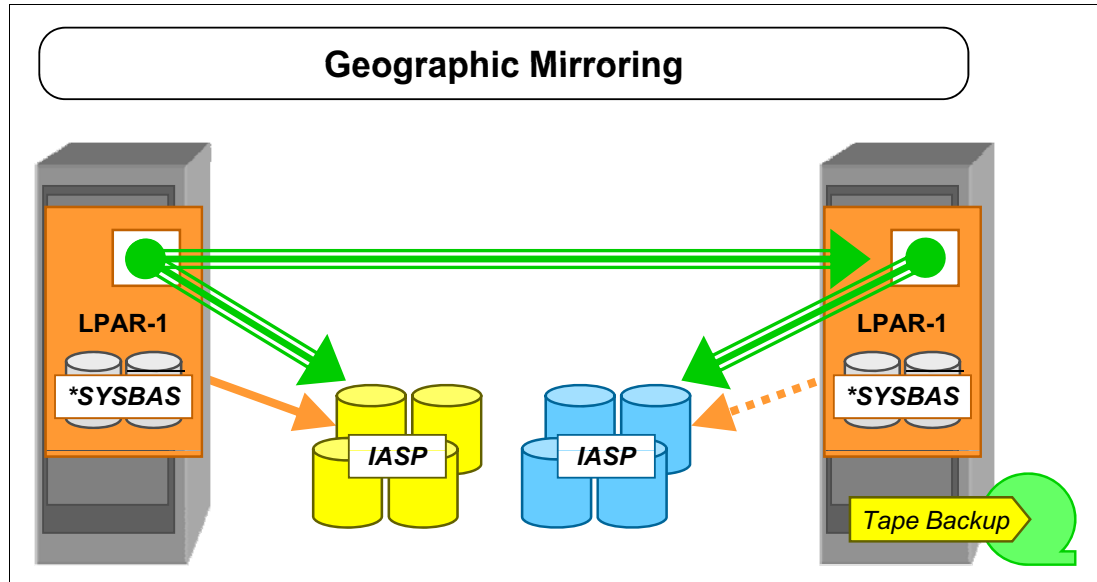
Figure 1-3   PowerHA and geographic mirroring

LUN-level switching

LUN-level switching switches a supported external storage unit connection between two systems to move an IASP from one cluster node to another. See Figure 1-4.

For more information about LUN-level switching, see "LUN-level switching" on page 57.
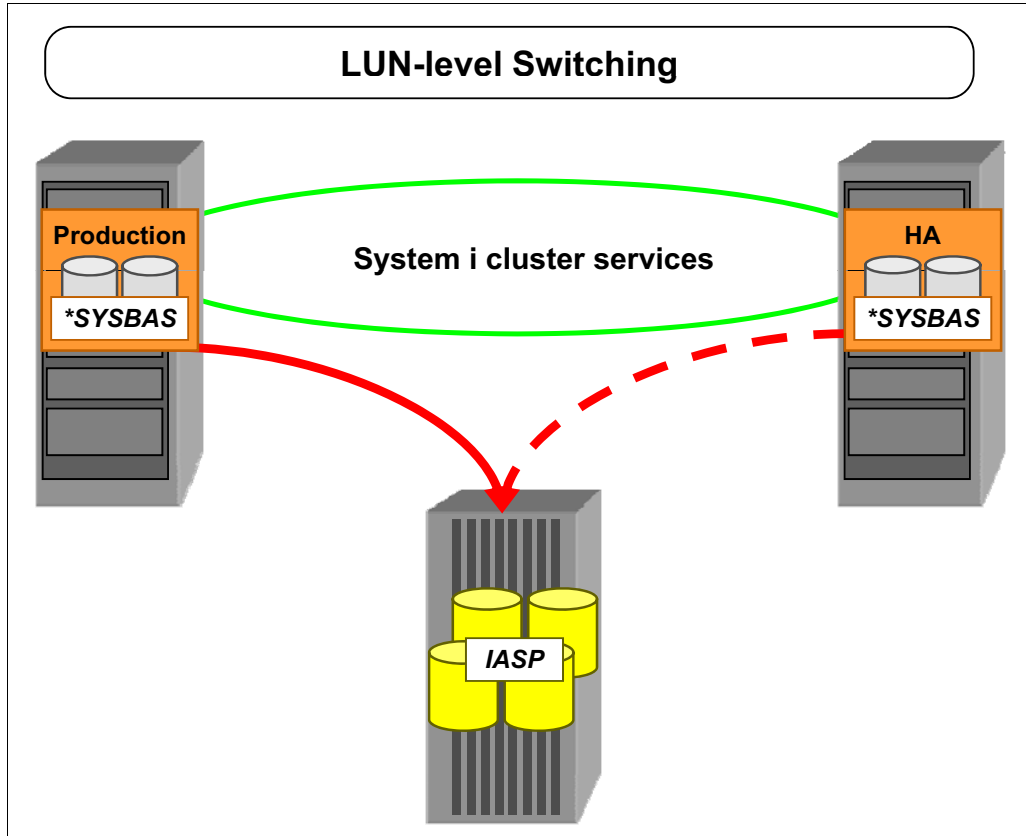
Figure 1-4   PowerHA and LUN-level switching

Metro Mirror

Metro Mirror is a synchronous copy services function on a supported external storage unit that provides hardware replication of an IASP from one cluster node to another (Figure 1-5).

For more information about Metro Mirror, see "Metro Mirror (synchronous)" on page 55.
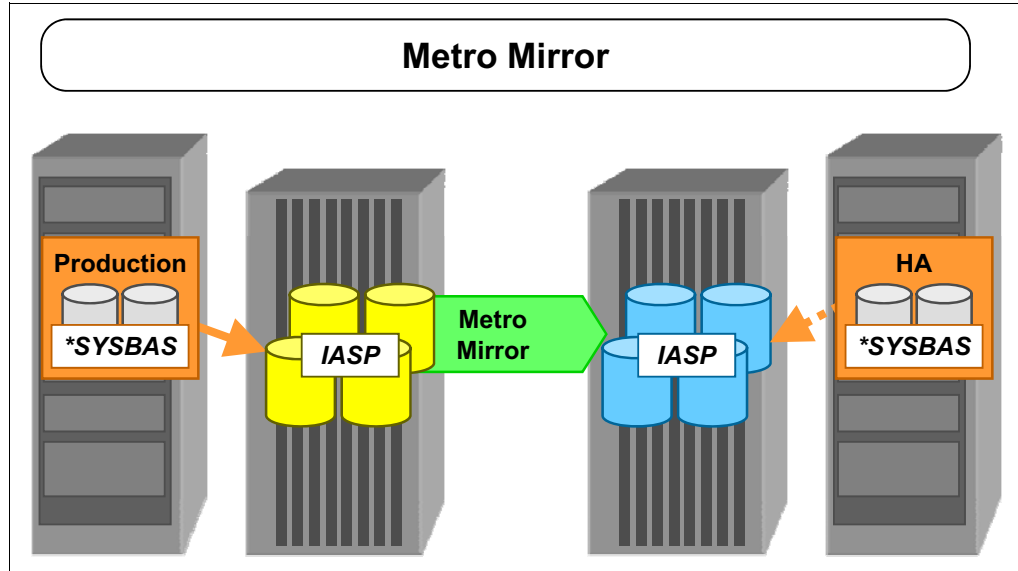
Figure 1-5   PowerHA and Metro Mirror

Global Mirror

Global Mirror is an asynchronous copy services function on a supported external storage unit that provides hardware replication of an IASP from one cluster node to another (Figure 1-6).

For more information about Global Mirror, see "Global Mirror (asynchronous)" on page 56.
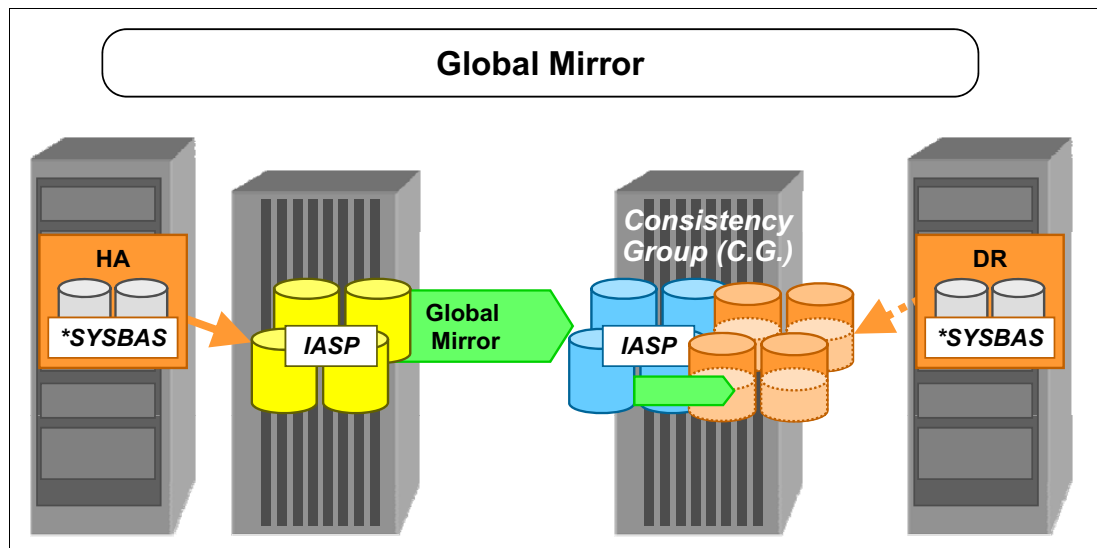
Figure 1-6   PowerHA and Global Mirror

Combined PowerHA solutions

You can combine several of the PowerHA solutions as shown in Figure 1-7.

Combined Power HA solutions

*SYSBAS

*SYSBAS   *SYSBAS

Cluster Admin Domain

Global Mirror

IASP

IASP

FlashCopy

LUN Level Switching
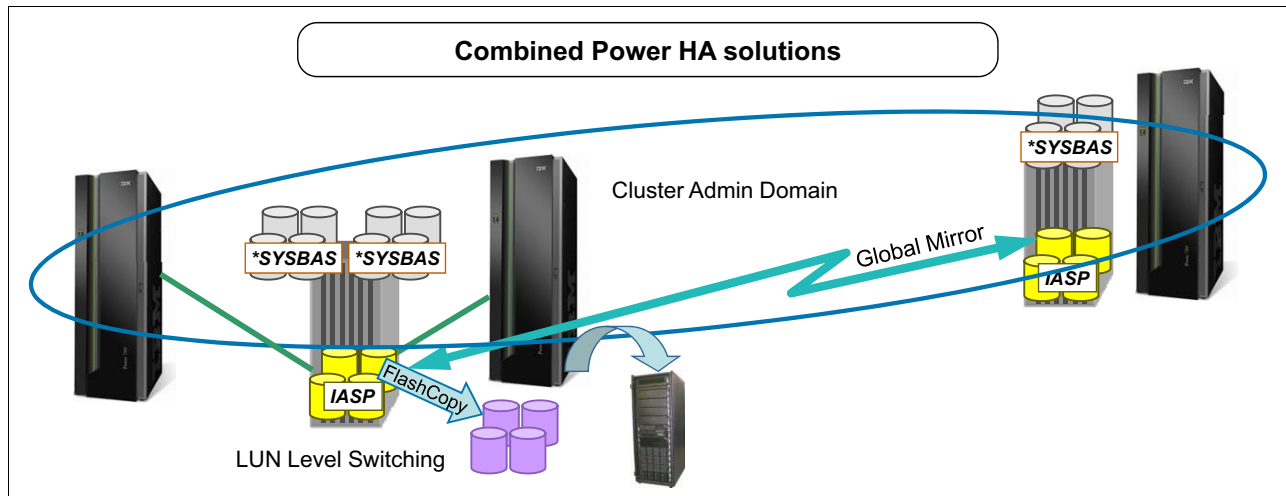
Figure 1-7   Combined PowerHA solution

1.1.2 IBM PowerHA SystemMirror for i availability capabilities

The following enhancements are included with PowerHA SystemMirror for i Version 7.2:

► IBM PowerHA SystemMirror for i Editions

IBM PowerHA SystemMirror for i is now offered in three editions:

– IBM PowerHA SystemMirror for i Express Edition (5770-HAS option 3) for full system IBM HyperSwap®. No IASP is required.

– IBM PowerHA SystemMirror for i Standard Edition (5770-HAS option 2) for local LUN switching, synchronous geographic mirroring options, and FlashCopy only.

– IBM PowerHA SystemMirror for i Enterprise Edition (5770-HAS option 1) for local, remote, or multiple site replication.

For more information about the IBM PowerHA SystemMirror for i Editions, see 3.6, "Licensing for IBM PowerHA SystemMirror for i" on page 63.

► DS8000 HyperSwap support for full system replication

A new Express Edition of PowerHA is available. The initial technology that is supported within this edition is the DS8000 HyperSwap technology. HyperSwap allows for the almost instantaneous switching between DS8000 storage servers that connect to the same IBM i instance.

HyperSwap can be used to eliminate outages because of planned storage server maintenance. The new control language (CL) commands, Display HyperSwap Status (`DSPHYSSTS`) and Change HyperSwap Status (`CHGHYSSTS`), can be used to manage the DS8000 that serves as the primary storage server. If HyperSwap is configured, a HyperSwap switch is also triggered if the production storage server fails or becomes unexpectedly unavailable.

HyperSwap can be used with Live Partition Mobility (LPM) to provide a minimal downtime solution to migrate from one server/storage combination to another. The Add HyperSwap Storage Description (`ADDHYSSTGD`) command can be used to define affinity between an IBM POWER® server and a storage server, so that when an LPM move is performed from one server to another, the correct HyperSwap switch also takes place as part of that process.

For more information about LPM, see Appendix B, "IBM Live Partition Mobility" on page 145.

► Synchronization of object authority and ownership

Object authority and ownership can now be synchronized with the administrative domain.

► Increased administrative domain limit

The number of monitored resource entries (MREs) that is supported within the PowerHA administrative domain is increased from 25,000 to 45,000.

► `DSPASPSTS` improvements

The Display ASP Status (`DSPASPSTS`) command is enhanced to preserve and display up to 64 vary-on or vary-off histories. New options display UDFS and STATFS information, and the user can query and analyze vary-on and vary-off history by using SQL table functions.

For more information about the Display ASP Status (`DSPASPSTS`) command, see the "Display ASP Status (DSPASPSTS)" topic in the IBM i 7.2 Knowledge Center:

https://ibm.biz/Bd4gjf

► Reduced unique identifier (UID)/group identifier (GID) processing time during vary-on

The processing time for the "UID/GID mismatch correction" step in the IASP vary-on process is reduced significantly. Processing time for traditional /QSYS.LIB objects is eliminated, and the processing time for Integrated File System (IFS) objects is minimized.

> **Tip:** It is still considered a preferred practice to synchronize the UID and GID attributes of user profiles that use the administrative domain to eliminate all UID/GID mismatch processing during vary-on.

► Independent ASP assignment for consolidated backups

Now, you can assign an existing IASP to a partition outside of the cluster device domain, which can be advantageous for many reasons.

For environments with multiple production clusters, each of which uses FlashCopy for system backups, this capability can be used to consolidate the number of required backup partitions. Before this enhancement, the FlashCopy was only attached to a partition within the cluster device domain so that each cluster required a dedicated partition for attaching the FlashCopy and completing the save. With this enhancement, one partition can be designated to attach multiple FlashCopies from different cluster environments. The requirements for this function are that only one IASP can be attached at a time and an IPL of the partition is required before a different IASP is attached.

Another use case for this function is for clients who are replicating to a DR site by using external storage replication. It is feasible that only the external storage server is active at the DR site until it is necessary to attach that copy of the IASP to an IBM i instance. At that point, this technology can be used to attach the IASP to that new IBM i instance.

For more information about attaching an IASP to a partition in a single system environment, see the "Attach Independent Disk pool" topic in the IBM i 7.2 Knowledge Center:

https://ibm.biz/Bd4gZE

1.2 PowerHA solution considerations

Today, clients have many choices and need to determine which option is their best HA and DR solution. The criteria for choosing the correct solution needs to be based on business needs, such as the recovery point objective (RPO), recovery time objective (RTO), geographic dispersion requirements, staffing, skills, and day-to-day administrative efforts.

Figure 1-8 shows typical RTOs for various recovery solutions that are associated with the seven tiers of business continuity (BC).
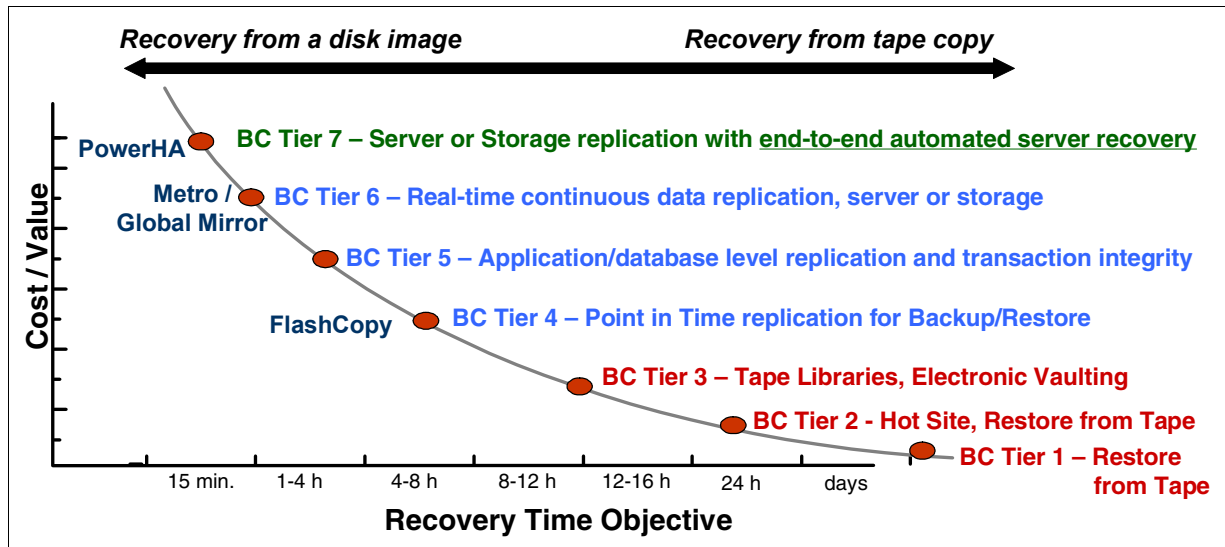
Figure 1-8   Seven tiers of disaster recovery

When you start to think about implementing HA for your IBM i environment, consider how this criteria applies to your situation before you determine the solution that fits your needs best:

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

IBM i data resiliency solutions are either based on logical replication or hardware replication (Figure 1-9). Unlike the previously mentioned PowerHA IASP hardware-based replication solutions, logical replication solutions send journal entries through TPC/IP from the production system to a backup system where the journal entries are then applied to the database.
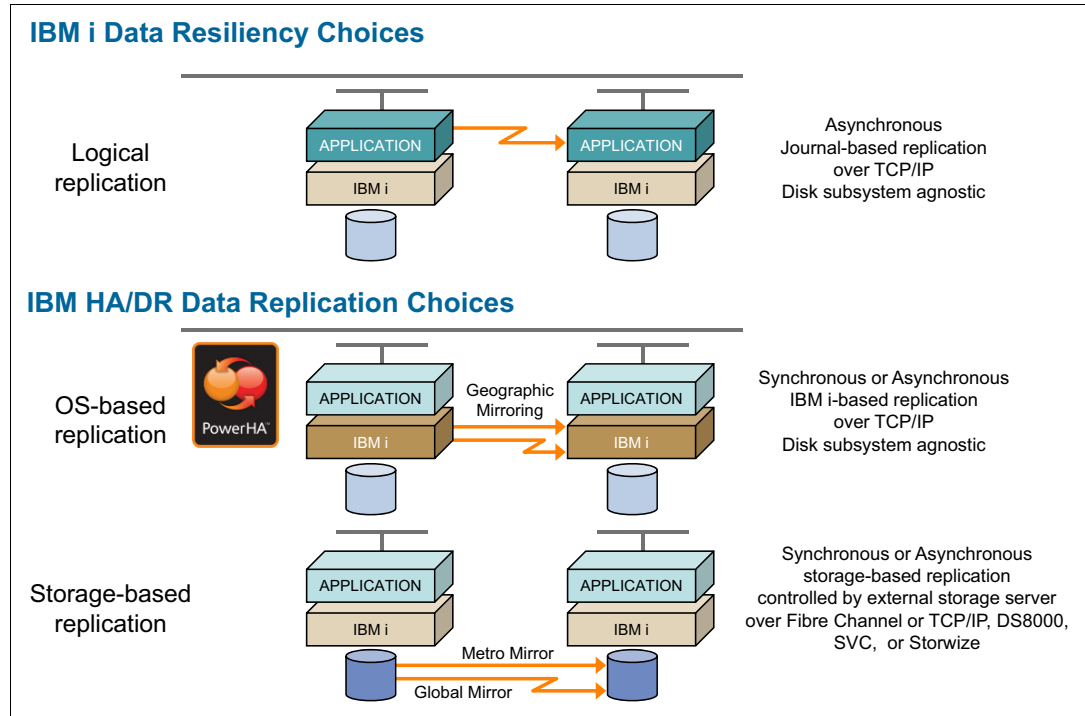
Figure 1-9   IBM i HA/DR data replication options

Solution concepts

This section explains concepts that can help you determine the best solution to address your company's requirements:

► Storage-based synchronous replication

A storage-based *synchronous* replication method is one in which the application state is directly tied to the act of data replication, as it is when it performs a write operation to local disk. You can think of the primary and secondary IASP copies as local disk from the application perspective. This aspect of a synchronous replication approach means that all data that is written to the production IASP is also written to the backup IASP copy and the application waits as though it were a write to local disk.

The two copies cannot be out of sync, and also the distance between the production and backup copies, in addition to the bandwidth of the communication link, has an influence on the application performance.The farther apart the production and backup copies, the longer the synchronous application steps need to wait before they proceed to the next application step.

► Storage-based asynchronous replication

For a longer distance that exceeds the limits of a metro area network, consider the use of an *asynchronous* hardware replication solution to prevent or minimize performance impacts for critical applications. The benefit in comparison to a logical replication approach is that the two copies are identical minus the data in the "pipe". Therefore, the secondary copy is ready to be varied on for use on a secondary node in the cluster.

► Cluster administrative domain

The *cluster administrative domain* is the PowerHA function that ensures that the set of objects that are not in an IASP are synchronized across the nodes in the cluster. Therefore, the application has the resources that it needs to function on each node in the cluster.

Clustering solutions, which are deployed with IASPs and use either storage-based copy services or geographic mirroring replication, require little day-to-day administrative maintenance. They were designed from the beginning for role-swap operations. An HA environment is defined as one in which the primary and secondary nodes of the cluster switch roles on a regular and sustained basis.

> **Rule:** If your business does not conduct regular and sustained role swaps, your business does not have an HA solution deployment.

► Logical replication

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

Synchronous remote journaling provides synchronous replication for the previously mentioned objects, but all other objects are captured by using the audit journal and replicated to the target system. The practical ramification of this type of replication approach is that administrative activities are required to ensure that the production and backup copes of data are the same before a role-swap operation.

Another issue is a significant out-of-sync condition that can occur between the primary and secondary copies of data while the backup server works to apply the data that is sent from the primary trying to catch up. The benefit of the logical replication approach is that the production and backup systems can be virtually any distance from each other and the backup copy can be used for read operations.

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

2

Independent auxiliary storage pools (IASPs)

Independent auxiliary storage pools (IASPs) are fundamental building blocks for implementing IBM PowerHA SystemMirror for i. This chapter provides a brief overview of the concept of IASPs in addition to step-by-step instructions to create an IASP.

The following topics are described in this chapter:

2.1 IASP technology

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

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

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

> **Terminology:** Throughout this book, the terms independent disk pool, independent auxiliary storage pool, independent ASP, and IASP are used interchangeably.

The following types of disk pools are available:

► System disk pool (disk pool 1)

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

► Basic disk pools (disk pool 2 - 32)

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

► Primary disk pool (disk pool in the range 33 to 255, but referred to by name)

The primary disk pool is an independent disk pool that defines a collection of directories and libraries that can have other secondary disk pools that are associated with it. Primary disk pools and any associated secondary pools are not automatically available at IPL and they can be brought online or taken offline independently of system activity on other disk pools. Data in a primary disk pool can be accessed by jobs on the system only hat only read data are still able to work, but any job that is trying to add data to theif the disk pool is brought online and the job has the IASP in its namespace.

► Secondary disk pools (disk pool in the range 33- 255, but referred to by name)

A secondary disk pool is an independent disk pool that defines a collection of directories and libraries and it must be associated with a primary disk pool. It is comparable to a basic disk pool because it can be used to separate specific application objects, such as journal receivers, from your main application objects.

► User-defined file system disk pool (disk pool in the range 33 - 255, but it is referred to by name)

A user-defined file system disk pool is an independent disk pool that contains only user-defined file systems (UDFS).

► Disk pool groups

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

A disk pool group is sometimes known as an *ASP group*.

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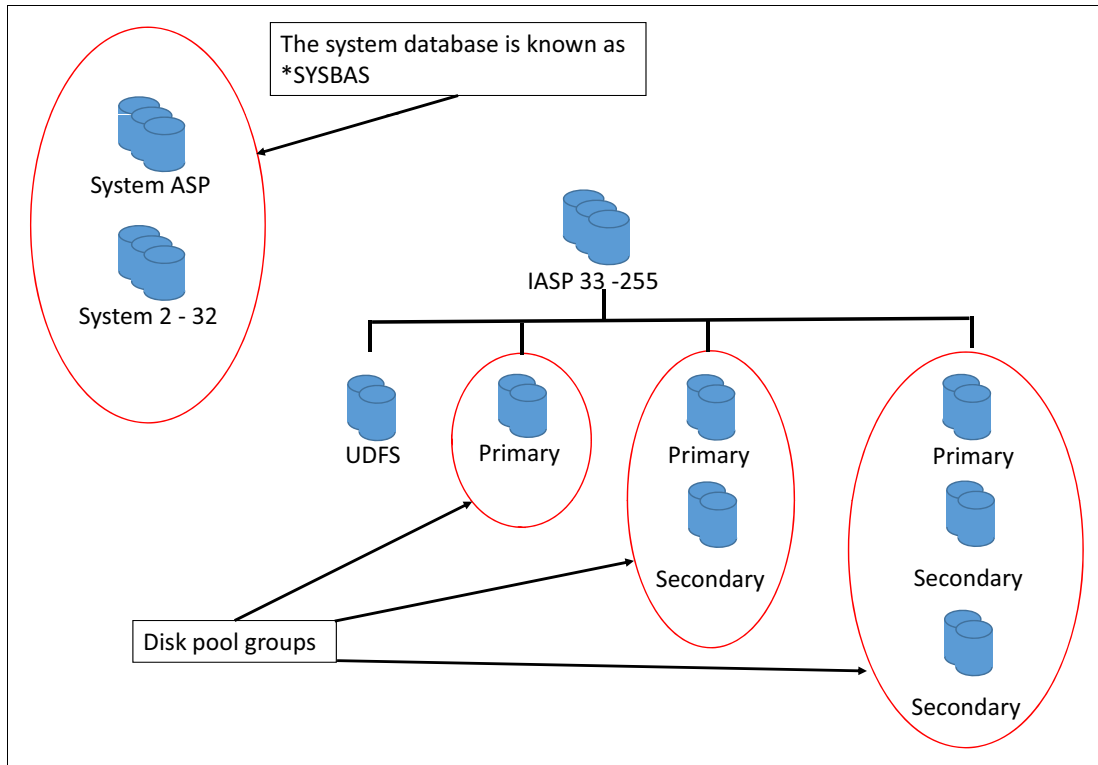
Figure 2-1   IASP hierarchy

Basic ASPs and independent ASPs differ when it comes to data overflow. Basic ASPs overflow and independent ASPs do not. When a basic user ASP fills, it overflows. The excess data spills into the system ASP. IASPs are designed so that they cannot overflow. Otherwise, they cannot be considered independent or switchable. If an IASP is allowed to fill up, the application that is responsible for filling it up simply halts. The responsible job is not automatically canceled. If this job is running from a single-threaded JOBQ, in a single-threaded subsystem, all further processing is stopped until user action is initiated.

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

2.1.1 Namespace

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

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

This set of libraries formed the library *namespace* for the thread and was the only possible component of that namespace. Although no formal term exists for this namespace component, it is now referred to as the **SYSBAS component* of the namespace. It is a required component of every namespace.

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

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

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

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

> **Note:** The namespace is a thread attribute and can be specified when a job is started. When it is referenced as a *job attribute*, it technically means the "thread attribute for the initial thread of a single-threaded job."

2.1.2 Relational database directory

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

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

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

Figure 2-2 shows an example of the RDB directory on a system with an IASP that is configured. One entry is present with a remote location of *LOCAL. This entry is the RDB entry that represents the database in SYSBAS. In addition, an RDB entry is created by the operating system when you vary on an IASP. In this example, this entry is the entry IASP1 with a remote location of LOOPBACK. By default, the relational database name of an IASP is identical to the IASP device name, but you can also specify another name.

Tip: When you migrate an application environment with many accesses through the RDB name, you might want to change the SYSBAS RDB name to a different value and use the "old" SYSBAS RDB name as the database name for the IASP. This way, you do not have to change RDB access to your environment.

```
              Work with Relational Database Directory Entries

  Position to  . . . . . .

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


                            Remote
  Option   Entry             Location                    Text

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



                                                                      Bottom
    F3=Exit    F5=Refresh    F6=Print list    F12=Cancel    F22=Display entire field
```

Figure 2-2   Work with Relational Database Directory Entries

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

► A view into a schema must exist in the same RDB as its referenced tables, views, or functions.

► An index into a schema must exist in the same RDB as its referenced table.

- A trigger or constraint into a schema must exist in the same RDB as its base table.
- The parent table and dependent table in a referential constraint both have to exist in the same RDB.
- A table into a schema has to exist in the same RDB as any referenced distinct types.

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

2.1.3 SQL Connections

In an SQL environment, SQL CONNECT is used to specify the correct database. To achieve the best performance, ensure that the database that is connected to corresponds to your current library namespace. You can use the INLASPGRP parameter in your job description or the Set Auxiliary Storage Pool Group (`SETASPGRP`) command to ensure that the SQL CONNECT function is operating within the same library namespace. If the SQL CONNECT function is not operating within the same library namespace, the application uses IBM Distributed Relational Database Architecture™ (IBM DRDA®) support, which can affect performance.

> **Tip:** When you create your IASP, you might want to consider the use of the original production system RDB name for the IASP name. The IASP then has the same RDB name as the original system, which can minimize the need for coding changes.

2.1.4 Object creation

Although it is possible to create files, tables, and so on in QSYS2, the corresponding library in the independent disk pool prevents the creation of these objects. Most applications that create data in QSYS2 do not realize it and fail when they run in an independent disk pool.

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

Example 2-1   Create view on SYSTABLES

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

2.1.5 System-wide statement cache (SWSC)

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

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

The IASP merged views are automatically re-created during the vary-on process of the IASP.

2.2 Creating an IASP

To create an IASP, you need at least one non-configured available disk on your system. The IASP can be created by using the Configure Device ASP (`CFGDEVASP`) command. The `CFGDEVASP` command performs all of the operations that are required at the operating system level and the System Licensed Internal Code (SLIC) level to create the new IASP.

Table 2-1 lists the parameters for the `CFGDEVASP` command. Consider your choices carefully for these parameters.

Table 2-1   Parameters for the CFGDEVASP command

Parameter	Values	Comments
ASPDEV	*Required value*	Name of the new IASP and also the RDB name for it.
ACTION	*CREATE, *DELETE, *PREPARE	► *CREATE is used to create a new IASP. ► *DELETE is used to delete an existing IASP. ► *PREPARE is used to prepare the system to attach a copy of an IASP from another system when the node is *not* a member of a cluster.
TYPE	*PRIMARY, *SECONDARY, *UDFS	► *PRIMARY is the first IASP in a disk pool group. ► *SECONDARY is an additional IASP for an existing disk pool group. ► *UDFS is an IFS-only IASP.
PRIASPDEV	*Required value if TYPE(*SECONDARY)*	Name of the primary IASP in the disk pool group.
PROTECT	*NO, *YES	Tells the system whether you want RAID only or mirrored disks in the IASP.
ENCRYPT	*NO, *YES	Tells the system whether you want the IASP contents encrypted by the operating system.
UNITS	*SELECT	Either list the logical unit numbers (LUNs) you want to use in the IASP or use *SELECT and the system will present a list of LUNs to select from. (See Figure 2-4 on page 20.)
CONFIRM	*YES, *NO	Indicates whether to use a confirmation panel before you create the IASP.

Figure 2-3 shows the required parameters for creating an IASP with the `CFGDEVASP` command.

```
                        Configure Device ASP (CFGDEVASP)

 Type choices, press Enter.

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


                           Additional Parameters


 Confirm  . . . . . . . . . . . .   *YES           *YES, *NO



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

Figure 2-3   CFGDEVASP command to create an IASP

Specifying *SELECT for the Disk units parameter displays the Select Non-Configured Disk Units panel that is shown in Figure 2-4. This panel provides you with a list of non-configured disks on your system. Select the disks that you want to add to your IASP and press Enter to create the IASP.

```
                    Select Non-Configured Disk Units

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

 Type options, press Enter.
   1=Select

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



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

Figure 2-4   CFGDEVASP command, selecting disks to put into an IASP

A message that shows the progress of the IASP creation is displayed on the bottom of the panel. Your panel is locked while the IASP is created. You can also follow the status of the IASP creation by using the Display ASP Status (`DSPASPSTS`) command.

2.3 Moving applications to an IASP

After you successfully created an IASP as described in 2.2, "Creating an IASP" on page 19, the next step is to look at the considerations for migrating your applications into it. This section describes the object types that can be moved into an IASP and that cannot, how you get access to objects in an IASP, and aspects of your application environment with an IASP.

The following topics are described in this section:

- ▶ 2.3.1, "Object considerations" on page 21
- ▶ 2.3.2, "Accessing objects in an IASP" on page 23
- ▶ 2.3.3, "Considerations for specific environments" on page 24
- ▶ 2.3.4, "Steps for application migration" on page 30

2.3.1 Object considerations

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

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

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

For several of these object types, special considerations apply:

► If network attributes reference the *alert table*, this object must reside in the system ASP.

► If an active subsystem references a *class object*, this class object must reside in the system ASP.

► *Database files* that are either multiple-system database files or that have DataLink fields that are created as link control cannot be located in an independent disk pool. If an active subsystem references the file object, *FILE must exist in the system disk pool, for example, the sign-on display file, unless the subsystem specified the IASP on the *ASPGRP* parameter.

► *Subsystem descriptions* can be placed into an IASP. However, if you want to start a subsystem, its subsystem description must be in the system ASP at that point.

> **Tip:** It is considered a preferred practice to put subsystem descriptions into an additional library that is located in the system ASP.

► The same is true for *job descriptions*. In many cases, job descriptions can be used only when they are located in the system ASP, for example, when they are used by a job scheduler that has no IASP access.

► *Job queues* in an IASP are operationally identical to job queues in the system ASP. Users can manipulate the jobs (submit, hold, release, and so on) or the job queues themselves (clear, hold, release, delete, and so on). However, the internal structures that hold the real content of the job queue still reside in the system ASP. Job entries in a job queue in an IASP do not survive the vary-off process and the vary-on process of an IASP and are lost when you switch over from the production system to the backup system.

► *Journals* and *journal receivers* must be located in the same IASP group as objects that are being journaled. Journal receivers can be moved to a secondary IASP within that IASP group.

► A *library* that is specified by the Create Subsystem Description (`CRTSBSD`) command must exist in the system ASP. In addition, libraries that are referenced in the system values QSYSLIBL or QUSRLIBL cannot be located in an IASP.

► If network attributes reference a *message queue*, that message queue must be located in the system ASP.

► *Programs* that are referenced in subsystem descriptions (for example, in routing entries or prestarted jobs) must be found in the system ASP when that subsystem is activated. The same is true if a program is associated with the attention key.

Table 2-3 shows the object types that cannot be put into an IASP.

Table 2-3   Object types that cannot be put into an IASP

Object types that cannot be put into an IASP			
*AUTHLR	*DDIR	*IMGCLG	*NWSD
*AUTL	*DEVD	*IPXD	*PRDAVL
*CFGL	*DOC	*JOBSCD	*PRDDFN
*CNNL	*DSTMF	*LIND	*PRDLOD
*COSD	*EDTD	*MODD	*RCT
*CRG	*EXITRG	*M36	*SOCKET
*CSPMAP	*FLR	*M36CFG	*SSND
*CSPTBL	*IGCSRT	*NTBD	*S36
*CTLD	*IGCTBL	*NWID	*USRPRF

If you look at the list of object types that are listed in Table 2-3, you will notice that most of them are either existing objects (such as folders or documents), configuration objects (such as device descriptions or line descriptions), or objects that closely relate to system security (such as user profiles or authority lists).

If you want to keep objects in the system ASP in sync between your production and your backup system (such as user IDs and passwords), you can use the administrative domain to help you with this task.

Considerations for moving IFS objects to the IASP: In addition to the object types that are listed in this section, you must not put IBM licensed program libraries into the IASP. Also, do not put any IFS directories that belong to IBM licensed programs into the IASP. This restriction includes the */home* directory and */home/username* directories, where *username* is the name of a user profile.

2.3.2 Accessing objects in an IASP

By default, each job on the system can access objects that are stored in the system ASP only. To access objects in an IASP, that IASP must be varied on. In addition, the IASP must be added to the namespace of the job by using one of the following ways:

► By changing the job description that the job is using
► By using the Set Auxiliary Storage Pool Group (`SETASPGRP`) command

Changing the job description is usually the simplest method. It fixes most situations in a single step. It is also applicable to prestarted jobs.

Important: You cannot change the QDFTJOBD job description to access an IASP. If you do, you are left with an unusable system after an IPL.

If you have user profiles that use QDFTJOBD, use the Create Duplicate Object (`CRTDUPOBJ`) command to create a job description and then modify that job description. You can then update the user profiles that need to access the IASP to use the new job description.

2.3.3 Considerations for specific environments

Now that you know how to generally access data in an IASP, this section describes specific environments and the impact of using an IASP on them. Specifically, the following environments are described:

Considerations for system values

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

► QALWUSRDMN: Allows user domain objects in libraries

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

► QATNPGM: Attention program

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

► QCFGMSGQ: Configuration message queue

You use this system value to specify the default message queue that the system uses when it sends messages for lines, controllers, and devices. The message queue must exist in the system ASP or in a basic user ASP.

► QCTLSBSD: Controlling subsystem

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

► QIGCCDEFNT: Double-byte code font

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

- ► QINACTMSGQ: Inactive job message queue

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

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

- ► QPRBFTR: Problem log filter

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

- ► QPWDVLDPGM: Password validation program

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

- ► QRMTSIGN: Remote sign-on control

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

- ► QSRTSEQ: Sort sequence

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

- ► QSTRUPPGM: Startup program

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

- ► QSYSLIBL: System part of the library list

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

- ► QUPSMSGQ: Uninterruptible power supply (UPS) message queue

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

- ► QUSRLIBL: User part of the library list

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

Considerations for network attributes

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

For more information about these commands, see the "CL Command Finder" topic in the IBM i 7.2 Knowledge Center. To access this function, type `CL Command Finder` in the Search field:

http://www.ibm.com/support/knowledgecenter/ssw_ibm_i_72/clfinder/finder.htm

Consider the keywords and parameters for the following system network attributes:

► Alert Filters (ALRFTR)

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

► Message Queue (MSGQ)

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

► Distributed Data Management Access (DDMACC)

This parameter specifies how the system processes distributed data management (DDM) and DRDA requests from remote systems for access to the data resources of the system. The DDM and DRDA connections refer to Advanced Program-to-Program Communication (APPC) conversations or active TCP/IP or OptiConnect connections. Changes to this parameter are immediate and apply to DRDA, DDM, or IBM DB2® Multisystem applications. However, jobs that are currently running on the system do not use the new value.

The DDMACC value is accessed only when a job is first started. You must specify a special value or program name that dictates how the requests are to be handled. If a program name is specified, the program must exist in the system ASP or in a basic user ASP.

► PC Support Access (PCSACC)

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

Considerations for ODBC/JDBC

Open Database Connectivity (ODBC) and Java Database Connectivity (JDBC) work with prestarted jobs that run under user profile QSYS. After a request comes into the system, that request gets connected to one of the prestarted jobs. That request has to use a user profile and password to authenticate. Then, the prestarted job is changed to run with this user profile and the environment that is defined in the user profile settings. If the user profile uses a job description that is associated with an IASP, the ODBC or JDBC connection can access objects that are located in the IASP.

If for any reason the use of a specific job description is not possible with certain ODBC or JDBC connections, both also provide parameters in the connection setup to explicitly include an IASP in their namespace for ODBC. Example 2-2 shows an example of giving the ODBC connection access to IASP1.

Example 2-2   Setting access to IASP1 with ODBC

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

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

Example 2-3   Setting access to IASP1 in JDBC

```
DriverManager.registerDriver(new AS400JDBCDriver());
AS400JDBCDataSource ds = new AS400JDBCDataSource("SYS1");
   ds.setUser("xxxxxxxxxx");
   ds.setPassword("yyyyyyyyyy");
   ds.setNaming("sql");
   ds.setDatabaseName("IASP1");
```

Considerations for FTP

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

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

```
quote rcmd SETASPGRP IASP1
```

Considerations for the integrated file system (IFS)

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

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

► The MOUNT operation will be successful if the existing directory is empty.

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

Tip: Check that no IFS directories have the same name as the primary IASP that you want to create.

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

You create a symbolic link with the following **ADDLNK** command, where *newobjpath* is the path to the directory or stream file after it is moved to the IFS and *originalobjectpath* is the path to the directory or stream file before it was moved.

```
ADDLNK OBJ(newobjpath) NEWLNK(originalobjectpath) LINKTYPE(*SYMBOLIC)
```

Note: Before you can use the **ADDLNK** command, you need to move the directory or stream file to its new location. Do not use the LINKTYPE(*HARD) parameter because it does not work across database boundaries.

Considerations for DRDA

Certain DRDA objects that relate to DRDA cannot be contained in an IASP. DDM user exit programs must reside in libraries in the system database, just as any other DRDA programs.

The process of varying on an IASP causes the RDB directory to be unavailable for a short period, which can delay or time out attempts by a DRDA application requester or application server to use the directory.

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

Considerations for database access by using SQL

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

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

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

Query Management Query and Query Management Procedures

You can resolve the SQL objects (tables, functions, views, and types) that are referenced in a Query Management Query (*QMQRY) object. You use the RDB that is specified on the RDB parameter or the RDB that is specified on the `CONNECT/SET CONNECTION` commands. This RDB might be an IASP. The query management objects that are referenced must be in the current RDB (namespace).

When output from the Start Query (`STRQQRY`) command is directed to an output file, Query Management ensures that the output file is created on the RDB (namespace) that was current at the time that `SRQMQRY` is executed.

Considerations for DB2 Web Query

DB2 Web Query references objects in the current RDB (namespace) only. A *QRYDFN object that was created in *SYSBAS might reference files in an IASP and vice versa. If a *QRYDFN object that was created to reference objects in an IASP runs when a different IASP is set as the current RDB (namespace), the *QRYDFN runs successfully if the new IASP contains objects with the same name and the file formats are compatible.

Considerations for spool files

Output queues can be put into an IASP so that spool files are available on the backup system after a switchover or failover situation. This capability works for output queues that are not attached to a physical printer device only because those output queues must be placed in QUSRSYS.

In addition, for output queues in an IASP, the connection between a job and its spool file ends when the job itself ends. Therefore, users can no longer access their old spool files by using the Work with Spooled Files (`WRKSPLF`) command, but instead they must use the Work with Output Queues (`WRKOUTQ`) command. This behavior is the same behavior that you get with output queues in SYSBAS when the system value QSPLFACN is set to *DETACH.

To print from an output queue in the IASP, you must ensure that the job that issues the `STRPRTWTR` command has the IASP in its namespace.

2.3.4 Steps for application migration

With the created IASP and an understanding of the behavior of an IASP, you can start to move your applications to an IASP environment.

Follow these general steps to migrate your applications to an IASP environment:

1. Ensure that your IASP is varied on and that its status is AVAILABLE.

2. Restore your application libraries into the IASP. *You cannot have identical library names in the system ASP and in the IASP.* If your application was installed in the system ASP of the system that you are using to do the migration, you must delete the original libraries before you can restore them into the IASP. Objects that are not supported in an IASP must be copied to a different library in the system ASP.

    > **Note:** Check the job log of the job that performs the restore to ensure that you capture any objects that were not restored. You need this information to complete the restore.

    When you restore objects, private authorities need to be restored with the `RSTAUT` command, also. The parameters that are shown will usually work, where *newasp* is the name of the IASP that you are restoring into.

    ```
    RSTAUT USRPRF(*ALL) SAVASPDEV(*ALLAVL) RSTASPDEV(newasp)
    ```

3. Copy IFS data that is part of your application into the new directory inside the IASP. Delete the original IFS data and create symbolic links that redirect access from the old directories to the new directories in the IASP.

4. If application objects are stored in QGPL or QUSRSYS, move them to a library inside the IASP. QGPL and QUSRSYS cannot be moved to the IASP.

5. Change your application job description to connect to the IASP by using the INLASPGRP parameter. If you use the QDFTJOBD job description, you must first copy it and change the copy.

    > **Important:** *Never* change QDFTJOBD to access an IASP. Because QDFTJOBD is used for the startup program during an IPL, the IPL will fail because the IASP is not yet available. Ensure that you set your library environment correctly in the job description because libraries in an IASP cannot be referenced by system values QSYSLIBL or QUSRLIBL.

6. If you created any new job descriptions, change the user profiles to use them. Ensure that you do not change your IBM i administration user profiles to use a job description with an included IASP. If the IASP is not varied on for any reason, you cannot sign on to the system if your profile points to a job description with an IASP.

7. Test your application.

8. Change your application environment to work with the IASP. Think of save and restore procedures or changes in the startup program to vary on the IASP.

9. Decide on procedures to synchronize objects in the system ASP from the primary system to the backup system.

10. Switch your application over to the backup system and test it there.

2.4 After you move your application to the IASP

Moving your application to the IASP was the first step. Now, you need to change your associated work management and processes to maximize the benefit to your business.

2.4.1 Starting and stopping your applications

Traditionally, in the non-IASP environment you use the QSTRUPPGM system value to run a program of your choice when you IPL the partition. This program starts any required subsystems and services before it starts the application environment.

This approach works well for a single system environment. However, when multiple partitions are involved, you immediately have the dilemma of whether or not you are the production system.

> **Note:** You will not normally expect to IPL the production partition because a role swap is normally a faster operation. A role swap means less disruption to your business.

In the IASP environment, you use a different structure. The program that is registered to the QSTRUPPGM system value starts the system subsystems and services, for example, QBATCH, QSPL, and TCP/IP, that are always required whenever the partition is running.

Then, you need to identify whether it is the current production node by using the `RTVASPSSN` command and retrieving the node role from the returned data. If the node that you IPL is the production node, it varies on the IASP.

Exit programs

Four exit points, which are documented in the "Vary Configuration exit programs" topic of the IBM i 7.2 Knowledge Center, are available for the `VRYCFG` command:

http://www.ibm.com/support/knowledgecenter/ssw_ibm_i_72/apis/XDCVRX400.htm?lang=en

The four exit points cover both the pre-vary on and post-vary on, and the pre-vary off and post-vary off. By using these exit points, you can automate the operations around your IASP.

A typical use for these exit programs is to use the post-vary-on exit to start the application or perform a backup (if a FlashCopy partition is used). These exit points can also be used to manage the takeover IP address. By using the exit points here and not on the cluster resource group (CRG), you have more flexibility to control when you start the IP address.

You can also consider the use of the pre-vary-off exit point to terminate any applications that must be shut down in a controlled manner before the IASP is varied off.

2.5 Journaling

It is difficult to achieve both the recovery time objective (RTO) and recovery point objective (RPO) with a hardware replication solution, such as IBM PowerHA SystemMirror for i, without seriously considering a journaling setup. This solution relies on disk replication that copies, from source disks to target disks, only data that exists on disks.

Everything that is still in main memory and that is not yet written to disk cannot be replicated. Journaling is not a matter for *planned* switches, where you vary off the IASP, which flushes memory content to disks. Journaling is a matter for *unplanned* switches, which can occur at any time, for any reason, and for which you must apply techniques to prevent losing data and to reduce recovery time.

For example, assume that an application uses a data area to track the next customer number of a database file to record registration information for the new customer to be assigned and an IFS file to capture a scanned image of the customer's signature. Three separate objects are used to store data. At the end of the transaction that enrolls this new customer, everything is consistent for those three objects in *main memory*. But you have no assurance that the related *disk image* received the updates. Most likely, they were not received yet. The disk update order might differ from the memory updates order, depending on the main memory-flushing algorithm.

This lack of consistency on disks drives clearly affects the ability to reach your planned RPO. Certain objects, such as database files, might also need time-consuming internal consistency checks and recovery at IASP vary-on time. Detecting and correcting them affects planned recovery time, too.

The use of journaling is the best way to achieve a consistent state and to reduce recovery times as much as possible. Journal protection must be enabled for all critical RPOs, data areas, database files, data queues, and IFS files. The use of journaling does not require an update to the application code. Starting, ending, and managing journal operations are performed outside of the application. The best protection for data consistency is achieved by changing the application to take care of transaction consistency with commitment control. The application decides when, from a functional standpoint, any transaction is finished and then it performs a commit. If an outage occurs during commitment control cycles, the system ensures, at IASP vary-on time, that data is consistent regarding the transactions. It undoes any update that is not committed.

Journaling was never mandatory with the IBM i operating system. Therefore, numerous users never tried to implement it, even when they used hardware replication solutions.

> **Note:** It is the user's responsibility to manage journaling. IBM PowerHA SystemMirror for i does not help you manage journals. But, you can use system-managed journals.

In the past, two main obstacles existed to using journaling:

► Performance impact
► Management effort

The following sections address these concerns:

2.5.1 Journal performance impact

A performance impact still exists and always will because the way that journaling works implies the use of more hardware resources, such as increased disk write operations. The objective of journaling is to write to disk, in the journal receiver object, on a synchronous manner, all updates to journaled objects before these updates are effective on disks.

However, in IBM i release after release, major improvements exist:

► When you use internal drives for IASP, write cache performance is a key factor for journal performance.

For more information about this topic, see the IBM Redbooks publication, *Journaling - Configuring for your Fair Share of Write Cache*, TIPS0653.

► If you decide to use a private ASP for receivers, do not skimp on the quantity of available disk arms for use by the journal.

For more information about this topic, see the IBM Redbooks publication, *Journaling - User ASPs Versus the System ASP*, TIPS0602.

> **Note:** Do not forget that a private (user) ASP can be created inside an IASP as a secondary ASP type and coexist with a primary type ASP.

► If you do not intend to use journal receiver entries for application purposes, or if you want to use journal receiver entries for application purposes and are ready for application programming interfaces (API) programming, consider minimizing receiver entries data. The journal parameters, *receiver size options*, *minimize entry-specific data*, and *fixed-length data*, play a role in this optimization step. For example, if you are using those journals only for recovery purposes, do you really need the entire database file record content in each record entry?

For more information about this topic, see the IBM Redbooks publication, *Journaling: How to View and More Easily Audit Minimized Journal Entries on the IBM System i Platform*, TIPS0626.

► Depending on your application needs, you cannot start journaling for work files that are used only to help data processing and do not contain any critical data.

► Ensure that your journal employs the most recent defaults. Many journals that were created before the release of IBM i Version 5.4 might still be locked into old settings, and these old settings might affect performance. One of the easiest ways to ensure that you remain in lock-step with the best journal settings is to let the system apply its own latest defaults. You can help ensure that these settings are employed by specifying the RCVSIZOPT(*SYSDFT) parameter on the Change Journal (`CHGJRN`) command.

► Consider installing and enabling the journal caching feature if journal performance (especially during batch jobs) is a major concern in your shop. Ensure that you understand the possible impacts on missing receiver writes in an outage.

> **Important:** Use of this function can result in lost data, so remember this possibility if your RPO is zero data loss.

For more information about this topic, see the IBM Redbooks publication, *Journal Caching: Understanding the Risk of Data Loss*, TIPS0627.

- ► The more actively modified objects (such as open files) that are associated with a journal, the higher you might want to set your journal recovery count. Failure to set your journal recovery count high enough slows your runtime performance and increases your housekeeping overhead on your production system without adding much benefit to your RTO. Increasing this value might make good sense, but do not get carried away.

  For more information about this topic, see the IBM Redbooks publication, *The Journal Recovery Count: Making It Count on IBM i5/OS*, TIPS0625.

- ► 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. The use of both a force-write-ratio and journal protection for the same file yields no extra recovery or survivability benefit and only slows down your application. The journal protection approach is the more efficient choice.

- ► If your applications tend to produce transactions that consist of fewer than 10 database changes, you might want to consider the use of *soft* commitment control. However, understand that the last committed transaction cannot be written to the journal.

  For more information about this topic, see the IBM Redbooks publication, *Soft Commit: Worth a Try on IBM i5/OS V5R4*, TIPS0623.

- ► Both before and after images of each updated record can be written to the journal receiver, depending on the way that the database file journaling is started. For recovery, you need the after image only. When the IASP is varied on, storage database recovery applies after images, if needed. Consider journaling only if after images are an option.

  Note: Commitment control automatically activates before images if they are needed for a database file that is involved in a transaction and for which only after images are journaled.

More in-depth treatments of a number of these preferred practices are in the IBM Redbooks publication *Striving for Optimal Journal Performance on DB2 Universal Database for iSeries*, SG24-6286.

2.5.2 Journal management effort

It is now much easier than in the past to manage the journal on a library-wide basis. New commands exist and certain restrictions that existed in the previous releases are gone.

The following topics are covered in this section:

- ► "Automatic journaling for changed objects within a library" on page 35
- ► "Start journaling for all or generic files or objects for one or more libraries" on page 36
- ► "End journaling locking restriction lifted" on page 37
- ► "Receiver name wrap" on page 37
- ► "Pseudo journal tool" on page 38

Automatic journaling for changed objects within a library

The Start Journal Library (**STRJRNLIB**) command allows the system to automatically start journaling for any object in this library that receives the operation create, move, restore, or all of these operations (Figure 2-5).

```
                      Start Journal Library (STRJRNLIB)

 Type choices, press Enter.

 Library  . . . . . . . . . . . .                    Name, generic*
             + for more values
 Journal  . . . . . . . . . . .                      Name
   Library  . . . . . . . . . .        *LIBL         Name, *LIBL, *CURLIB
 Inherit rules:
   Object type  . . . . . . . .        *ALL          *ALL, *FILE, *DTAARA, *DTAQ
   Operation  . . . . . . . . .        *ALLOPR       *ALLOPR, *CREATE, *MOVE...
   Rule action  . . . . . . . .        *INCLUDE      *INCLUDE, *OMIT
   Images . . . . . . . . . . .        *OBJDFT       *OBJDFT, *AFTER, *BOTH
   Omit journal entry . . . . .        *OBJDFT       *OBJDFT, *NONE, *OPNCLO
   Remote journal filter  . . . .      *OBJDFT       *OBJDFT, *NO, *YES
   Name filter  . . . . . . . .        *ALL          Name, generic*, *ALL
             + for more values
 Logging level  . . . . . . . .        *ERRORS       *ERRORS, *ALL



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

Figure 2-5   Start Journal Library (STRJRNLIB) command

By using the **STRJRNLIB** command, all objects that are created in, moved to, and restored in the library will get an *automatic journaling start* to the same journal, with the same settings. You no longer need to take care of them. However, you must take care of those objects that you do not want to journal. End journaling must be run for them.

Note: The **STRJRNLIB** command does not take care of existing objects in the library. It applies to changes to the library only.

A panel exists in the Display Library Description (`DSPLIBD`) command to display the current settings for the automatic journaling of a library. Figure 2-6 shows an example with a journaled library that is called MYLIB.

```
                       Display Library Description

 Library  . . . . . . . :    MYLIB          Type . . . . . . . . . :    TEST

 Journaling information:
   Currently journaled  . . . . . . . . :    YES
   Current or last journal  . . . . . . :    MYJOURNAL
     Library  . . . . . . . . . . . . . :      MYLIB
   Journal images . . . . . . . . . . . :    *AFTER
   Omit journal entry . . . . . . . . . :    *NONE
   New objects inherit journaling . . . :    *YES
   Inherit rules overridden . . . . . . :    NO
   Journaling last started date/time  . :    09/20/11  17:00:25
   Starting receiver for apply  . . . . :
     Library  . . . . . . . . . . . . . :
     Library ASP device . . . . . . . . :
     Library ASP group  . . . . . . . . :


                                                                      Bottom
   F3=Exit    F10=Display inherit rules    F12=Cancel    Enter=Continue
```

Figure 2-6   Journaling information for MYLIB library

Considerations apply with another method to automatically start journaling with the QDFTJRN data area. For more information about these considerations and about the `STRJRNLIB` command, see the IBM Redbooks publication, *Journaling at Object Creation with i5/OS V6R1M0*, TIPS0662.

Start journaling for all or generic files or objects for one or more libraries

The Start Journal Physical File (`STRJRNPF`), Start Journal Access Path (`STRJRNAP`), and Start Journal Object (`STRJRNOBJ`) CL commands are enhanced to allow the start of journaling for all or generic files or objects at one time. In older releases, you wrote a program to build a list of objects, read the list, and started journaling for each entry of this list.

For example, all database files that are included in a library, or a set of libraries, can be journaled with one command as shown in Figure 2-7.

```
                   Start Journal Physical File (STRJRNPF)

 Type choices, press Enter.

 Physical file to be journaled  .                    Name, generic*, *ALL
   Library  . . . . . . . . . .       *LIBL          Name, *LIBL, *CURLIB
                + for more values
                                      *LIBL
 Journal  . . . . . . . . . . .                      Name
   Library  . . . . . . . . . .       *LIBL          Name, *LIBL, *CURLIB
 Record images  . . . . . . . .       *AFTER         *AFTER, *BOTH
 Journal entries to be omitted  .     *NONE          *NONE, *OPNCLO
 Logging level  . . . . . . . .       *ERRORS        *ERRORS, *ALL




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

Figure 2-7   Start Journal Physical File (STRJRNPF) command

Therefore, for an existing library, by using four CL commands, it is possible to start journaling for all existing applicable objects and for all changes that apply in the future. Example 2-4 shows an example with the MYLIB library.

Example 2-4   Automatic journaling starting for MYLIB library

```
STRJRNPF FILE(MYLIB/*ALL) JRN(MYLIB/MYJOURNAL) IMAGES(*AFTER) OMTJRNE(*OPNCLO)
STRJRNOBJ OBJ(MYLIB/*ALL) OBJTYPE(*DTAARA) JRN(MYLIB/MYJOURNAL) IMAGES(*AFTER)
STRJRNOBJ OBJ(MYLIB/*ALL) OBJTYPE(*DTAQ) JRN(MYLIB/MYJOURNAL) IMAGES(*AFTER)
STRJRNLIB LIB(MYLIB) JRN(MYLIB/MYJOURNAL) INHRULES((*ALL *ALLOPR *INCLUDE *AFTER *OPNCLO))
```

End journaling locking restriction lifted

You can end the journaling of a physical file or an object, even if the object was opened by an application. You do not need to shut down the application to perform this action. It remains impossible to end journaling for files in the middle of a commitment control cycle. The End Journal Physical File (**ENDJRNPF**) and End Journal Object (**ENDJRNOBJ**) CL commands also support generic names and all objects.

Receiver name wrap

In the past, for a user-created journal, when the receiver name got its maximum sequence (for example, JRNRCV9999), the system was unable to create a new one, and no more entries were able to be written through the journal. All application programs waited for an answer to an inquiry message in the QSYSOPR message queue that requested an action by the user. The Change Journal (**CHGJRN**) command processing is enhanced to support receiver name wrapping.

Therefore, if your currently attached receiver name is at the maximum value, the system now generates a new receiver with a receiver number value that is wrapped (for example, JRNRCV0000). And if a journal receiver with the name JRNRCV0000 lingers (often due to poor housekeeping practices), the system simply increments the numerical suffix to skip over potential lingering receivers from earlier and continues.

Pseudo journal tool

The pseudo journal tool is a stand-alone set of software. It helps to estimate the quantity of journal traffic that ensures that journal protection is enabled for a set of designated physical files. It is useful to answer these questions:

► How many journals must I configure?

► Will the total quantity of journal/disk traffic justify the use of more than one journal?

► Does it make sense to configure the journal caching feature on the production system and, if so, how much benefit will be gained for the particular applications?

The benefit of the pseudo journal tool is that it not only helps answer these questions, it answers them without affecting your system while it performs the analysis. Better yet, it produces a customized analysis of projected additional disk traffic, which is tuned to your particular application and its database reference pattern.

For more information about the pseudo journal tool, the software to download, and a tutorial, see the IBM DB2 for i: Code example website:

https://ibm.biz/Bd4hri

2.6 Reducing IASP vary-on times

This section describes tips that you can use to reduce IASP vary-on times.

The following topics are covered in this section:

► 2.6.1, "Keeping as few database files in SYSBAS as possible" on page 38
► 2.6.2, "Synchronizing UIDs/GIDs" on page 38
► 2.6.3, "Access path rebuild" on page 39
► 2.6.4, "System-managed access-path protection" on page 39

2.6.1 Keeping as few database files in SYSBAS as possible

The system disk pool and basic user disk pools (SYSBAS) need to contain primarily operating system objects, licensed program libraries, and few-to-no user libraries. This structure yields the best possible protection and performance. Application data is isolated from unrelated faults and it can also be processed independently of other system activity. IASP vary-on and switchover times are optimized with this structure. Expect longer IASP vary-on and switchover times if many database objects reside in SYSBAS because additional processing is required to merge database cross-reference information into the disk pool group cross-reference table.

2.6.2 Synchronizing UIDs/GIDs

In a high-availability environment, a user profile is considered 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).

The UID and GID are used in relation to object ownership. To reduce the amount of internal processing that occurs during a switchover, where the IASP is made unavailable on one system and then made available on a different system, the UID and GID values must be synchronized across the recovery domain of the device cluster resource group. If not, each object that is owned by a user profile with a non-matching UID needs to be accessed, and the UID needs to be changed as part of the vary-on process of the IASP after each switchover or failover. Synchronization of user profiles, including UID and GID, can be accomplished by using the cluster administrative domain support.

2.6.3 Access path rebuild

To reduce the amount of time that an IASP vary-on waits for access path rebuilds, consider performing the rebuilds in the background (after the vary-on completes) by using the RECOVER attribute of a logical file. If RECOVER(*IPL) is specified, the vary-on waits for the rebuild to complete when one is necessary. If RECOVER(*AFTIPL) is specified, the vary-on does not wait and the access path is built in the background after the vary-on process is complete.

Access paths are not available while they are rebuilt. Therefore, if an application requires that specific logical files are valid shortly after a vary-on completes, the user needs to consider specifying RECOVER(*IPL) to avoid a situation in which jobs try to use these files before they are valid.

Another option to consider is the journaling of access paths so that they are recovered from the journal during a vary-on and they do not need to be rebuilt. Access paths are journaled implicitly by system-managed access-path protection (SMAPP). To ensure that specific, critical access paths are journaled, they must be explicitly journaled.

When the system rebuilds an access path after the IPL, you can use the Edit Rebuild of Access Paths (**EDTRBDAP**) command to modify rebuild sequences to select those access paths that need to be rebuilt now. However, by default, this command applies only to SYSBASE.

To allow the **EDTRBDAP** command to work with IASP access paths, follow these steps:

1. Enter the following command, where *YY* is the IASP number:

   ```
   CRTDTAARA DTAARA(QTEMP/QDBIASPEDT) TYPE(*DEC) LEN(4) VALUE(YY)
   ```

2. Run the **EDTRBDAP** command when the IASP becomes ACTIVE or AVAILABLE. It might be necessary to invoke the command multiple times while the IASP is varied on and while it is ACTIVE. A CPF325C message is sent until the command is allowed to be used.

3. Enter the following command to delete the data area:

   ```
   DLTDTAARA DTAARA(QTEMP/QDBIASPEDT)
   ```

2.6.4 System-managed access-path protection

Analyze your system-managed access-path protection (SMAPP) setting and ensure that you are not locked into an outdated setting that was inherited from the past. A SMAPP setting that is too high can significantly increase your recovery duration and cause you to miss your RTO.

Tip: SMAPP is a form of behind-the-scenes journaling. If you see a SMAPP setting longer than, for example, 50 minutes, seriously consider lowering the value. An original default setting nearly a decade ago was 150 minutes, and many companies did not revisit this setting as hardware speeds improved. These companies might still operate with outdated settings. If you continue to use outdated settings, the vary-on duration for an IASP will exceed your RTO.

SMAPP applies to the IASP vary-on (or IPL for the SYSBASE) step, which is responsible for rebuilding access paths if they are damaged after an outage. For access paths (or indexes in the SQL world) that depend on large physical files (or tables in the SQL world), it might take considerable time (several hours). To avoid rebuilding the access paths, SMAPP uses existing journals for access path update transparent recording, as though the access paths are journaled. They are recorded so that the IASP vary-on step can use them to update access paths instead of rebuilding them.

Access path update recording occurs automatically at each ASP level (independent or not, and including SYSBASE) and depends on the following parameters:

► Access path recovery time target for each ASP.
► Access path recovery estimate for each ASP. Each access path for which the estimated rebuild time is higher than the target is protected by SMAPP.

SMAPP affects the overall system performance. The lower the target recovery time that you specify for access paths, the greater this effect. Typically, the effect is not noticeable, unless the processor is nearing its capacity.

In Figure 2-8, you can see an example of an existing IASP installation.

```
                   Display Recovery for Access Paths              SYSTEMA
                                                          05/10/11  21:15:58
Estimated system access path recovery time  . . . :          44   Minutes
Total not eligible recovery time  . . . . . . . . :           0   Minutes
Total disk storage used . . . . . . . . . . . . . :     909,073   MB
% of disk storage used  . . . . . . . . . . . . . :       0,035
System access path recovery time  . . . . . . . . :          50
Include access paths  . . . . . . . . . . . . . . :        *ALL


          ------Access Path Recovery Time------  -----Disk Storage Used-----
ASP             Target           Estimated          Megabytes      ASP %
1               *NONE                    1             34,611      0,009
IASP1           *NONE                    9            265,965      0,022
IASP2           *NONE                   33            608,497      0,059



                                                                    Bottom
F3=Exit   F5=Refresh   F12=Cancel    F13=Display not eligible access paths
F14=Display protected access paths   F15=Display unprotected access paths
```

Figure 2-8   Display Recovery for Access Paths (DSPRCYAP) command result

No target exists for any IASP. Therefore, the system, by itself, starts access path journaling to achieve the overall system target, which is set to 50 minutes. By pressing the F14 key, you can see the access paths that are currently protected (Figure 2-9).

For this example, the user needs to review the targets to specify a better recovery time target, for example, by specifying *MIN, which means the lower possible recovery time.

```
                    Display Protected Access Paths              SYSTEMA
                                                    05/10/11  21:21:06
                                              Estimated
                                              Recovery
  File          Library       ASP         If Not Protected
  OSASTD10      M3EPRD        IASP1           00:03:03
  OSASTD90      M3EPRD        IASP1           00:02:57
  OSASTD00      M3EPRD        IASP1           00:02:55
  QADBIFLD      QSYS00033     IASP1           00:02:50
  MITTRA30      MVXCDTA800    IASP2           00:02:00
  MITTRA35      MVXCDTA800    IASP2           00:01:55
  OODOCU00      MVXCDTA800    IASP2           00:01:53
  MITTRAZ9      MVXCDTA800    IASP2           00:01:50
  MITTRA50      MVXCDTA800    IASP2           00:01:49
  QADBIFLD      QSYS00034     IASP2           00:01:49
  MITTRA20      MVXCDTA800    IASP2           00:01:48
  MITTRA70      MVXCDTA800    IASP2           00:01:48
                                                          More...

  F3=Exit F5=Refresh F12=Cancel F17=Top F18=Bottom
```

Figure 2-9   Protected access paths

3

IBM PowerHA SystemMirror for i architecture

This chapter explains the key components of IBM i cluster technology. Before you explore the implementation of IBM PowerHA SystemMirror for i, it is important to first understand IBM i clustering technology and capabilities.

The following topics are described in this chapter:

3.1 Cluster

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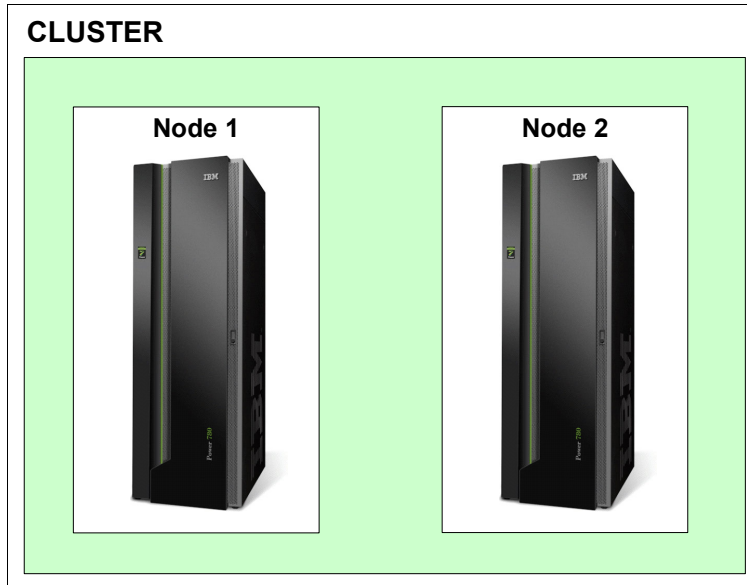
Figure 3-1   A simple cluster that consists of two nodes

Clustering offers a business the following major benefits:

► Simplified administration of servers by allowing the management of a group of systems as a single system or single database

► Continuous or high availability (HA) of systems, data, and applications

The following attributes are normally associated with the concept of clustering:

► Simplified single system management
► HA and continuous availability
► High-speed interconnect communication
► Scalability and flexibility

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

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

For more information about device domains, see 3.2, "Device domain" on page 49.

For more information about device CRGs, see 3.3, "Cluster resource group" on page 50.
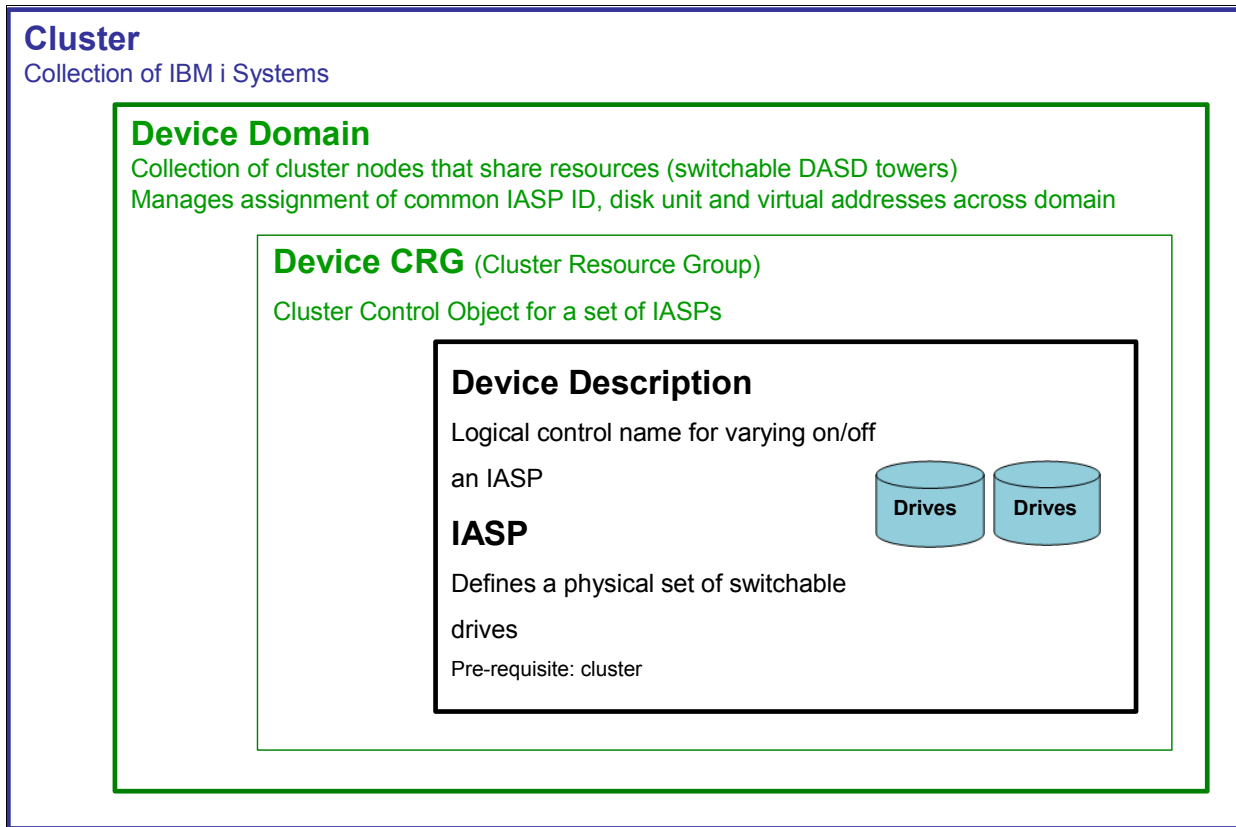
Figure 3-2   Cluster structure and IASP relationship

Basic cluster functions

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

► Heartbeat monitoring

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

► Reliable message function

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

Why you want clustering

The concept of HA in the sense of disaster recovery (DR) is an important consideration. However, disasters are not the only reason why HA is important. Disasters or unplanned outages account for only about 20% of all outages. Most outages are planned, such as a shutdown to perform an upgrade or complete a total system backup. A relatively straightforward action, such as the backup of databases and other objects, accounts for approximately 50% of all planned outages.

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

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

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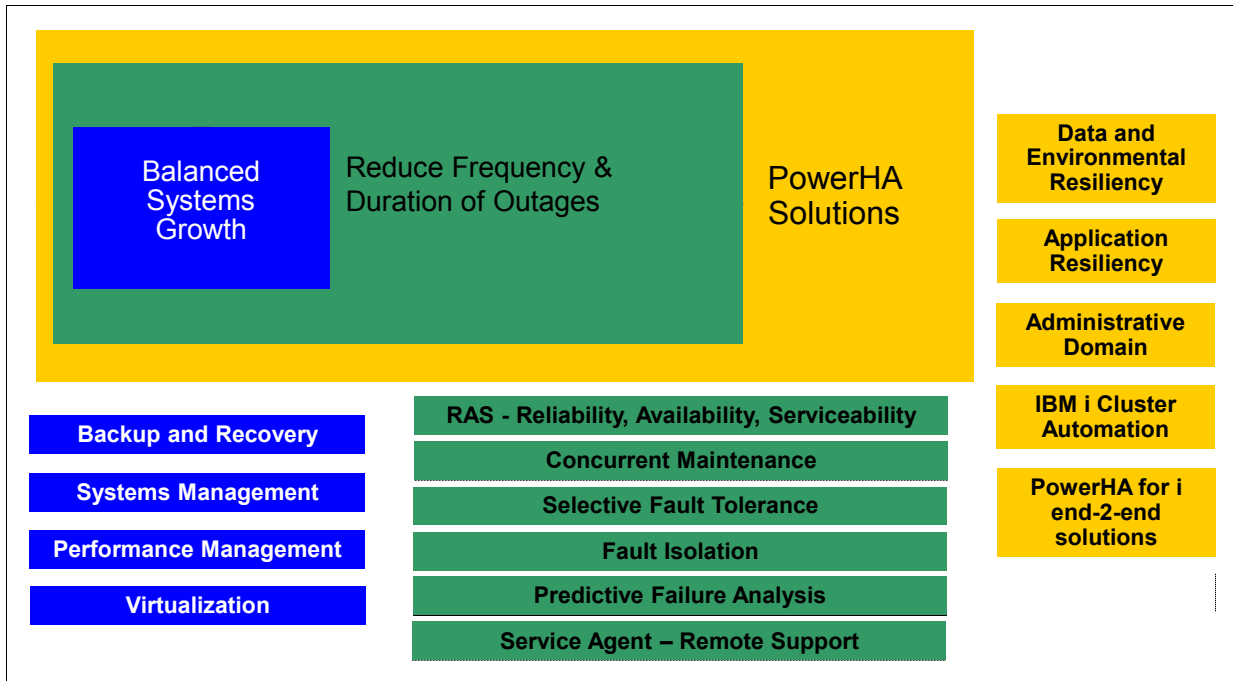
Figure 3-3   IBM i clustering technology

3.1.1 Cluster nodes

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

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

A cluster consists of a minimum of two nodes. The environment can be extended to a cluster with a maximum of 128 nodes. However, the number needs to be considered because management of a cluster with many nodes can be difficult.

A node of a cluster can fill one of three possible roles within a recovery domain:

► Primary/source node:
  – The point of access for a resilient device.
  – Contains the principal copy of any replicated resource.
  – The current owner of any device resource.
  – All CRG objects can fail over to a backup node.

► Backup node:
  – Can take over the role of primary access at the failure of the current primary node.
  – Contains a copy of the cluster resource.
  – Copies of data are kept current through replication.

► Replicate node:
  – Has copies of cluster resources.
  – Unable to assume the role of primary or backup.

Figure 3-4 shows the naming convention in the normal replication direction.
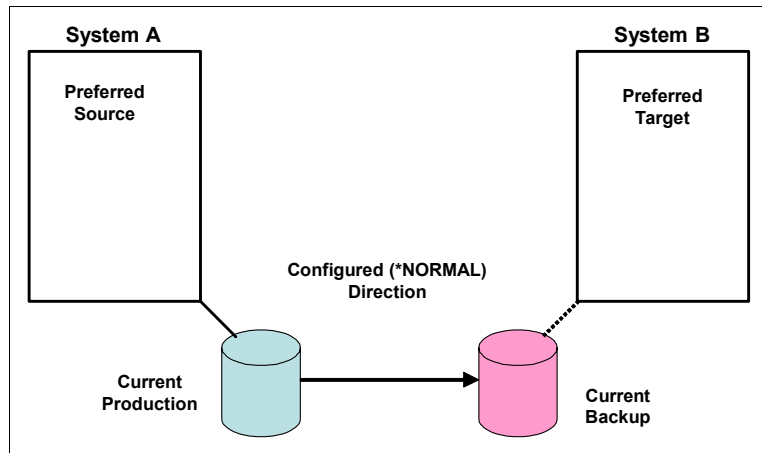
Figure 3-4   Naming convention for replication in the normal direction

Figure 3-5 shows the replication and naming convention in the reverse direction.
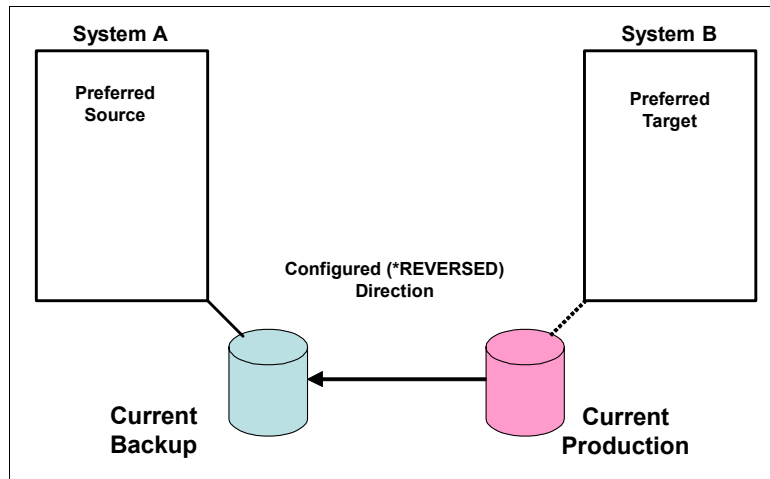
Figure 3-5   Naming convention in replication in the reverse direction

3.1.2 Advanced node failure detection

IBM PowerHA SystemMirror for i allows advanced node failure detection by cluster nodes by registering with a Hardware Management Console (HMC) or Virtual I/O Server (VIOS) management partition on systems that are managed by Integrated Virtualization Manager (IVM). Clustering is notified in a severe partition failure or system failure to trigger a cluster failover event instead of causing a cluster partition condition.

> **Important:** For node failure detection, ensure that you install the latest group program temporary fix (PTF) SF99776 - 720 High Availability for IBM i. For more information about the latest PTFs, see the IBM Support Fix Central website:
>
> http://ibm.co/1qJSF8Z

If a node detects a failure, it notifies the other nodes in the cluster by using a distress message. This message triggers a failover if the failing node is the primary node of a CRG. If instead the failure is a heartbeating failure, a partition occurs. Partitions are usually the result of network failures, and the nodes automatically merge after the problem is fixed. However, the node can fail too quickly for a distress message to go out, which can result in a "false" partition.

For LPAR failure conditions, the IBM POWER Hypervisor™ (PHYP) notifies the HMC that an LPAR failed. For system failure conditions other than a sudden system power loss, the Flexible Service Processor (FSP) notifies the HMC of the failure. Then, the Common Information Model (CIM) server on the HMC or VIOS can generate an IBM Power state change CIM event for any registered CIM clients.

> **Important:** For systems that can be managed with an HMC, the CIM server runs on the HMC. The HMC affords the most complete node failure detection because it is not part of the system and it can continue to operate when a system fails completely.

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

> **Note:** Notification of failures by an HMC or VIOS depends on a TCP/IP application server that is running on the cluster node that will receive the notification. If the application server is not running, the advanced node failure detection is unaware of node failures.

3.2 Device domain

A *device domain* is the first of the cluster constructs to define when you create a switchable IASP. It is a logical construct within cluster resource services that is used to ensure that no configuration conflicts prevent a switchover or failover.

The device domain is a subset of cluster nodes. For more information about cluster nodes, see 3.1.1, "Cluster nodes" on page 46.

The set of configuration resources that are associated with a collection of resilient devices can be switched across the nodes in the device domain. Resource assignments are negotiated to ensure that no conflicts exist. The configuration resources that are assigned to the device domain must be unique within the entire device domain. Therefore, even though only one node can use a resilient device at any time, that device can be switched to another node and brought online (Figure 3-6).
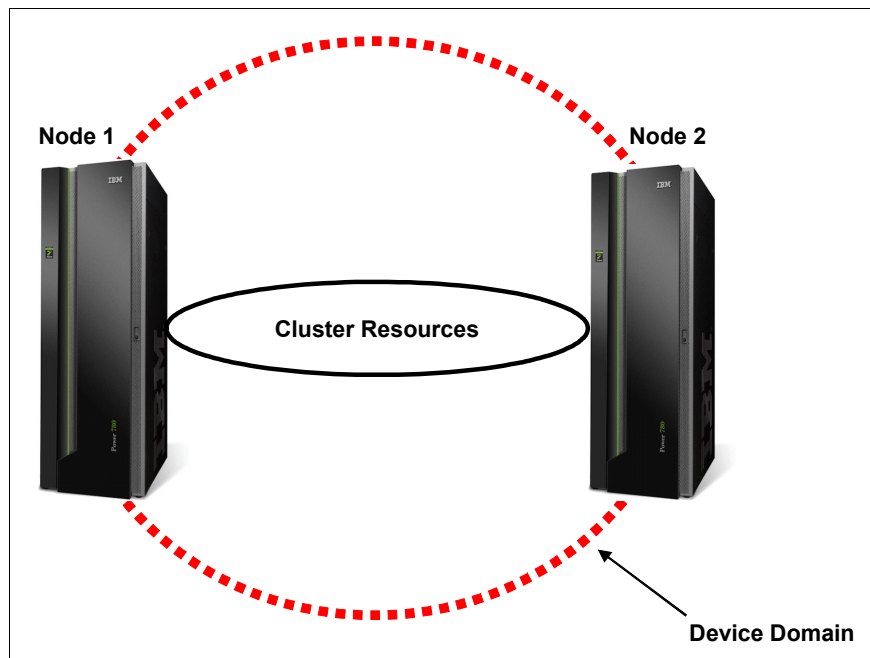
Figure 3-6   Device domain

The following cluster resources are negotiated across a device domain to ensure that no conflicts exist:

► IASP number assignments. IASPs are automatically assigned a number. The user can assign the IASP name.

► DASD unit number assignments. To keep from conflicting with the permanently attached disk units of each node, all IASP unit numbers begin with a four (4).

► Virtual address assignments.

3.3 Cluster resource group

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

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

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

Four CRG object types are used with cluster resource services:

► Application CRG

An *application CRG* enables an application (program) to be restarted on either the same node or a different node in the cluster. The takeover IP address allows access to the application without regard for the system on which the application is running. This capability allows resilient applications to be switched from one node to another.

► Data CRG

A *data CRG* enables data resiliency so that multiple copies of data can be maintained on more than one node in a cluster. A data CRG does not perform the replication, but it uses the exit program to inform a replication program when to start or end replication, and on which nodes to replicate. A data CRG does not monitor for a data resource failure.

► Device CRG

A *device CRG* supports device resiliency in IBM i high-availability environments. Device CRGs can be used to control switchable resources in an IBM i high-availability environment. The device CRG contains a list of switchable devices. The switchable devices include device descriptions, such as an independent disk pool, tape or optical device, line description, or network server. The entire collection of devices is switched to the backup node when a planned or unplanned outage occurs. Optionally, the devices can also be made available (varied on) as part of the switchover or failover process.

► Peer CRG

A *peer CRG* is a non-switchable cluster resource group in which each IBM i node in the recovery domain plays an equal role in the recovery of the node. The peer cluster resource group provides peer resiliency for groups of objects or services.

The CRG defines the recovery or accessibility characteristics and behavior for that group of resources. A CRG describes a recovery domain and supplies the name of the cluster resource group exit program that manages cluster-related events for that group.

3.3.1 Recovery domain

A *recovery domain* is a subset of nodes in the cluster that are grouped in a cluster resource group for purposes, such as performing a recovery action. Each cluster resource group has a recovery domain that is a subset of the nodes in the cluster. Recovery domains have these characteristics:

► The nodes within a recovery domain participate in any recovery actions for the resources of the domain.

► Different CRGs might have different recovery domains.

► As a cluster goes through operational changes (for example, nodes end, nodes start, and nodes fail), the current role of a node might change. Each node has a preferred role that is set when the CRG is created.

► A recovery domain can be a subset of the nodes in a cluster, and each cluster node might participate in multiple recovery domains.

3.3.2 CRG exit programs

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

An exit program is called when a CRG detects certain events, such as a node that is added to the recovery domain or the current primary node fails. The exit program is called with an action code that indicates the event. Furthermore, the exit program can indicate whether to process the event. *User-defined* means that the IBM i cluster technology does not provide the exit program. Typically, the exit program is provided by the application or data replication provider. The exit program is the way that a CRG communicates cluster events to the exit program provider. The exit program can perform the correct action based on the event, such as allowing a resource access point to move to another node.

The exit program is optional for a resilient device CRG but it is required for the other CRG types. When a cluster resource group exit program is used, it is called on the occurrence of cluster-wide events.

For more information about the cluster resource group exit programs, including what information is passed to them for each action code, see the "Cluster Resource Group Exit Program" topic in the IBM i 7.2 Knowledge Center:

http://www.ibm.com/support/knowledgecenter/ssw_ibm_i_72/apis/clrgexit.htm

3.4 IBM PowerHA SystemMirror for i technologies

This section describes the technologies for exchanging application data between IBM PowerHA SystemMirror for i cluster nodes. Clustering solutions are built on IBM PowerHA SystemMirror for i, based on the technologies of the physical exchange of the data that is included on the IASP.

The following mechanisms that are used by IBM PowerHA SystemMirror for i are described in this section:

► 3.4.1, "Host-based replication (geographic mirroring)" on page 53
► 3.4.2, "Storage-based replication" on page 54
► 3.4.3, "Cluster administrative domain" on page 59

Figure 3-7 and Figure 3-8 on page 53 provide information about IBM PowerHA SystemMirror for i feature availability in relation to the hardware that is used.

	Internal SAS/SSD [1]	DS8000	SVC / Storwize	EMC [2]
	POWER7/8	POWER7/8	POWER7/8	POWER7/8
PowerHA SystemMirror for IBM i				
FlashCopy	No	**Yes**	**Yes**	No (Timefinder)
Metro Mirror	No	**Yes**	**Yes**	No (SRDF)
Global Mirror	No	**Yes**	**Yes**	No (SRDF)
LUN Level Switching	No	**Yes**	**Yes**	No
Geographic Mirroring	**Yes**	**Yes**	**Yes**	**Yes**
IASP HyperSwap	No	**Yes** [3]	**Yes** [3]	No
PowerHA SystemMirror for IBM i *plus* IASP Copy Services Manager (ICSM, formerly Advanced Copy Services (ACS))				
FlashCopy	No	**Yes**	**Yes**	No (Timefinder)
Metro Mirror	No	**Yes**	**Yes**	No (SRDF)
Global Mirror	No	**Yes**	**Yes**	No (SRDF)
LUN Level Switching	No	**Yes**	**Yes**	No
Metro/Global Mirror	No	**Yes**	**No**	No
Full Copy Services Manager (FCSM)				
FlashCopy	No	**Yes**	**Yes**	No
Global Mirror	No	**Yes**	**Yes**	No
Metro Mirror	No	**Yes**	**Yes**	No
HyperSwap	No	**Yes**	**Yes**	No

(1) SSD requires POWER7 or later
(2) EMC (DMX, VMAX) are not supported by PowerHA except with Geographic Mirror
(3) IASP HyperSwap requires IBM i V7R3

Note. Native attach means the partition contains a SCSI, SAS, or Fibre Channel card used to connect to the storage

Figure 3-7   IBM i native attach storage and resiliency

	DS8000	XIV	SVC / Storwize
	POWER7/8	POWER7/8	POWER7/8
PowerHA SystemMirror for IBM i			
FlashCopy	Yes	No	Yes
Metro Mirror	Yes	No	Yes
Global Mirror	Yes	No	Yes
LUN Level Switch	Yes	No	Yes [1]
Geographic Mirroring	Yes	Yes	Yes
IASP HyperSwap	Yes [2]	No	Yes [2]
PowerHA SystemMirror for IBM i *plus* IASP Copy Services Manager (ICSM, formerly Advanced Copy Services (ACS))			
FlashCopy	Yes	No	Yes
Metro Mirror	Yes	No	Yes
Global Mirror	Yes	No	Yes
LUN Level Switch	Yes	No	Yes
Full Copy Services Manager (FCSM)			
FlashCopy	Yes	Yes	Yes
Metro Mirror	Yes	Yes	Yes
Global Mirror	Yes	Yes	Yes
HyperSwap (full system)	Yes	No	Yes

1   Requires NPIV capable Fibre Channel
2   IASP HyperSwap requires IBM i V7R3

Figure 3-8   IBM i PowerVM® VIOS storage and resiliency

3.4.1 Host-based replication (geographic mirroring)

Host-based replication is used to keep a consistent copy of an IASP that is assigned to one of the cluster nodes on an IASP assigned to another node. The processes that keep the IASP data in synchronization run on the nodes that own the IASPs. The Internet Protocol network is used for sending the changes to the backup IASP.

In the case of IBM PowerHA SystemMirror for i, the host-based replication feature is called *geographic mirroring*.

Figure 3-9 shows the concept of geographic mirroring.
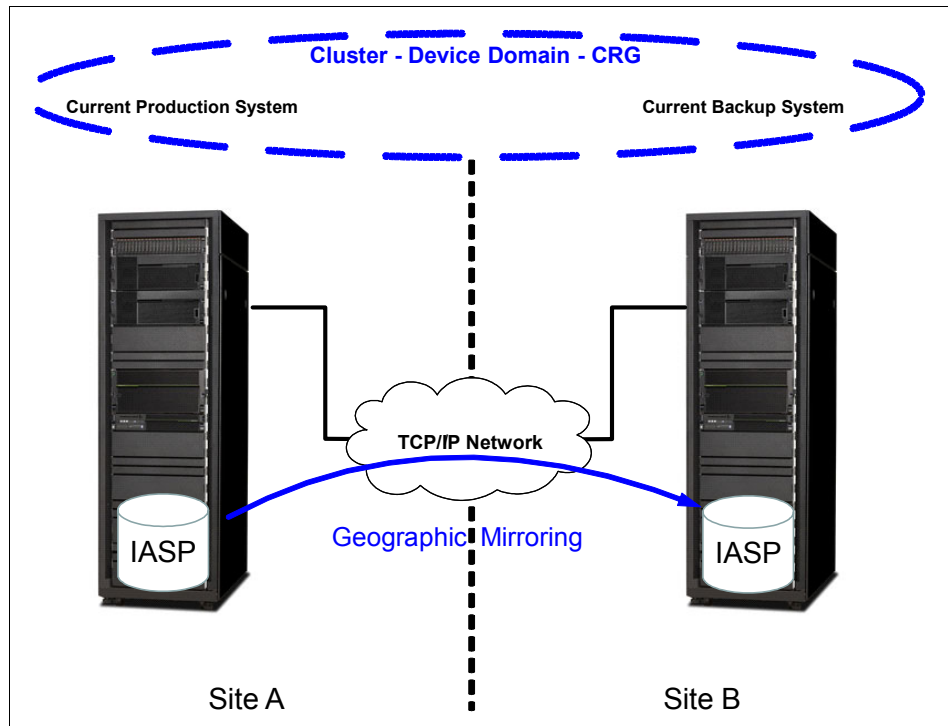
Figure 3-9   Host-based replication in IBM PowerHA SystemMirror for i: Geographic mirroring

Note: IASPs can be built on internal disk, external disk storage, or both.

Geographic mirroring guarantees that the changes to the replicated IASP are applied to the target copy in the original order.

Geographic mirroring allows for the production and mirrored copies to be in the same site for high-availability protection in a server failure. It is also possible to separate the two systems geographically for DR protection in a site-wide outage.

3.4.2 Storage-based replication

In contrast to host-based replication (geographic mirror), storage-based replication occurs at the storage subsystem device level.

The following types of the storage-level replication are supported by IBM PowerHA SystemMirror for i:

► "Metro Mirror (synchronous)" on page 55
► "Global Mirror (asynchronous)" on page 56
► "LUN-level switching" on page 57
► "FlashCopy" on page 58

Metro Mirror (synchronous)

Metro Mirror is a replication solution that is based on the synchronous data replication between external IBM storage systems that connect to IBM i. Therefore, all of the operations are acknowledged at the source system when they are completed on both storage subsystems. Data between storage systems is replicated over the storage area network (SAN) by using either Fibre Channel (FC) or FC over Internet Protocol (IP).

Synchronous replication in the case of Metro Mirror means that the source and a copy of the IASP are in a consistent state when an unexpected failure occurs on any of the cluster nodes.

When you use a Metro Mirror replication with IBM PowerHA SystemMirror for i, each node has a locally attached external IBM Storage System, and the replicated IASP must be located in it.

Figure 3-10 shows the most common Metro Mirror configuration. The system ASP can reside on internal or external disks.
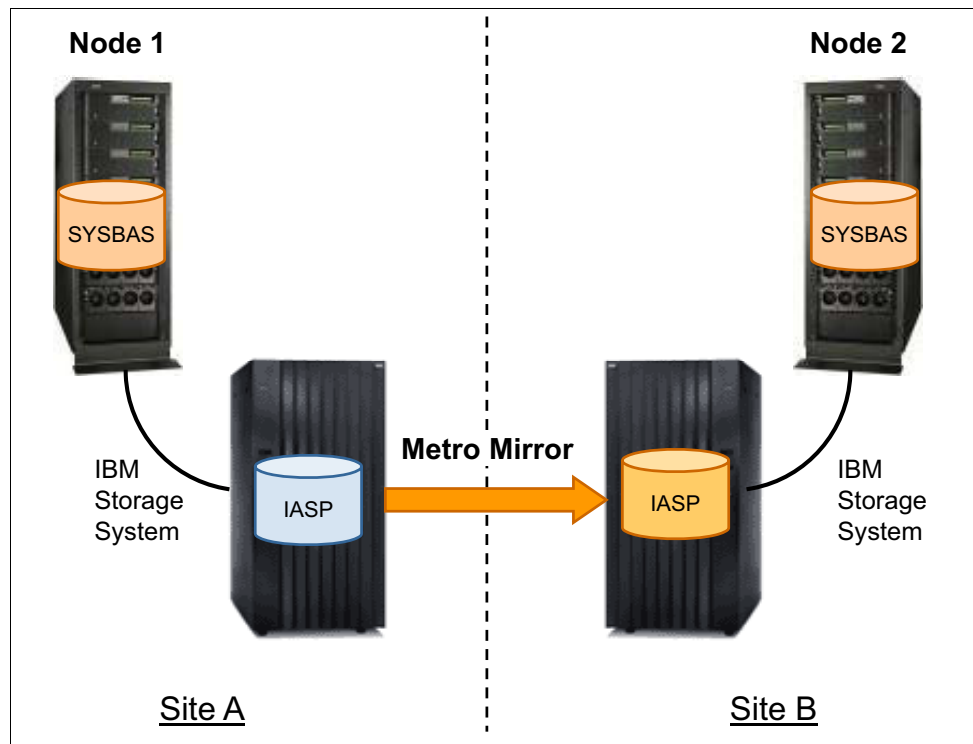
Figure 3-10   Metro Mirror architecture

In the synchronous mode of operation, the distance between the replication sites is limited to metro distances. For more information about using Metro Mirror with the DS8000 storage servers, see the IBM Redbooks publication *IBM PowerHA SystemMirror for i: Using DS8000 (Volume 2 of 4)*, SG24-8403.

In addition to IBM DS8000 storage servers, you can also build a Metro Mirror solution by using the IBM Storwize family. For more information about this type of implementation, see the IBM Redbooks publication *IBM PowerHA SystemMirror for i: Using IBM Storwize (Volume 3 of 4)*, SG24-8402.

Global Mirror (asynchronous)

Global Mirror is an asynchronous replication between local and remote IBM Storage Systems. Similar to Metro Mirror, this replication is also controlled by the storage system and data is replicated at the disk level.

Due to asynchronous replication, this solution can be used for replication between storage systems that are located a long distance from each other because the delay that is introduced by the remote write does not affect the local operations.

Global Mirror is based on the following copy services functions of a storage system:

► Global Copy
► FlashCopy
► FlashCopy consistency group

You can use FlashCopy and the FlashCopy consistency group with Global Copy to maintain a consistent copy of the IASP in the backup site. Figure 3-11 shows the general architecture for the Global Mirror replication.
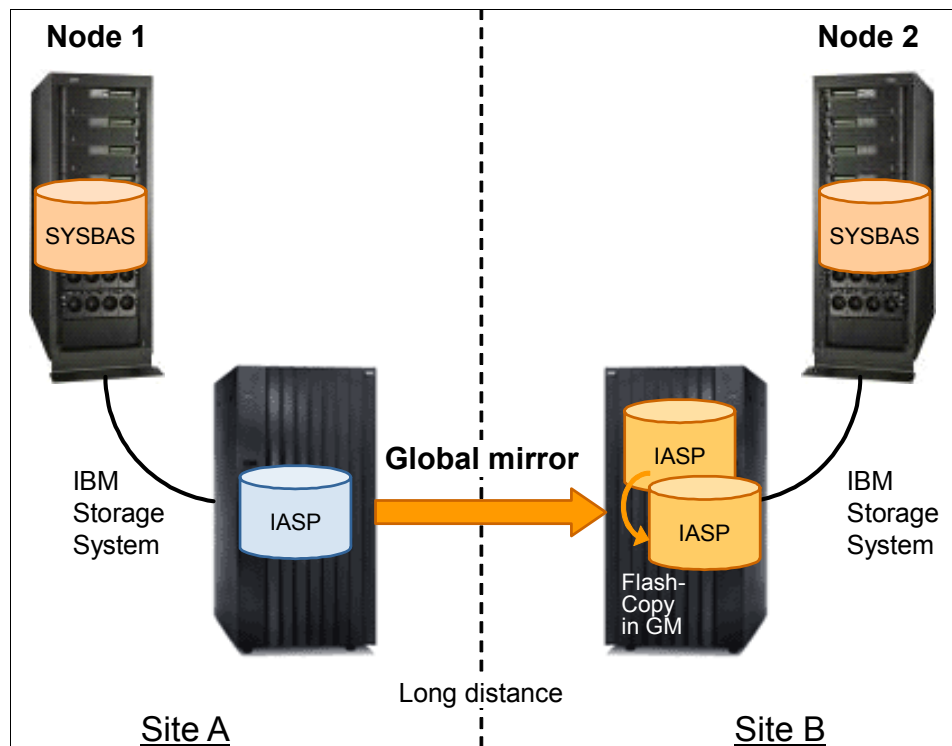
Figure 3-11   Global Mirror architecture

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 use Global Copy with certain FlashCopy operations to attain consistent data at the remote site.

For more information about using the Global Mirror solution, see the IBM Redbooks publication *IBM PowerHA SystemMirror for i: Using DS8000 (Volume 2 of 4)*, SG24-8403, which describes Global Mirror for DS8000 Copy Services.

Global Mirror can also be used in solutions that are based on the IBM Storwize Family. For more information about this type of implementation, see the IBM Redbooks publication *IBM PowerHA SystemMirror for i: Using IBM Storwize (Volume 3 of 4)*, SG24-8402.

LUN-level switching

You can use logical unit number (LUN)-level switching, which is provided by IBM PowerHA SystemMirror for i, to switch a set of LUNs (a volume group) between systems. This solution is good when a cluster resides in a single data center.

LUN-level switching supports an automated IBM i high-availability solution by using a single copy of an IASP that can be switched between two IBM i cluster nodes for a local high-availability solution.

LUN-level switching is supported for N_Port ID Virtualization (NPIV) attachment and for native attachment. It is also supported in a SAN Volume Controller Stretched Cluster environment. This solution is attractive in heterogeneous environments where a SAN Volume Controller Stretched Cluster is used as the basis for a cross-platform, two-site, high-availability solution.

Figure 3-12 shows the architecture of LUN-level switching.
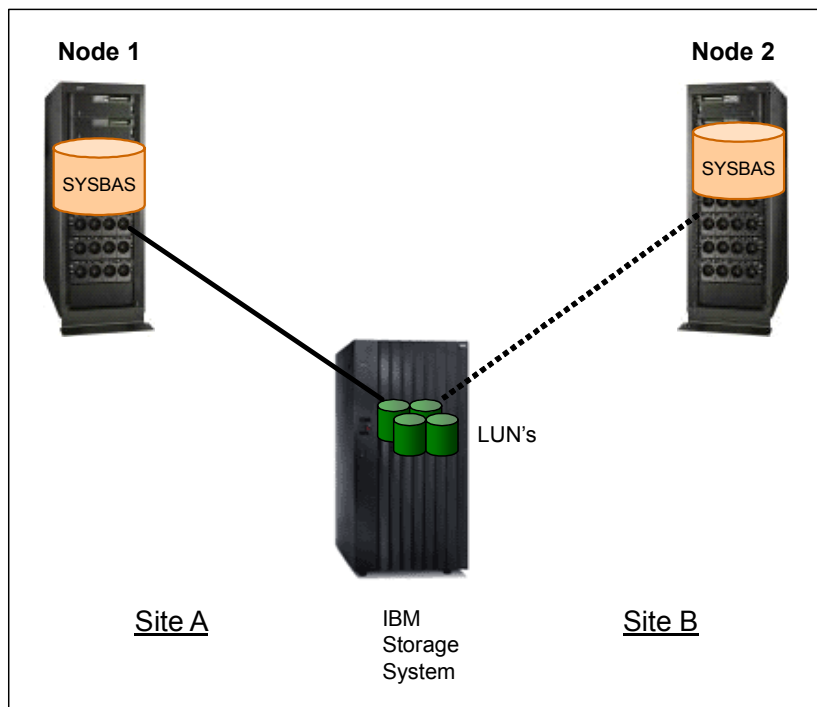
Figure 3-12   LUN-level switching architecture

SAN Volume Controller Stretched Cluster

The SAN Volume Control supports Stretched Cluster configurations where nodes within an I/O group can be physically separated from one another. This capability allows nodes to be placed in separate failure domains, which provides protection against failures that affect a single failure domain.

Figure 3-13 shows a SAN volume configuration with LUN-level switching.
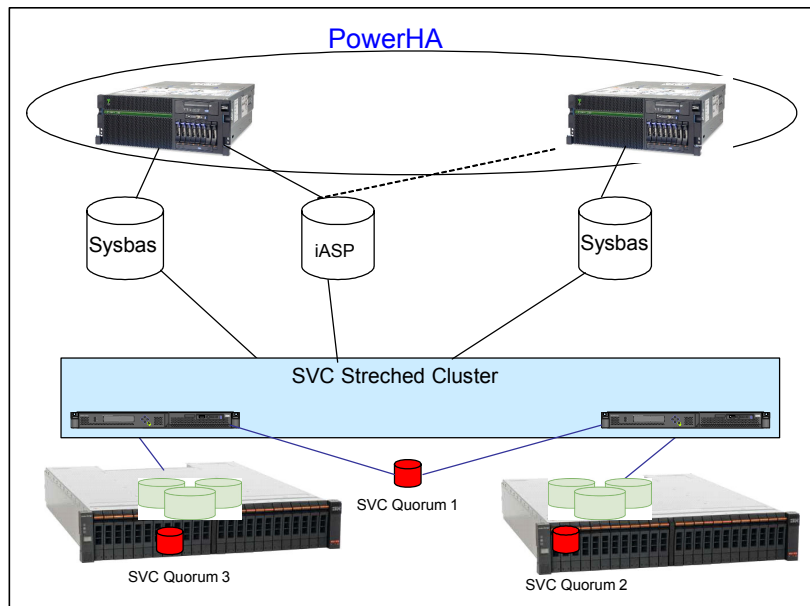
Figure 3-13   SAN volume control with LUN-level switching

The SAN Volume Controller Stretch Cluster configuration provides a continuous availability platform where host access is maintained if any single failure domain is lost. This availability is accomplished through the inherent active architecture of the SAN Volume Controller with the use of volume mirroring. During a failure, the SAN Volume Controller nodes and associated mirror copy of the data remain online and available to service all host I/O.

The Stretched Cluster configuration uses the SAN Volume Controller volume mirroring function. Volume mirroring allows the creation of one volume with two copies of managed disk (MDisk) extents. The two data copies, if placed in different MDisk groups, allow volume mirroring to eliminate the effect on volume availability if one or more MDisks fail. The resynchronization between both copies is incremental, and it is started by the SAN Volume Controller automatically. A mirrored volume has the same functions and behavior as a standard volume.

In the SAN Volume Controller software stack, volume mirroring is below the cache and copy services. Therefore, FlashCopy, Metro Mirror, and Global Mirror are unaware that a volume is mirrored. All operations that can be run on non-mirrored volumes can also be run on mirrored volumes.

FlashCopy

Use *FlashCopy* to create a point-in-time copy of the logical volume. By performing a FlashCopy, a relationship is *established* between a source volume and a target volume. Both volumes are considered to form a FlashCopy *pair*.

As a result of the FlashCopy, either all physical blocks from the source volume are copied (when you use the copy option) to the target volume, or when you use the nocopy option, only those parts are copied where data was changed in the source data since the FlashCopy was established.

The target volume needs to be the same size or larger than the source volume whenever FlashCopy is used to flash an entire volume.

Figure 3-14 shows an example of the FlashCopy architecture. In this example, when the production IASP FlashCopy is established, the copy can be accessed by another LPAR. The copy of the IASP is read and write capable, so you can use it for backup or testing purposes.
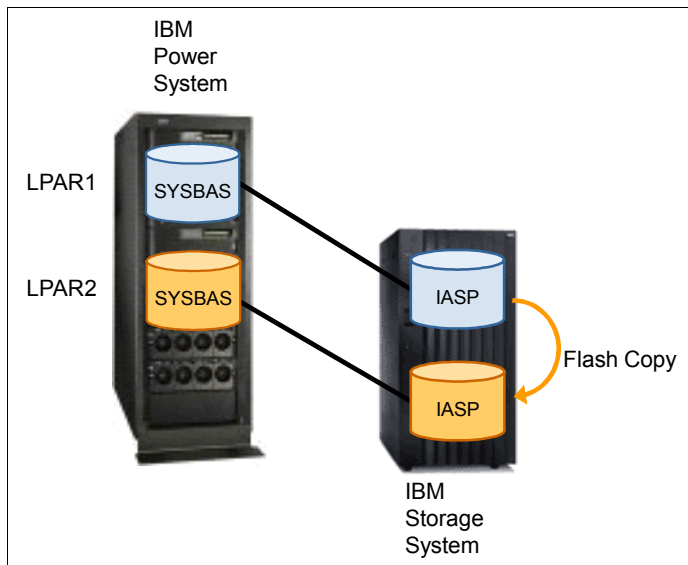
Figure 3-14   FlashCopy architecture example

IBM PowerHA SystemMirror for i supports FlashCopy and Extended FlashCopy (formerly Space Efficient). When you use classic FlashCopy, the target volume size that is reported to the system must be allocated in the storage device. When you use Extended FlashCopy, you can create virtual volumes that are space efficient and use them as a target for the FlashCopy operation. The space-efficient volumes are volumes with a defined virtual size, but the physical storage is not physically allocated to them.

3.4.3 Cluster administrative domain

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

Configuration parameters or data is associated with applications and application data. This information is known collectively as the operational environment for the application. Examples of this type of data include user profiles that are used for accessing the application or its data, or system environment variables that control the behavior of the application. With a high-availability environment, the operational environment needs to be the same on every system where the application can run, or where the application data resides. When a change is made to one or more configuration parameters or data on one system, the same change needs to be made on all systems.

Use a cluster administrative domain to identify resources that need to be maintained consistently across the systems in an IBM i high-availability environment. The cluster administrative domain monitors for changes to these resources and synchronizes any changes across the active domain.

Figure 3-15 shows a cluster architecture with a cluster administrative domain.
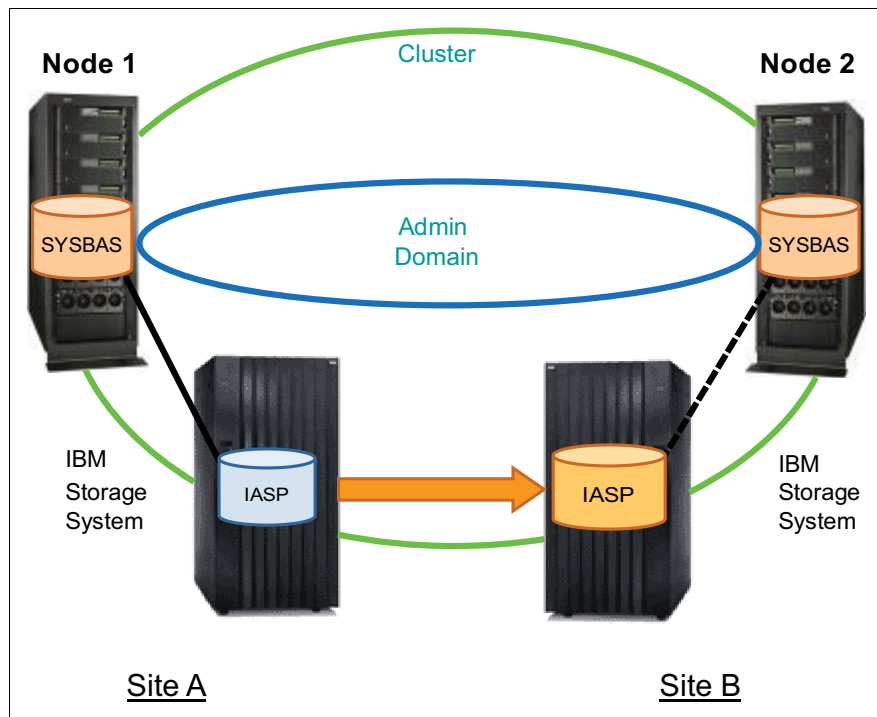
Figure 3-15   Cluster administrative domain architecture

When a cluster administrative domain is created, the system creates a peer CRG with the same name. The nodes that make up the cluster administrative domain are defined by the CRGs recovery domain. (See 3.3.1, "Recovery domain" on page 51.) Each cluster node can be defined in only one cluster administrative domain within the cluster.

After the cluster administrative domain is created, it can be managed with CL commands or the IBM PowerHA SystemMirror for i graphical user interface (GUI) in IBM Navigator for i.

Monitored resources

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

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

To add MREs, the resource can exist on a different node from the node on which the MREs are added. If the resource does not exist on every node in the administrative domain, the monitored resource is created. If a node is later added to the cluster administrative domain, the monitored resource is created. MREs cannot be added in the cluster administrative domain if the domain has a status of `Partitioned`.

The IBM PowerHA SystemMirror for i administrative domain can monitor up to 45,000 resource entries.

Table 3-1 lists the type of objects or monitored resources that can be managed in a cluster administrative domain.

Table 3-1   Supported monitored resource object types

Object or attribute description	Type
Authorization lists	*AUTL
Classes	*CLS
Ethernet line descriptions	*ETHLIN
Independent disk pool 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
Printer device descriptions for LAN connections	*PRTDEV
Printer device descriptions for virtual connections	*PRTDEV
Subsystem descriptions	*SBSD
System environment variables	*ENVVAR
System values	*SYSVAL
Tape device descriptions	*TAPDEV
Token-ring line descriptions	*TRNLIN
TCP/IP attributes	*TCPA
User profiles	*USRPRF

3.5 ASP copy descriptions

ASP copy descriptions are used by IBM PowerHA SystemMirror for i to manage Geographic Mirror, Metro Mirror, Global Mirror, and FlashCopy copies. The copy description defines a copy of an IASP. The parameters of an IASP depend on the IBM PowerHA SystemMirror for i solution.

Figure 3-16 shows the relationships between the managed objects that are described in this section. The IASPs that are in the device domain for the cluster share the same ASP copy descriptions on all nodes in the cluster.
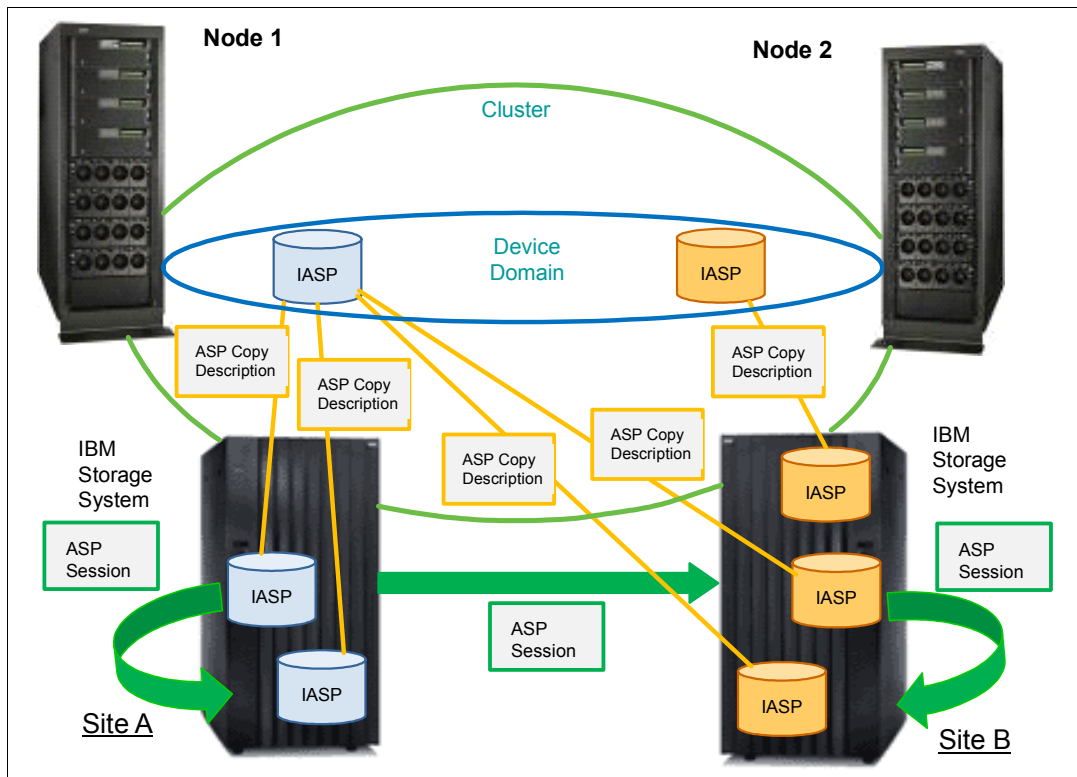
Figure 3-16   General view of the ASP copy descriptions and ASP sessions

3.5.1 ASP sessions

An *ASP session* is used to link two ASP copy descriptions and start the Copy Services functions between them. The IBM IBM PowerHA SystemMirror for i sessions allow IBM PowerHA SystemMirror for i to manage and monitor their activity. The status of the ASP session describes the status of the replication.

Figure 3-17 shows an example of the status of the ASP session.

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


                            Copy Descriptions

 ASP
 device          Name                    Role          State     Node
 IASP1           IASP1CPYD1              SOURCE         UNKNOWN   NODE1
 IASP1           IASP1CPYD2              TARGET          ACTIVE   NODE2
                                                                       Bottom
 Press Enter to continue
 F3=Exit   F5=Refresh   F12=Cancel   F19=Automatic refresh
```

Figure 3-17   ASP session example

3.6 Licensing for IBM PowerHA SystemMirror for i

This section describes the licensing information for IBM PowerHA SystemMirror for i. The following topics are covered in this section:

► 3.6.1, "Licensing considerations" on page 63
► 3.6.2, "License considerations with Live Partition Mobility" on page 67

3.6.1 Licensing considerations

Before you install the IBM PowerHA SystemMirror for i licensed product (5770-HAS), check whether the following products are in place:

► IBM i is installed on all system nodes (servers or LPARs) that will be part of your HA or DR solution.

► The HA Switchable Resources (option 41 of 5770-SS1) product is installed on all system nodes (server or LPARs) that will be part of your HA or DR solution.

Note: HA Switchable Resources is included in 5770-HAS IBM PowerHA SystemMirror for i. You do not need to order it separately. However, it still must be installed separately.

Three editions of IBM PowerHA SystemMirror for IBM i are available:

► Express Edition PID 5770-HAS, option 3

The Express Edition is the foundation for a class of HA/DR offerings based on restarting the LPAR into another LPAR on a different server for HA operations. The Express Edition allows for the primary LPAR to be restarted or IPLed to a target LPAR on another server. The Express Edition offering enables a single node, full system HyperSwap, which provides continuously available storage through either planned or unplanned storage events.

► Standard Edition PID 5770-HAS, option 2

The Standard Edition helps you to protect your critical business applications from planned or unplanned outages in the data center (single-site solution). The Standard Edition provides reliable monitoring, failure detection, and automated recovery of business application environments to backup resources, taking advantage of the IBM suite of disk storage solutions that provide the data resiliency foundation.

Use the Standard Edition to monitor numerous soft and hard errors from various event sources, such as the HMC, system operations, an application failure, or utility power loss, enabling either automated or operator-initiated actions.

The Standard Edition supports geographic mirroring in synchronous mode, FlashCopy for DS8000, and the Storwize family.

► Enterprise Edition PID 5770-HAS, option 1

The Enterprise Edition includes all of the capabilities of the Standard Edition and more. Use the Enterprise Edition package to extend your data center solution across two sites. IBM PowerHA SystemMirror for i Enterprise Edition includes support for the DS8000, SVC, and Storwize family of storage servers with either Metro Mirror or Global Mirror.

Note: The Standard Edition must be installed also along with Enterprise Edition.

Table 3-2 provides a comparison of the IBM PowerHA SystemMirror for i editions.

Table 3-2   Comparison of the IBM PowerHA SystemMirror for i editions

IBM i HA/DR clustering	Express edition	Standard edition	Enterprise edition
Centralized cluster management		✓	✓
Cluster resource management		✓	✓
Centralized cluster configuration		✓	✓
Automated cluster validation		✓	✓
Cluster administration domain		✓	✓
Cluster device domain		✓	✓
Integrated heartbeat		✓	✓
Application monitoring		✓	✓
IBM i event/error management		✓	✓
Automated planned failover		✓	✓
Managed unplanned failover		✓	✓
Centralized FlashCopy		✓	✓
LUN-level switching		✓	✓
Geomirror Sync mode		✓	✓
Geomirror Async mode		✓	✓
Multisite HA/DR management			✓
DS8000/SVC/V7000 Metro Mirror			✓
DS8000/SVC/V7000 Global Mirror			✓
HyperSwap	✓		✓

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

You can order a *Power System capacity BackUp (CBU)* model for DR and HA when you order a Power System or a model upgrade. The terms for the CBU for i allow a production system processor's optional i license entitlements and 5250 Enterprise Enablement license entitlements to be temporarily transferred to the CBU Edition when the primary system processor cores are not in concurrent use.

For an IBM i environment, you must register the order for a new CBU at the IBM Capacity Backup for Power Systems website:

https://ibm.biz/Bd4hax

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

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

For example, as shown in Figure 3-18, eight active cores with IBM i and IBM PowerHA SystemMirror for i entitlements on the primary server are shown. The CBU system contains one IBM i entitlement, one IBM PowerHA SystemMirror for i entitlement, and a temporary key (good for two years) with a usage limit of eight cores for IBM PowerHA SystemMirror for i.
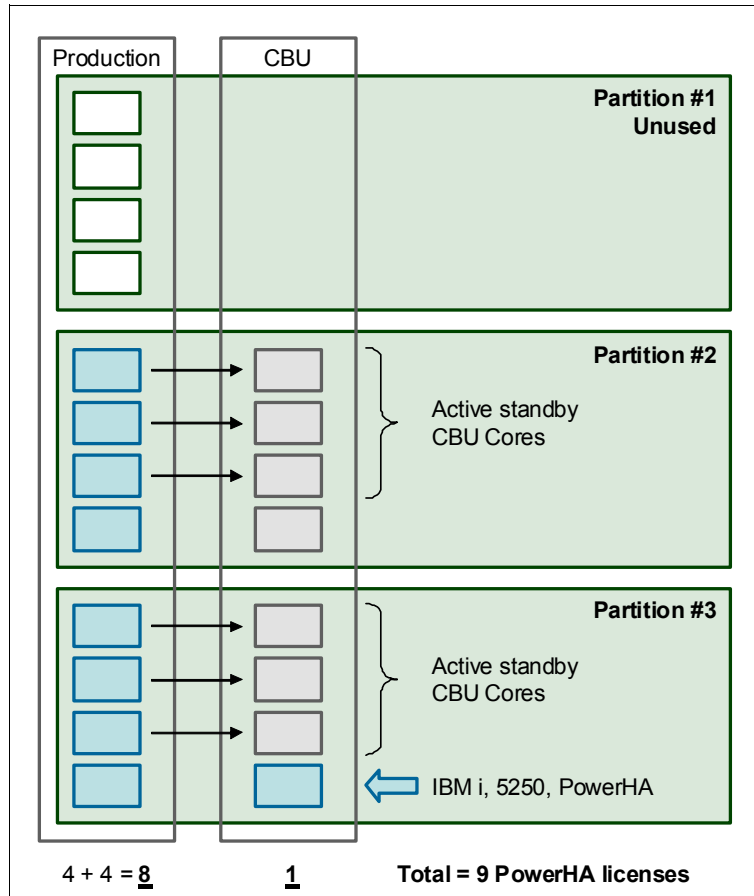
Figure 3-18   IBM PowerHA SystemMirror for i licensing for CBU

3.6.2 License considerations with Live Partition Mobility

When partitions are moved from one system to another, you must ensure that you still adhere to license requirements in your environment.

For IBM i licenses, the following rules apply:

- ► When an LPAR is moved to another system permanently, the considerations when you use Live Partition Mobility are the same as when a manual move to another system is performed. IBM i entitlements are licensed to the hardware serial number. Therefore, you need a new license on the system that the LPAR is migrated to or you can transfer entitlements from the old system to the new system under certain circumstances. Licenses for IBM i licensed programs can be moved to the new system.

- ► When an LPAR is moved to another system temporarily, the following scenarios are possible:

  - – All necessary licenses are available on both systems. No special considerations apply.

  - – The secondary system is a CBU edition. In this case, normal CBU regulations apply.

No licenses are available on the secondary system. In this case, the 70-day grace period applies. You can use the system for 70 consecutive days. When you perform multiple Live Partition Mobility operations, this 70-day grace period restarts after each migration. To work in this way, licenses must be available on the source system. Source and target systems must belong to the same company or must be leased by the same company. In addition, the source system must be in a higher or equal processor group as the target system. To receive support when you work on the target system, we suggest that a minimum of one IBM i license with a valid software maintenance contract exists on that system.

4

Creating the cluster framework

This chapter shows you how to create the following basic components that are common to most IBM PowerHA SystemMirror for i configurations:

- ► Cluster
- ► Cluster nodes
- ► Advanced node failure detection
- ► Device domain
- ► Administrative domain

Note: PowerHA licensed program 5770-HAS with Standard or Enterprise edition and IBM i 5770-SS1 base option 41 must be installed on your system in advance.

The following topics are described in this chapter:

4.1 Gathering information for your cluster configuration

You might find it helpful to determine parameter values or settings before you run the commands to configure your cluster. A blank example worksheet with parameter descriptions is provided in Appendix E, "Worksheet to configure the cluster framework" on page 169.

4.2 Creating a cluster and adding cluster nodes

This section shows you how to set up a basic cluster environment with two nodes.

Note: The configuration examples that are shown in this chapter use the values from Table 4-1. Your information will likely differ.

Table 4-1   Sample worksheet for the values that are used in the examples that are shown in this chapter

Parameter	Keyword and type	Value	Description	Commands where parameter is used
IASP	ASPDEV CHAR(10)	IASPHA	IASP name or device description.	CFGDEVASP
Cluster	CLUSTER CHAR(10)	PWRHACLU	Cluster name.	CRTCLU ADDCLUMON ADDDEVDMN CRTCAD ADDCADMRE
Node identifier (primary)	NODE CHAR(8)	ITSO1NOD	Primary or source node name.	CRTCLU ADDCLUMON ADDDEVDMNE CRTCAD
Node identifier (backup)	NODE CHAR(8)	ITSO2NOD	Backup or target node name.	CRTCLU ADDCLUMON ADDDEVDMNE CRTCAD
Node identifier (additional)	NODE CHAR(8)	N/A	Additional node name, if required.	CRTCLU ADDCLUMON ADDDEVDMNE CRTCAD
IP address (primary node)	dotted IP	192.168.80.172	Cluster IP address for the primary node. It can have 1 or 2 addresses.	CRTCLU
IP address (backup node)	dotted IP	192.168.80.182	Cluster IP address for the backup node. It can have 1 or 2 addresses.	CRTCLU ADDCLUNODE
IP address (additional node)	dotted IP	N/A	Cluster IP address for additional node. It can have 1 or 2 addresses. Not applicable for all configurations.	CRTCLU ADDCLUNODE
Device domain	DEVDMN CHAR(10)	PWRHADMN	Device domain name.	ADDDEVDMNE
Cluster administrative domain	ADMDMN CHAR(10)	PWRHACAD	Cluster administrative domain name.	CRTCAD ADDCADMRE

4.2.1 Prerequisites

Use the following prerequisite steps for creating a cluster and adding cluster nodes:

1. For cluster communications, which are also known as a *heartbeat*, a minimum configuration of one IP address is required. However, for redundancy, we suggest that you use a shared Ethernet adapter, a virtual IP address across two Ethernet adapters, or two IP addresses with two Ethernet adapters. In addition, all cluster IP addresses on a certain node must be able to communicate with all cluster IP addresses on all nodes in the cluster.

    The Cluster IP addresses for both source and target nodes must be configured and active before you proceed.

2. Run the following Start TCP Server (**STRTCPSVR**) command on all of the nodes that you plan to use in your cluster to start the InetD server, if it is not started:

    ```
    STRTCPSVR *INETD
    ```

    Note: The InetD server setting must be changed to auto start *YES or you need to ensure that the command is added to the system startup program.

3. Run the Change Network Attribute (**CHGNETA**) command on all nodes that you want in your cluster to change the Allow Add to Cluster (**ALWADDCLU**) parameter from the default of *NONE, if not already set:

    – Any other system can add this system as a node in a cluster with no required authentication:

    ```
    CHGNETA ALWADDCLU(*ANY)
    ```

    – Any other system can add this system as a node in a cluster only after the request is authenticated:

    ```
    CHGNETA ALWADDCLU(*RQSAUT)
    ```

    The ALWADDCLU network attribute is checked to see whether the node that is added is allowed to be part of the cluster and whether to validate the cluster request by using X.509 digital certificates. A *digital certificate* is a form of personal identification that can be verified electronically. If validation is required, the following products must be installed on the systems for the requesting node and the node that is being added:

    – IBM i option 34 (Digital Certificate Manager)

    – IBM i option 35 (CCA Cryptographic Service Provider)

    When *RQSAUT is specified for the ALWADDCLU parameter, the certificate authority trust list for the IBM i cluster security server application must be correctly set up. The server application identifier is QIBM_QCST_CLUSTER_SECURITY. At a minimum, add certificate authorities for those nodes that you allow to join the cluster.

4.2.2 Creating a cluster

To create a cluster, you need to specify at least one node to be a part of the cluster and you must have access to that node. A cluster is created by using the Create Cluster (**CRTCLU**) command.

Tip: We suggest that you run the **CRTCLU** command on the system or partition that is designated as the primary or source node.

Follow these steps to create a cluster:

1. Create the cluster by using the **CRTCLU** command that is shown in Figure 4-1. This step also starts the cluster node.

   The current (*CUR) target cluster version for IBM i 7.2 is 8. If you are adding nodes that are an earlier version to the cluster, this setting must be set to *PRV, which is also true for the Target PowerHA version parameter.

   The cluster message queue option defines a cluster-wide failover message queue. If a node fails, a message is sent to this message queue on the node that the cluster fails to. You can also specify how long this message will wait for an answer and the default action if no answer is provided within that time frame. For the initial setup, we suggest that you use the *NOMAX and *CANCEL options to prevent an unintended failover. These settings can be changed later.

   > **Note:** Several of the TCP/IP servers that are used by clustering require that the QUSER user profiles show STATUS = *ENABLED and that no *SECADM or *ALLOBJ special authorities exist. It must also not be expired.

```
                        Create Cluster (CRTCLU)

 Type choices, press Enter.


 Cluster  . . . . . . . . . . . . > PWRHACLU      Name
 Node list:
   Node identifier  . . . . . . . > ITSO1NOD      Name
   IP address . . . . . . . . . . > '192.168.80.172'


              + for more values
 Start indicator  . . . . . . . .   *YES          *YES, *NO
 Target cluster version . . . . .   *CUR          *CUR, *PRV
 Target PowerHA version . . . . .   *CUR          *CUR, *PRV
 Cluster message queue  . . . . . > QSYSOPR       Name, *NONE
   Library  . . . . . . . . . . . >   QSYS        Name
 Failover wait time . . . . . . . > *NOMAX        Number, *NOWAIT, *NOMAX
 Failover default action  . . . . > *CANCEL        *PROCEED, *CANCEL


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

Figure 4-1 Creating a cluster

2. Add the second node to the cluster by using the Add Cluster Node Entry (**ADDCLUNODE**) command as shown in Figure 4-2. If the start indicator parameter is set to *YES, the node is started immediately after it is added to the cluster.

```
                    Add Cluster Node Entry (ADDCLUNODE)

 Type choices, press Enter.

 Cluster  . . . . . . . . . . . > PWRHACLU      Name
 Node list:
   Node identifier  . . . . . . > ITSO2NOD      Name
   IP address . . . . . . . . . > '192.168.80.182'


 Start indicator  . . . . . . . .   *YES          *YES, *NO


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

Figure 4-2   Adding a cluster node to the cluster

3. Use the cluster menu WRKCLU, as shown in Figure 4-3, option 6 to verify that both nodes are configured and active.

```
                          Work with Cluster
                                                    System:    ITSOHA1
 Cluster  . . . . . . . . . . . . . . . . :   PWRHACLU

 Select one of the following:

      1. Display cluster information
      2. Display cluster configuration information

      6. Work with cluster nodes
      7. Work with device domains
      8. Work with administrative domains
      9. Work with cluster resource groups
     10. Work with ASP copy descriptions

     20. Dump cluster trace



 Selection or command
 ===> 6

 F1=Help   F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel
```

Figure 4-3   Work with cluster menu

Figure 4-4 shows the Work with Cluster Nodes display that shows that both nodes are active.

```
                        Work with Cluster Nodes

 Local node . . . . . . . . . . . . . . . :    ITSO1NOD
 Consistent information in cluster  . . . :    Yes


 Type options, press Enter.
   1=Add   2=Change   4=Remove   5=Display more details   6=Work with monitors
   8=Start   9=End   20=Dump trace


 Opt     Node         Status         Device Domain


         ITSO1NOD     Active
         ITSO2NOD     Active



                                                                    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 4-4   WRKCLU option 6 that shows both nodes are active

4.3 Configuring advanced node failure detection

Note: This section is only applicable to a Hardware Management Console (HMC)-managed system or Virtual I/O Server (VIOS) partitions on an Integrated Virtualization Manager (IVM)-managed system.

Setting up advanced node failure detection can help to avoid cluster partition situations that occur because a cluster node was not able to send out a panic/notification message before the node failed. Advanced node failure detection helps to allow specified failover procedures to occur.

To allow the HMC or VIOS CIM server to notify registered IBM i cluster nodes of sudden partition or system failures, you need to set up Secure Shell (SSH) communications between the HMC or VIOS and your cluster nodes. A digital certificate from the HMC or VIOS will also be copied and installed on every node in the cluster that you want to monitor.

Important: Before you set up advance node failure detection, ensure that the latest IBM PowerHA SystemMirror for i Group PTF is installed on your system.

The following steps are covered in this section:

► 4.3.1, "Setting up the IBM i partitions" on page 75
► 4.3.2, "Managed by an HMC" on page 76
► 4.3.3, "Managed by a VIOS" on page 78
► 4.3.4, "Installing a certificate on the IBM i partition" on page 78
► 4.3.5, "Adding monitors to the cluster nodes" on page 79
► 4.3.6, "Multiple HMC or VIOS environments" on page 79

4.3.1 Setting up the IBM i partitions

Follow these steps to set up SSH and other required components on the IBM i partitions. These steps must be performed on all nodes in the cluster that you want to monitor:

1. Log on to the IBM i cluster node with the QSECOFR profile or another profile that has sufficient authority.

2. You need to designate a user profile to use for secure copy operations. This profile must have a home directory in the integrated file system (IFS) on the IBM i partition. For example, if you use QSECOFR as the profile name, you need to verify that a `/home/QSECOFR` directory exists. If this directory does not exist, create it by using the command:

   `md '/home/QSECOFR'`

3. Licensed program 5733-SC1 (IBM Portable Utilities for i) with options *BASE and 1 must to be installed on your cluster nodes.

   > **Note:** If you are installing this product for the first time, be aware that it does not appear on the GO LICPGM menu. You must use the `RSTLICPGM` command. As of this writing, it is on volume B_GROUPx_04.

   > **Tip:** If you upgraded from a previous IBM i release, ensure that licensed program 5733-SC1 (IBM Portable Utilities for i) is also upgraded to the current level.

4. Licensed program 5770-UME (IBM Universal Manageability Enablement for i) must be installed on your cluster nodes.

5. Start the TCP/IP server *CIMOM if it is not started by using the following command:

   `STRTCPSVR *CIMOM`

6. The default configuration of the *CIMOM server that is provided by the installation of the 5770-UME licensed program must be changed so that the IBM i partition can communicate with the CIM server. You need to change two configuration attributes that control security aspects by running the `cimconfig` command within a PASE shell:

   a. With the *CIMOM server running, start a PASE shell:

      `call qp2term`

   b. Enter or copy and paste the following command into the PASE shell. This command is a single command string:

      `/QOpenSys/QIBM/ProdData/UME/Pegasus/bin/cimconfig -s enableAuthentication=false -p`

   c. Enter or copy and paste the following command into the PASE shell. This command is a single command string:

      `/QOpenSys/QIBM/ProdData/UME/Pegasus/bin/cimconfig -s sslClientVerificationMode=optional -p`

Figure 4-5 shows the PASE panel with the commands and expected results.

```
                            /QOpenSys/usr/bin/-sh
 #
 > /QOpenSys/QIBM/ProdData/UME/Pegasus/bin/cimconfig -s
       enableAuthentication=false -p
   PGC00218: The planned value for property enableAuthentication is set to
   false in the CIM server.
   #
 > /QOpenSys/QIBM/ProdData/UME/Pegasus/bin/cimconfig -s
       sslClientVerificationMode=optional -p
   PGC00218: The planned value for property sslClientVerificationMode is set to
   optional in the CIM server.
   #
 ===>
 F3=Exit      F6=Print    F9=Retrieve    F11=Truncate/Wrap
 F13=Clear    F17=Top     F18=Bottom     F21=CL command entry
```

Figure 4-5   PASE panel that shows the configuration commands

    d. End the PASE shell by pressing F3.

    e. Restart the *CIMOM server by using the following commands. The **ENDTCPSVR** command might take several minutes to complete.

```
ENDTCPSVR *CIMOM
STRTCPSVR *CIMOM
```

    f. Verify or change the system value QSHRMEMCTL to '1'.

    g. Start the TCP/IP server *SSHD, if not started:

```
STRTCPSVR *SSHD
```

7. Go to 4.3.2, "Managed by an HMC" on page 76 if you are using an HMC or 4.3.3, "Managed by a VIOS" on page 78 if you are using a VIOS.

4.3.2 Managed by an HMC

This section describes the required steps for an HMC-managed environment.

Local HMC

For a local HMC setup, follow these steps:

1. You need to use the physical monitor and keyboard that are attached to your HMC.

2. Select **HMC Management** on the left panel of the HMC graphical user interface (GUI). A list of options appears on the right panel.

3. Select **Open Restricted Shell Terminal**. A new shell window opens on the desktop on which you can enter commands.

4. Use the secure copy command to copy the default server.pem file to your IBM i cluster node:

```
scp /etc/Pegasus/server.pem QSECOFR@system-name:/cert-name.pem
```

> **Note:** You might consider changing the QSECOFR user profile to another IBM i user profile. Change system-name to the Domain Name System (DNS) TCP name of the IBM i partition. Change *cert-name*.pem to a name that uniquely identifies the HMC. However, you must keep the .pem extension.

5. Sign off the HMC.

6. Continue to 4.3.4, "Installing a certificate on the IBM i partition" on page 78.

Remote HMC

For a remote HMC configuration, follow these steps:

1. SSH client software must be installed on the device that you are using to connect to the HMC. PuTTY is an example of this software, and it is freely available.

2. For the HMC that you will connect to, the SSH port must be open through the local area network (LAN) adapter firewall and you must enable Remote Command execution by using SSH.

3. By using the device with the SSH client, open an SSH connection to the HMC and log in with an administrator account.

4. Use the secure copy command to copy the default server.pem file to your IBM i cluster node:

   ```
   scp /etc/Pegasus/server.pem QSECOFR@system-name:/cert-name.pem
   ```

> **Note:** Consider changing the QSECOFR user profile to another IBM i user profile. Change system-name to the TCP name of the IBM i partition. Change *cert-name*.pem to a name that uniquely identifies the HMC. However, you must keep the .pem extension.

Figure 4-6 shows an example of the SSH session.

```
login as: hscroot
Using keyboard-interactive authentication.
Password:
hscroot@ITSOHMC3:~> scp /etc/Pegasus/server.pem QSECOFR@192.168.80.171:/itsohmc3.pem
The authenticity of host '192.168.80.171 (192.168.80.171)' can't be established.
ECDSA key fingerprint is 1c:d8:c5:25:5f:e8:61:d8:86:0a:23:30:32:ab:15:29.
Are you sure you want to continue connecting (yes/no)? yes
Failed to add the host to the list of known hosts (/home/hscroot/.ssh/known_hosts).
QSECOFR@192.168.80.171's password:
server.pem                                    100% 1306     1.3KB/s   00:00
hscroot@ITSOHMC3:~> scp /etc/Pegasus/server.pem QSECOFR@192.168.80.181:/itsohmc3.pem
The authenticity of host '192.168.80.181 (192.168.80.181)' can't be established.
ECDSA key fingerprint is 4d:db:d1:a2:49:18:95:5c:87:6b:a6:03:38:66:14:14.
Are you sure you want to continue connecting (yes/no)? yes
Failed to add the host to the list of known hosts (/home/hscroot/.ssh/known_hosts).
QSECOFR@192.168.80.181's password:
server.pem                                    100% 1306     1.3KB/s   00:00
```

Figure 4-6   Sample SSH session that shows scp commands to both cluster nodes

5. Sign off the SSH session by using the `quit` command.

6. Continue to 4.3.4, "Installing a certificate on the IBM i partition" on page 78.

4.3.3 Managed by a VIOS

For a VIOS partition, follow these steps:

1. Telnet to and sign on to the VIOS partition.

2. Change to a nonrestricted shell by entering the command:

   ```
   oem_setup_env
   ```

3. Use the secure copy command to copy the default `cert.pem` file to your IBM i cluster node.

   ```
   /usr/bin/scp  /opt/freeware/cimom/pegasus/etc/cert.pem
   QSECOFR@system-name:/cert-name.pem
   ```

   > **Note:** You might consider changing the QSECOFR user profile to another IBM i user profile. Change system-name to the TCP name of the IBM i partition. Change *cert-name*.pem to a name that uniquely identifies the HMC. However, you must keep the .pem extension.

4. Start the CIM server in the VIOS partition:

   ```
   startnetsvc cimserver
   ```

5. Sign off the VIOS partition.

6. Continue to 4.3.4, "Installing a certificate on the IBM i partition" on page 78.

4.3.4 Installing a certificate on the IBM i partition

Follow these steps to install a certificate on the IBM i partition:

1. On your IBM i cluster node, start the PASE shell environment:

   ```
   call qp2term
   ```

2. Move the digital certificate that was copied from the HMC or VIOS to the truststore by using the following **mv** command:

   ```
   mv /cert-name.pem
   /QOpenSys/QIBM/UserData/UME/Pegasus/ssl/truststore/cert-name.pem
   ```

   Replace both *cert-name*.pem file name instances with the specific file name that was specified in the previous HMC or VIOS section.

3. Add the digital certificate to the truststore by using the following command:

   ```
   cd/QOpenSys/QIBM/ProdData/UME/Pegasus /bin/cimtrust -a -U QSECOFR -f
   /QOpenSys/QIBM/UserData/UME/Pegasus/ssl/truststore/cert-name.pem -T s
   ```

   Replace the *cert-name*.pem file name with the specific file name that was chosen in the previous HMC or VIOS section.

   You must run these commands on each node in the cluster that you want to monitor.

4. Exit the PASE shell by pressing F3.

5. Restart the CIM server to include the new certificate by running the following commands:

   ```
   ENDTCPSVR *CIMOM
   STRTCPSVR *CIMOM
   ```

4.3.5 Adding monitors to the cluster nodes

Use the Add Cluster Monitor (**ADDCLUMON**) command to add cluster nodes to the monitor as shown in Figure 4-7. The CIM server entries refer to the HMC or VIOS host name and logon credentials.

```
                    Add Cluster Monitor (ADDCLUMON)

 Type choices, press Enter.

 Cluster  . . . . . . . . . . . > PWRHACLU      Name
 Node identifier  . . . . . . . > ITSO1NOD      Name
 Monitor type . . . . . . . . .   *CIMSVR       *CIMSVR
 CIM server:
   CIM server host name . . . . .   userhmc.company.com
   CIM server user id . . . . . .   hmcuser
   CIM server user password . . .   hmcpassword
```

Figure 4-7   Adding the first node to the cluster monitor

Add more nodes to the cluster monitor as shown in Figure 4-8.

```
                    Add Cluster Monitor (ADDCLUMON)

 Type choices, press Enter.

 Cluster  . . . . . . . . . . . > PWRHACLU      Name
 Node identifier  . . . . . . . > ITSO2NOD       Name
 Monitor type . . . . . . . . .   *CIMSVR       *CIMSVR
 CIM server:
   CIM server host name . . . . .   userhmc.company.com
   CIM server user id . . . . . .   hmcuser
   CIM server user password . . .   hmcpassword
```

Figure 4-8   Adding more nodes to the cluster monitor

4.3.6 Multiple HMC or VIOS environments

If your cluster contains multiple HMCs or VIOS partitions and you want to set up advanced node failure detection on them, you must run these procedures with each HMC or VIOS. Begin again at 4.3.2, "Managed by an HMC" on page 76, or 4.3.3, "Managed by a VIOS" on page 78. Be certain to specify a unique *cert-name*.pem file name for each HMC or VIOS certificate.

4.4 Completing the cluster configuration

This section adds the nodes to a device domain. A device domain is required to allow nodes to participate in a highly available environment.

Next, you set up the cluster administrative domain. Although the cluster administrative domain is not required, we suggest that you use an administrative domain to synchronize various resources across the cluster nodes that are unable to be replicated otherwise.

Follow these steps:

1. Add the nodes to the device domain by using the Add Device Domain Entry (**ADDDEVDMNE**) command as shown in Figure 4-9.

> **Important:** The node that contains the independent auxiliary storage pool (IASP), which is referred to as the primary or source node typically, must be added first.

```
                    Add Device Domain Entry (ADDDEVDMNE)

 Type choices, press Enter.

 Cluster  . . . . . . . . . . . . > PWRHACLU      Name
 Device domain  . . . . . . . . . > PWRHADMN      Name
 Node identifier  . . . . . . . . > ITSO1NOD      Name



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

Figure 4-9   Adding the first node to the device domain

2. Add the second node to the device domain as shown in Figure 4-10. An IASP cannot be configured for this node.

```
                    Add Device Domain Entry (ADDDEVDMNE)

 Type choices, press Enter.

 Cluster  . . . . . . . . . . . . > PWRHACLU      Name
 Device domain  . . . . . . . . . > PWRHADMN      Name
 Node identifier  . . . . . . . . > ITSO2NOD      Name



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

Figure 4-10   Adding the second node to the device domain

3. Create the cluster administrative domain by using the Create Cluster Admin Domain (**CRTCAD**) command as shown in Figure 4-11. One or more nodes can be specified when you create the cluster administrative domain.

```
                   Create Cluster Admin Domain (CRTCAD)

 Type choices, press Enter.

 Cluster  . . . . . . . . . . .   PWRHACLU      Name
 Cluster administrative domain  .   PWRHACAD      Name
 Admin domain node list . . . . .   ITS01NOD      Name
               + for more values   ITS02NOD
 Synchronization option . . . . .   *LASTCHG      *LASTCHG, *ACTDMN




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

Figure 4-11   Creating the administrative domain that contains the cluster nodes

The Synchronization option determines how monitored resource changes are applied when a node is added to the administrative domain:

– If *LSTCHG is specified, the monitored resource on the node with the last or most recent change is applied, including the node that is being added, to all active nodes in the domain.

– If *ACTDMN is specified, only changes from active nodes are applied. Changes on the node that is being added, even if those changes are the latest or most recent, are discarded and the monitored resource values are synchronized with the active domain.

> **Important:** The "last change" of a monitored resource is determined by the time and date on that object. You must verify that the time, time zone, and date settings are accurate for all nodes in the administrative domain.

More nodes can be added to the administrative domain by using the Add Cluster Admin Domain Node Entry (**ADDCADNODE**) command.

4. Start the administrative domain by using the Start Cluster Admin Domain (**STRCAD**) command as shown in Figure 4-12.

```
                    Start Cluster Admin Domain (STRCAD)

 Type choices, press Enter.

 Cluster  . . . . . . . . . . . . > PWRHACLU      Name
 Cluster administrative domain  . > PWRHACAD      Name




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

Figure 4-12   Starting the administrative domain

5. Monitored resources are added to the administrative domain by using the Add Admin Domain MRE (**ADDCADMRE**) command. Figure 4-13 shows the IASP description example. You might use this example to create the IASP device description on all other nodes in the cluster.

```
                    Add Admin Domain MRE (ADDCADMRE)

 Type choices, press Enter.

 Cluster  . . . . . . . . . . . . > PWRHACLU      Name
 Cluster administrative domain  . > PWRHACAD      Name
 Monitored resource . . . . . . . > IASPHA        Character value
 Monitored resource type  . . . . > *ASPDEV       *ASPDEV, *AUTL, *CLS...
 Node identifier  . . . . . . . .   *             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 4-13   Adding the monitored resource example for the IASP

The Node identifier specifies the cluster node that provides the monitored resource and its attributes. The default is the local node. All or several attributes of a resource can be monitored. If only specific attributes are required, each attribute name must be listed.

Figure 4-14 shows a subsystem description that is added as a monitored resource.

```
                    Add Admin Domain MRE (ADDCADMRE)

 Type choices, press Enter.

 Cluster  . . . . . . . . . . . . > PWRHACLU      Name
 Cluster administrative domain  . > PWRHACAD      Name
 Monitored resource . . . . . . . > QBATCH        Character value
 Monitored resource type  . . . . > *SBSD         *ASPDEV, *AUTL, *CLS...
 Node identifier  . . . . . . . .   *             Name, *
 Library  . . . . . . . . . . . . > QSYS          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 4-14   Adding a monitored resource example for a subsystem description

Figure 4-15 shows a user profile that is added as a monitored resource.

Certain attributes of a user profile, including but not limited to, job descriptions, initial menus/programs, and authorization lists, must exist on all active nodes in the domain before you add the profile as a monitored resource.

```
                    Add Admin Domain MRE (ADDCADMRE)

 Type choices, press Enter.

 Cluster  . . . . . . . . . . . . > PWRHACLU      Name
 Cluster administrative domain  . > PWRHACAD      Name
 Monitored resource . . . . . . . > QSECOFR       Character value
 Monitored resource type  . . . . > *USRPRF       *ASPDEV, *AUTL, *CLS...
 Node identifier  . . . . . . . .   *             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 4-15   Adding a monitored resource example for user profile

With the **ADDCADMRE** command, you can add only one monitored resource at a time. The PowerHA GUI might be more efficient if many common resources (such as user profiles) need to be added. An example is shown in Figure 4-16.
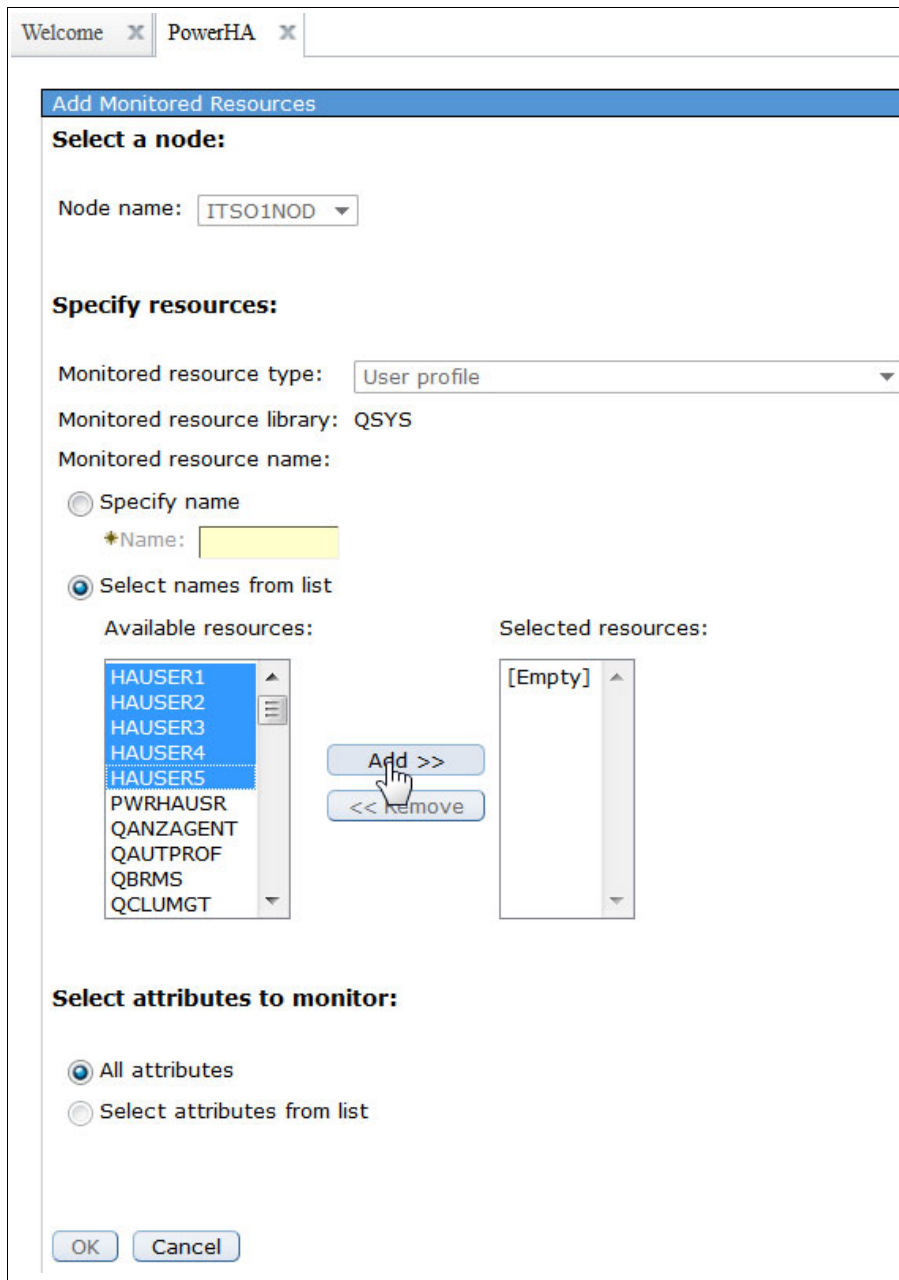

Figure 4-16   Adding multiple resources at one time by using the PowerHA GUI

4.5 Verifying requirements for PowerHA

On the main PowerHA GUI window, use the drop-down Select Action menu next to the cluster name to select **Check Requirements**. An example is shown in Figure 4-17.

Any listed items must be reviewed to determine whether action is needed. You must review system and application implications before you change anything. Selecting the two-arrow icon (>>) at the end of a line presents the Fix option so that PowerHA makes the suggested change.

Figure 4-17   PowerHA check requirements display

5

Choosing a replication solution

This chapter provides an overview of the various IBM PowerHA SystemMirror for i replication solutions. The intent of this chapter is to help you determine a suitable solution for your environment that meets your business needs.

The following topics are described in this chapter:

- ► 5.1, "Establishing your business needs" on page 88
- ► 5.2, "Synchronous solutions" on page 90
- ► 5.3, "Asynchronous solutions" on page 92
- ► 5.4, "Three-site solutions" on page 93
- ► 5.5, "Backup window reduction solutions" on page 95
- ► 5.6, "Where do you go next" on page 96

5.1 Establishing your business needs

The most important consideration before you choose an availability solution is *"What am I trying to achieve for the business?"* Four considerations are key:

► Recovery point objective (RPO)
► Recovery time objective (RTO)
► Cost of downtime to the business
► Independent auxiliary storage pool (IASP) or full system solution

Unless you understand these points well, it is difficult to identify your required solution. This kind of information is normally documented in a service level agreement (SLA) between the IT operation and the business or customer. If you do not have this kind of agreement, consider implementing it. Unless you document the business requirements, you can guarantee that the business expectation will be for zero downtime and zero loss of data.

When you think about your business needs, consider the other systems in your enterprise. You need to be able to switch systems individually to alternative sites. However, this capability introduces considerations about how the ancillary systems still communicate. Although this problem is not a large problem, it often is overlooked.

5.1.1 Recovery point objective and recovery time objective

RPO and RTO are the two principal metrics for availability solutions. An example of a failure timeline is show in Figure 5-1. This timeline indicates the RPO extending back in time from the point of failure on the left side, and the RTO extending forward in time on the right side.
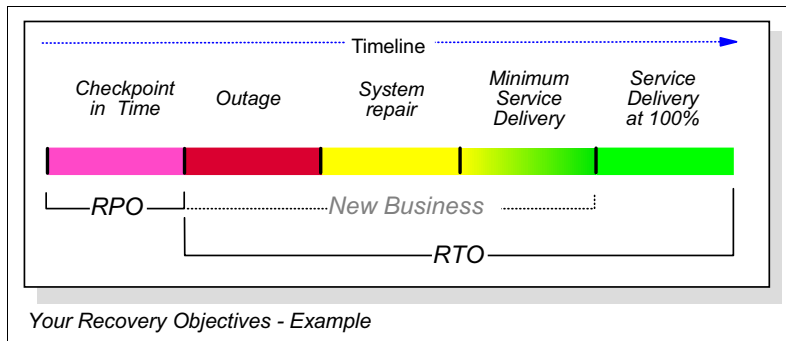
Figure 5-1   Timeline of a failure

Recovery point objective

The RPO is how much data can you afford to lose.

If you cannot tolerate any loss of committed data, your RPO is zero, which means that you need to look at a synchronous solution (5.2, "Synchronous solutions" on page 90), logical unit number (LUN) switching, Metro Mirror, HyperSwap solution, or a three-site solution (5.4, "Three-site solutions" on page 93).

Or, if you can tolerate partial data loss, for example, if you can identify missing transactions and reprocess them, consider a Global Mirror solution (5.3, "Asynchronous solutions" on page 92).

Either solution might require you to eliminate, as much as possible, your current backup window. Sometimes, you use While Active technology. However, if you choose to implement an external storage solution, consider backup window reduction solutions (5.5, "Backup window reduction solutions" on page 95).

Recovery time objective

The RTO is how long (from the point of failure) can you take to recover the system.

Every client situation differs. Certain clients cannot afford to be down any longer than necessary. Other clients can be down if it is less than a total of *x* hours in a day.

When you define your RTO, be careful to consider what you want to include in your definition. For example, the system might be recovered within the required time, but you cannot connect to the outside world as you move IP addresses by reconfiguring the Domain Name System (DNS) server, which takes time. Often, these considerations can be resolved easily (in this example, by using a virtual IP address and RIP2 or Open Shortest Path First (OSPF) to dynamically relocate the IP connectivity).

The shorter desired RTO that you want, the more you need to consider an independent auxiliary storage pool (IASP) solution, which might deliver availability. A full system solution can deliver a disaster recovery (DR) solution only. The difference is the length of time that it takes to fail over to the backup node.

With an IASP solution, you recover only the database and the system automatically ensures that uncommitted transactions are rolled back. Then, you are ready to start your applications.

With a full system solution, you recover the database the same way, and the system must sort out anything that was in flight when the operating system crashed. This situation is known as an *abnormal IPL* and it can take several hours to recover.

5.1.2 Cost of downtime

Whenever your systems are not available, the business and either direct or indirect costs are affected. For example, for an online store, you can calculate the amount of business each hour, so it is a simple cost to determine. Conversely, your accountants might need to work out the costs. Either way, a cost is associated with downtime.

You need to know the cost of downtime because this cost helps you identify the budget that you potentially can spend on eliminating downtime.

Different solution costs vary, depending on where you are starting from. For example, if DS8000 storage is installed, those same solutions are less expensive than internal disks only.

5.1.3 IASP or a full system solution

The final consideration is whether you want to migrate to an IASP solution or stay with a full system solution.

5.2 Synchronous solutions

This section contains a brief overview of the following synchronous options that are available for IBM i:

- ► 5.2.1, "Synchronous geographic mirroring" on page 90
- ► 5.2.2, "LUN-based switching" on page 90
- ► 5.2.3, "Metro Mirror" on page 91
- ► 5.2.4, "HyperSwap" on page 91

5.2.1 Synchronous geographic mirroring

Geographic mirroring is host-based replication that is handled within operating system storage management. It works by replicating changed pages of memory. By working at the system storage management level, it is not affected by any issues that result from object locking.

Geographic mirroring is a sensible option for smaller systems. As the size of the IASP grows, the demand on the communications lines grows. The limitation is in terms of the ability to perform a full resynchronization if needed in an acceptable time frame. Unless you are prepared to provide the required communications infrastructure, do not consider this solution for IASPs that are larger than 5 Tb.

Consider the CPU and memory requirements on both the source and target systems because CPU and memory can affect application performance in a synchronous transmission and synchronous mirroring mode.

Table 5-1 provides a comparison of the synchronous geographic mirroring options that are available for IBM i.

Table 5-1 Synchronous geographic mirroring options

Type of storage	Supported	IASP	Full system
DS8000	Yes	Yes - HA	No
Storwize/SVC	Yes	Yes - HA	No
Other storage	Yes	Yes - HA	No

5.2.2 LUN-based switching

LUN-based switching is not a replication solution, but you can use it to move storage from one system to another in a similar way to a replication switch. This approach protects you against a system failure, but not from a storage failure. By providing the crash consistent image to the secondary system unit quickly, it provides a fast switching high-availability solution with a single point of failure in the storage unit.

Table 5-2 provides a comparison of the LUN-based switching options that are available for IBM i.

Table 5-2   LUN switching options

Type of storage	Supported	IASP	Full system
DS8000	Yes	Yes - HA	No
Storwize/SVC	Yes	Yes - HA	No
Other storage	No	No	No

5.2.3 Metro Mirror

Metro Mirror is storage-based replication that is handled within the external storage unit. It works on replicating changed sectors on disk. Because it is a storage function, it does not suffer from any issues that result from object locking and it also does not require any CPU or system memory for its operation.

Metro Mirror is a sensible option for a large range of systems, including large quantities of data over metropolitan distances. Consider what storage system will provide the best level of availability and function, especially when you need enterprise levels of availability.

Table 5-3 provides a comparison of the Metro Mirror options that are available for IBM i.

Table 5-3   Metro Mirror options

Type of storage	Supported	IASP	Full system
DS8000	Yes	Yes - HA	Yes - DR[1]
Storwize/SVC	Yes	Yes - HA	Yes - DR[1]
Other storage	No	No	No
[1] Requires IBM Systems Lab Services software for automation. For more information, see Appendix A, "PowerHA Tools for IBM i" on page 141.			

5.2.4 HyperSwap

HyperSwap is an automated storage failover solution for synchronous storage-based replication. It protects your IBM i system when it is attached to both the primary and secondary storage system from a primary storage failure by either a host-managed (DS8000 series) or storage-managed (clustered Storwize and SVC family) automated storage failover to the secondary storage system, but not from a system failure.

If a primary storage failure occurs, the failover to the secondary storage is transparent to the applications and no disruption occurs. For additional protection against system failure, HyperSwap can be combined with IBM PowerHA SystemMirror for i and IASP storage-based replication.

Table 5-4 provides a comparison of the HyperSwap options that are available for IBM i.

Table 5-4   HyperSwap options

Type of storage	Supported	IASP	Full system
DS8000	Yes	Yes - HA	Yes - HA[1]
Storwize/SVC	Yes	No	Yes - HA
Other storage	No	No	No
[1] Requires PowerHA Express Edition.			

5.3 Asynchronous solutions

This section contains a brief overview of the asynchronous options that are available for IBM i.

5.3.1 Asynchronous geographic mirroring

Geographic mirroring is host-based replication that is handled within operating system storage management. It works on replicating changed pages of memory. Because it is at the system storage management level, it does not suffer from any issues that result from object locking.

Geographic mirroring is a sensible option for smaller systems. As the size of the IASP grows, the demand on the communications lines grows. The limitation is in terms of the ability to perform a full resynchronization if necessary in an acceptable time frame. Unless you are prepared to provide the required communications infrastructure, do not consider this solution for IASPs that are larger than 5 Tb.

Consider the CPU and memory requirements on both the source and target systems because CPU and memory can affect application performance in asynchronous transmission and asynchronous mirroring mode.

Table 5-5 provides a comparison of the asynchronous geographic mirroring options that are available for IBM i.

Table 5-5   Asynchronous geographic mirroring options

Type of storage	Supported	IASP	Full system
DS8000	Yes	Yes - DR	No
Storwize/SVC	Yes	Yes - DR	No
Other storage	Yes	Yes	No

5.3.2 Global Mirror

Global Mirror is storage-based replication that is handled within the external storage unit. It works on replicating changed sectors on disk. Because it is a storage function, it does not suffer from any issues that result from object locking. It also does not require any CPU or system memory for its operation.

Global Mirror is a sensible option for a large range of systems, including large quantities of data over long distances. Consider which storage system will provide the best level of availability and function, especially when you need enterprise levels of availability.

Table 5-6 provides a comparison of the Global Mirror options that are available.

Table 5-6   Global Mirror options

Type of storage	Supported	IASP	Full system
DS8000	Yes	Yes - DR	Yes - DR[1]
Storwize/SVC	Yes	Yes - DR	Yes - DR[1]
Other storage	No	No	No
[1] Requires IBM Systems Lab Services software for automation. For more information, see Appendix A, "PowerHA Tools for IBM i" on page 141.			

5.4 Three-site solutions

This section contains a brief overview of the following three-site options that are available for IBM i:

- ► 5.4.1, "LUN switching + Metro Mirror or Global Mirror (IASP)" on page 93
- ► 5.4.2, "Metro Mirror + Metro Mirror" on page 93
- ► 5.4.3, "Metro Mirror + Global Mirror" on page 94

5.4.1 LUN switching + Metro Mirror or Global Mirror (IASP)

It is possible to combine LUN switching with a Metro Mirror or a Global Mirror solution. This solution provides a fast high availability (HA) switch at the local site that is combined with a DR switch at the remote site. LUN switching is possible on each end of the replication link.

Table 5-7 provides a comparison of LUN switching that is combined with replication options that are available for IBM i.

Table 5-7   LUN switching that is combined with replication options

Type of storage	Supported	IASP	Full system
DS8000	Yes[1]	Yes[1]	No
Storwize/SVC	Yes	Yes	No
Other storage	No	No	No

5.4.2 Metro Mirror + Metro Mirror

The Metro Mirror + Metro Mirror solution is a three-site solution that consists of three separate copies of the IASP on disk that has a Metro Mirror relationship among all three of the nodes.

Figure 5-2 shows an example of this kind of configuration. An active Metro Mirror relationship is between Site 1 and Site 2 and a separate active Metro Mirror relationship is between Site 1 and Site 3. In addition, an idle Metro Mirror relationship is between Site 2 and Site 3.

Figure 5-2   Metro Mirror + Metro Mirror three site solution

Table 5-8 provides a comparison of the Metro Mirror + Metro Mirror replication options that are available for IBM i.

Table 5-8   Metro Mirror + Metro Mirror replication options

Type of storage	Supported	IASP	Full system
DS8000	Yes[1]	Yes[1]	Yes (no PowerHA support)
Storwize/SVC	No	No	No
Other storage	No	No	No
[1] Requires IBM Systems Lab Services software for automation. For more information, see Appendix A, "PowerHA Tools for IBM i" on page 141.			

5.4.3 Metro Mirror + Global Mirror

Metro Mirror + Global Mirror is a three-site solution that consists of three separate copies of the IASP on disk with a Metro Mirror relationship among all three of the nodes.

Figure 5-3 shows an example of this kind of configuration. An active Metro Mirror relationship is between Site 1 and Site 2. A separate active Global Mirror relationship is between Site 1 and Site 3. In addition, an idle Global Mirror relationship is between Site 2 and Site 3.
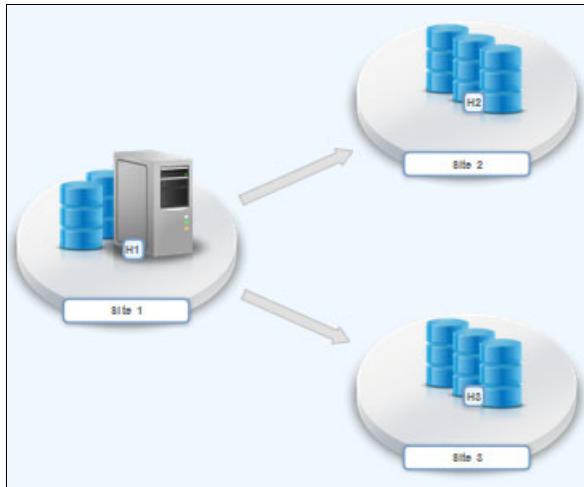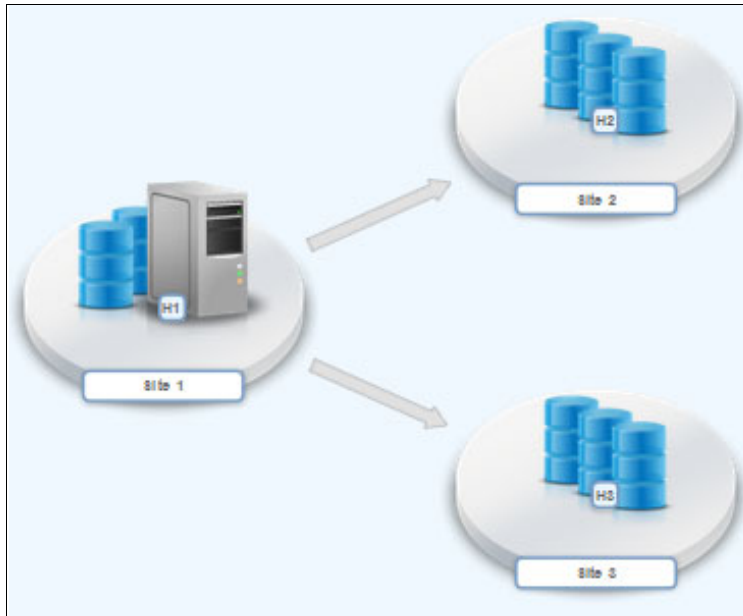

Figure 5-3   Metro Mirror + Global Mirror three-site solution

Table 5-8 on page 94 provides a comparison of the Metro Mirror + Global Mirror replication options that are available for IBM i.

Table 5-9   Metro Mirror + Global Mirror replication options

Type of storage	Supported	IASP	Full system
DS8000	Yes[1]	Yes[1]	Yes (no PowerHA support)
Storwize/SVC	No	No	No
Other storage	No	No	No
[1] Requires IBM Systems Lab Services software for automation. For more information, see Appendix A, "PowerHA Tools for IBM i" on page 141.			

5.5 Backup window reduction solutions

This section contains a brief overview of the point-in-time copy options that are available for IBM i.

5.5.1 FlashCopy

FlashCopy provides a near instantaneous copy of storage. When FlashCopy is combined with the IBM i quiesce function, it provides a reliable copy of data that can be used for backups, queries, or other functions that do not require that the data survives indefinitely.

Table 5-10 provides a comparison of the FlashCopy options that are available for IBM i.

Table 5-10   FlashCopy options

Type of storage	Supported	IASP	Full system
DS8000	Yes	Yes	Yes[1]
Storwize/SVC	Yes	Yes	Yes[1]
Other storage	No	No	No
[1] Requires IBM Systems Lab Services software for automation. For more information, see Appendix A, "PowerHA Tools for IBM i" on page 141.			

5.6 Where do you go next

Now, you have a good idea about the type of solution that you want to implement. If you want to proceed with the implementation of a PowerHA solution, read the following IBM Redbooks publications to help with this process:

► *IBM PowerHA SystemMirror for i: Using DS8000*, SG24-8403
► *IBM PowerHA SystemMirror for i: Using IBM Storwize*, SG24-8402
► *IBM PowerHA SystemMirror for i: Using Geographic Mirroring*, SG24-8401

Alternatively, you can contact your IBM Business Partner or the IBM Systems Lab Services organization to plan and deploy an appropriate availability solution for you. IBM Systems Lab Services can be contacted from the following website:

http://ibm.com/systems/services/labservices

6

Troubleshooting and collecting data for problem determination

This chapter describes methods that help you with troubleshooting and collecting data for problem determination for issues that might arise when you use IBM SystemMirror PowerHA for i.

The following websites are also available to help troubleshoot problems with PowerHA:

- ► IBM Support Portal:

  https://ibm.biz/BdXqvs

- ► IBM i 7.2 Knowledge Center:

  https://ibm.biz/Bd4Vd8

- ► IBM PowerHA SystemMirror for i DeveloperWorks page:

  https://ibm.biz/Bd4VdJ

- ► IBM Full System Copy Services Manager (FSCSM) for PowerHA DeveloperWorks page:

  https://ibm.biz/Bd4vwr

The following topics are described in this chapter:

6.1 Maintaining current PTF levels

It is important to ensure that program temporary fix (PTF) levels on your IBM i system are always current to help minimize the possibility of issues on your system.

6.1.1 Creating a fix strategy

Although the scope of this book cannot go into detail about creating a fix strategy for every environment, the following *Guide to fixes* provides detailed information about this topic:

https://ibm.biz/Bd4VDS

6.1.2 PowerHA group PTF and recommended fixes

The PowerHA group PTF and the high availability (HA) recommended fixes are the primary fixes that you must keep up-to-date on any system that is using PowerHA.

PowerHA group PTF

The PowerHA group PTF is typically updated three times a year, although no schedule is set. This group PTF contains not only fixes for the 5770-HAS product, but also for other related products.

For example, Licensed Internal Code storage management fixes resolve issues with varying on an independent auxiliary storage pool (IASP). These fixes can be for the Licensed Machine Code (5770-999), but they are still part of the group PTF. For the latest group PTF for IBM PowerHA SystemMirror for i, see this website:

https://ibm.biz/Bd4VDm

IBM i Support: Recommended fixes for PowerHA

Because the PowerHA group PTF is updated only a few times a year, when newer fixes become available, these PTFs are listed on the IBM i Support: Recommended Fixes website. From this website, you can view all recommended fixes:

https://ibm.biz/Bd4VDb

To view the current recommended fixes for PowerHA, see the following IBM i Recommended Fixes - High Availability website:

https://ibm.biz/Bd4VDp

By using the IBM MustGather Data Collector tool (QMGTOOLS), you can compare the PowerHA PTFs on the system with the current recommended PowerHA PTFs. This comparison is described in 6.2.3, "Comparing system PowerHA PTFs with available PowerHA PTFs" on page 101.

6.2 Using QMGTOOLS for diagnostic data collection

This section describes a tool that is called the IBM MustGather Data Collector (QMGTOOLS), which can be used to collect diagnostic data for PowerHA to send to IBM.

The problem data collection that needs to be sent to IBM for PowerHA issues is most easily gathered by using QMGTOOLS. QMGTOOLS has a wider scope than simply PowerHA. It can be used to collect data for database, communications, performance, and many other issues. This tool evolves continually to make it easier to collect the data that IBM Support requires to resolve issues across this wide range of areas. QMGTOOLS was originally designed to collect diagnostic data for PowerHA issues and it is still the best tool for this purpose. This tool collects all of the data that is typically needed for IBM Support and development teams to diagnose an issue.

The following QMGTOOLS topics are described in this section:

► 6.2.1, "Downloading and installing QMGTOOLS" on page 99
► 6.2.2, "Verifying the build date of QMGTOOLS" on page 99
► 6.2.3, "Comparing system PowerHA PTFs with available PowerHA PTFs" on page 101
► 6.2.4, "Collecting diagnostic data with QMGTOOLS" on page 103
► 6.2.5, "Development of QMGTOOLS" on page 109

6.2.1 Downloading and installing QMGTOOLS

Before you use QMGTOOLS to collect PowerHA diagnostic data, QMGTOOLS must be downloaded and installed on each node in the cluster. Several methods are available. The following *MustGather: How To Obtain and Install QMGTOOLS* document provides the steps to download and install these tools:

http://www.ibm.com/support/docview.wss?uid=nas8N1011297

6.2.2 Verifying the build date of QMGTOOLS

Before you collect the data for PowerHA for any issue, verify to determine whether QMGTOOLS is at the latest build date. Because updates are continually added to QMGTOOLS, you need to keep it up-to-date on all nodes in the cluster.

Check whether QMGTOOLS is at the latest build date by following these steps:

1. All of the QMGTOOLS commands reside in the QMGTOOLS library. From a 5250 session on your system, add that library to the library list for this session by using the following Add Library List Entry (`ADDLIBLE`) command:

   ADDLIBLE QMGTOOLS

2. Display the main menu for QMGTOOLS by running the following command:

   GO MG

The MustGather Data Collector menu is displayed as shown in Figure 6-1.

```
 MG                      Must Gather Data Collector
                    (C) COPYRIGHT IBM CORP. 2009, 2012
 Select one of the following:


      1. HA (High Availability)         14. External Storage
      2. Performance/Misc collection    15. Work Management
      3. Client/Server                  16. Internals
      4. Communications menu            17.
      5. Database menu                  18.
      6. CTA/EWS (JAVA/HTTP/DCM/WAS)     19. Hardware data collection
      7. Save/Restore menu              20. HMC menu
      8.
      9. Misc tools
     10. FTP data to IBM
     11. View FTP to IBM statuses
     12. Display build date
     13. Check IBM for updated QMGTOOLS


 Selection or command
 ===>


 F3=Exit    F4=Prompt    F9=Retrieve    F12=Cancel
 F13=Information Assistant  F16=System main menu
```

Figure 6-1   QMGTOOLS main menu

3. Enter option 12 (Display build date) from the QMGTOOLS main menu and press Enter to display the build date as shown in Figure 6-2.

```
 12. Display build date
 13. Check IBM for updated QMGTOOLS


  Selection or command
  ===>


  F3=Exit    F4=Prompt    F9=Retrieve    F12=Cancel
  F13=Information Assistant  F16=System main menu
  Build date and version : 09/15/15 V7R2M0
```

Figure 6-2   Displaying the current QMGTOOLS build date

4. Enter option 13 (Check IBM for updated QMGTOOLS) from the QMGTOOLs main menu and press Enter to check for an updated QMGTOOLS. This option connects through File Transfer Protocol (FTP) to the IBM site where the QMGTOOLS tool is hosted. It checks the build date of that tool and compares it with the build date of the QMGTOOLS tool on the local system. If the build on the IBM site is newer than the build on the IBM i system, a prompt is displayed to download the current build as shown Figure 6-3.

```
   Local build date  : 091515
   Current build date: 102715

   Build date mismatch, would you like to download current version?
   If so, enter library to store savefile, press F1. Else Exit (F3)

   Library to store save file :  QGPL

 F1=Continue  F3=Exit
```

Figure 6-3   Checking for updated QMGTOOLS

5. Enter a library to store the save file and press F1 to download the latest QMGTOOLS build.

6. After the save file is downloaded, restore the save file by running the following Restore Library (**RSTLIB**) command where QGPL is the library where the QMGTOOLs save file was downloaded:

RSTLIB SAVLIB(QMGTOOLS) DEV(*SAVF) SAVF(QGPL/QMGTOOL720)

6.2.3 Comparing system PowerHA PTFs with available PowerHA PTFs

QMGTOOLS provides a way to compare the level of PowerHA PTFs on your system with the level of PowerHA PTFs that are available on the IBM i Recommended Fixes website for PowerHA ("IBM i Support: Recommended fixes for PowerHA" on page 98).

To compare your level of PowerHA PTFs, follow these steps:

1. From a 5250 session on your system, add the QMGTOOLS library to the current session library list by using the following Add Library List Entry (**ADDLIBLE**) command:

ADDLIBLE QMGTOOLS

2. Display the main menu for QMGTOOLS by running the following command:

GO MG

The MustGather Data Collector menu is displayed as shown in Figure 6-1 on page 100.

3. Enter option 1 (HA (High Availability)) from the QMGTOOLs main menu and press Enter to display the QHASTOOLS menu.

4. On the QHASTOOLS menu (Figure 6-4), enter option 12 (Compare HA PTFs from IBM public FTP site) and press Enter to compare the PowerHA PTFs on your system with the current level that is available from the IBM public FTP site.

A connection is made to IBM to perform the comparison. If this comparison is successful, the following message is displayed:

`Complete. Check QSYSPRT for results.`

Press Enter to go back to the IBM i command line.

```
 HASMNU                          QHASTOOLS menu


 Select one of the following:

      1. Collect and retrieve cluster data from multiple nodes
      2. Dump cluster data on local node only
      3.
      4.
      5. Cluster Debug Tool
      6.
      7. Dump SST macros XSM/Cluster
      8. Collect GUI data
      9. Node status trap
     10.
     11. Additional XSM tools
     12. Compare HA PTFs from IBM public FTP site

     14. Install/Uninstall DSCLI SSLv3 fix
 Selection or command
 ===>


 F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel
 F13=Information Assistant  F16=System main menu
```

Figure 6-4   QHASTOOLS menu

5. Run the Work with Job (`WRKJOB`) command to find the QSYSPRT spooled file and enter option 4 to work with spooled files.

6. By displaying the QSYSPRT spooled file, you can see the list of PTFs that are missing from the system as compared to the IBM i Recommended Fixes website for PowerHA.

7. See the example that is shown in Figure 6-5.

   This tool helps to provide information about how current this node is in comparison to the available fixes.

```
                                Display Spooled File
  File . . . . . :    QSYSPRT
  Control . . . . .
  Find . . . . . .

 *...+....1....+....2....+....3....+....4....+....5....+....6....+....7....+.
  System CTCIHA4A
  PTFs missing :
  -----------------
  SI57154  Not found
  SI57016  Not found
  SI56917  Not found
  SI56648  Not found
  SI56614  Not found
  SI56562  Not found
  SI56561  Not found
  SI56541  Not found
  SI55982  Not found
  SI55886  Not found
  SI55603  Not found
  SI55306  Not found
  SI55303  Not found
  SI55031  Not found
  SI55011  Not found
  MF60243  Not found
                                                                    More...

  F3=Exit    F12=Cancel    F19=Left    F20=Right    F24=More keys
```

Figure 6-5   QMGTOOLS PTF comparison output

6.2.4 Collecting diagnostic data with QMGTOOLS

After QMGTOOL is downloaded and updated on all nodes in the cluster, it is a simple process to collect the problem determination data. Two options can be used to collect the QMGTOOLS data:

► The data can be collected across all nodes at one time. For more information, see "Option 1: Gathering QMGTOOLS data from all cluster nodes at one time" on page 104.

► The data can be collected separately for each node. For more information, see "Option 2: Gathering QMGTOOLS data from each cluster node" on page 107.

In addition, this section describes data collection data when problems arise when you use IBM Navigator for i PowerHA graphical user interface (GUI). For more information, see "Collecting diagnostic data for the PowerHA graphical user interface" on page 108.

Tip: IBM Support might instruct you to collect additional data that is not included in these instructions. However, if a PowerHA issue is encountered, it is a good starting point to gather QMGTOOLS data and upload it when you create a problem management record (PMR) with IBM Support.

The following list is a partial listing of the data that is gathered by QMGTOOLS for PowerHA issues. This list is not an all-inclusive list:

► System cluster and cluster resource group (CRG) job logs and dumps
► Admin domain job logs and dumps
► Call stacks for currently running cluster-related jobs
► Display Cluster Info (`DSPCLUINF`) output
► Display CRG Info (`DSPCRGINF`) output
► Cross-site mirroring trace and log (`XSMTRACE` and `XSMLOG`) output
► Dump Information (`DMPINF`) for IBM IASP Copy Services Manager (ICSM) Toolkit output
► System value settings
► TCP/IP configuration settings
► Job queue listing
► Work with Disk Status (`WRKDSKSTS`) output
► Recent product activity log (PAL) entries
► Recent vertical Licensed Internal Code logs (VLOGS)
► History log
► Service documents (Docs) output

Option 1: Gathering QMGTOOLS data from all cluster nodes at one time

When you gather QMGTOOLS, IBM Support usually requires that data from all nodes is collected and uploaded. By using this option, all of this data can be gathered from a single node in the cluster. This method uses the File Transfer Protocol (FTP) to submit remote commands to other nodes and also to transfer data back to the node where these steps are run. The following steps can be performed from any node in the cluster:

1. From a 5250 session on your system, add the QMGTOOLS library to the current session library list by using the following Add Library List Entry (`ADDLIBLE`) command:

   ADDLIBLE QMGTOOLS

2. Display the main menu for QMGTOOLS by running the following command:

   GO MG

   The MustGather Data Collector menu is displayed as shown in Figure 6-1 on page 100.

3. Enter option 1 (HA (High Availability)) from the QMGTOOLs main menu and press Enter to display the QHASTOOLS menu.

4. On the QHASTOOLS menu (Figure 6-4 on page 102), enter option 1 (Collect and retrieve cluster data from multiple nodes) and press Enter.

5. From the DMPCLU window (Figure 6-6), accept the defaults and press Enter. This option is typically needed by IBM Support. This option dumps the data to the QTILIB library.

```
                          (DMPCLU)


 Type choices, press Enter.

 Library to store data  . . . . .   QTILIB        Library Name
 Dump cluster trace only  . . . .   N             Y, N
 Collect system snapshot  . . . .   Y             Y, N
 Data compression during save . .   *YES          *YES, *NO, *LOW, *MEDIUM...


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

Figure 6-6   DMPCLU command prompt

6. If the same user profile exists on all nodes in the list, press F6 on the panel that is shown in Figure 6-7 so that you need to enter the user profile name and password a single time. Otherwise, specify a user ID (user profile), password, and confirm the password for each node in the cluster and press F1 to continue to step 8. We suggest that you use QSECOFR or an equivalent user for this user ID.

```
  Note:  Only nodes on this page will be processed. If more nodes exist,
           process this page first. Go back and process the next pages.

     Nodes          UserID        Password        Confirm Password
     ----------     ----------    --------------  ----------------

     CTCIHA4A
     CTCIHA4D
     CTCIHA4C
     CTCIHA4B


 F1=Continue  F3=Exit  F6=Options
```

Figure 6-7   Entering sign-on information to authenticate to each node in the cluster

7. If you press F6 (Options) to specify the same user profile for all nodes in the cluster, the Advanced Options panel that is shown in Figure 6-8 is displayed. Specify a `Y` for the "Use same user/pass for all nodes" option and enter the User ID, Password, and Confirm Password values. Press F1 to continue.

```
                         Advanced Options

     Use same user/pass for all nodes . . . . . . . . . . . .: Y
     User ID : POWERHAUSR   Password :
                            Confirm  :

 F1=Continue  F3=Exit
```

Figure 6-8   Specifying the same user ID and password for all nodes

8. The previous window is displayed with the completed user profiles and passwords that were specified for all nodes (Figure 6-9). Press F1 to begin the collection of the diagnostic data.

```
  Note:  Only nodes on this page will be processed. If more nodes exist,
           process this page first. Go back and process the next pages.

     Nodes          UserID        Password        Confirm Password
     ----------     ----------    --------------  ----------------

     CTCIHA4A       POWERHAUSR
     CTCIHA4D       POWERHAUSR
     CTCIHA4C       POWERHAUSR
     CTCIHA4B       POWERHAUSR

 F1=Continue  F3=Exit  F6=Options
```

Figure 6-9   Completed user IDs and passwords for all nodes

9. After the jobs are submitted to all nodes, QMGTOOLS gathers its data from each node as shown in Figure 6-10.

```
      Nodes         Status
      -----         ------

      CTCIHA4A      Waiting...   Grabbing additional joblog
      CTCIHA4D      Waiting...   Grabbing additional joblog
      CTCIHA4C      Waiting...   Grabbing additional joblog
      CTCIHA4B      Waiting...   Grabbing additional joblog

   0000000004
```

Figure 6-10   QMGTOOLS gathers initial data from all nodes in the cluster

10. QMGTOOLS continues through 14 steps of gathering various data from each node in the cluster. Figure 6-11 shows one of the nodes on step 11 of 14, and the other three nodes on step 12 of 14.

```
      Nodes         Status
      -----         ------

      CTCIHA4A      Waiting...   Getting network attributes (11/14)
      CTCIHA4D      Waiting...   Getting System Snapshot (12/14)
      CTCIHA4C      Waiting...   Getting System Snapshot (12/14)
      CTCIHA4B      Waiting...   Getting System Snapshot (12/14)

   0000000008
```

Figure 6-11   QMGTOOLS gathers data for steps 11 and 12 of 14

11. After QMGTOOLS collects the data on all nodes and transfers back all of that data from each node to the node that is running QMGTOOLS, the data is ready to save to a save file as shown in Figure 6-12. Press F1 to create the save file.

```
      Nodes         Status
      -----         ------

      CTCIHA4A      Done         Local system collection done...
      CTCIHA4D      Done         Done retrieving...
      CTCIHA4C      Done         Done retrieving...
      CTCIHA4B      Done         Done retrieving...

   Data collected into library QTILIB
   FTP done, press F1 to put into savefile or F3 to exit
```

Figure 6-12   QMGTOOLS data collection complete for all cluster nodes

12. After you save the data to a save file, a message is displayed at the bottom of the session as shown in Figure 6-13.

The resulting file is called CLUDOCS*xxx*, where *xxx* is a sequential, unique number that is assigned to the current collection. The CLUDOCS*xxx* save file contains data from all nodes that were specified in this set of steps. This data must be uploaded to IBM and attached to the problem management record (PMR) when you initially contact IBM about the issue that you encountered. By including this data when you open a PMR, you help to expedite the problem resolution for the issue that was captured.

```
Selection or command
===>


F3=Exit    F4=Prompt    F9=Retrieve    F12=Cancel
F13=Information Assistant  F16=System main menu
Data saved into save file CLUDOCS001 in library QTILIB
```

Figure 6-13   QMGTOOLS data that is saved to a save file

Option 2: Gathering QMGTOOLS data from each cluster node

This option is more time-consuming than "Option 1: Gathering QMGTOOLS data from all cluster nodes at one time" on page 104 because this option requires that the steps in this section are followed for each node in the cluster. Therefore, it is best to use option 1 if possible.

However, certain circumstances can prevent you from using option 1. For example, network or system configurations can prevent the FTP connections from being established to one or more of the nodes in the cluster. Therefore, the data collection is prevented.

Follow these steps to collect this data from each node:

Important: You must perform the steps in this procedure on each node in the cluster.

1. From a 5250 session on your system, add the QMGTOOLS library to the current session library list by using the following Add Library List Entry (**ADDLIBLE**) command:

   ADDLIBLE QMGTOOLS

2. Display the main menu for QMGTOOLS by running the following command:

   GO MG

   The MustGather Data Collector menu is displayed as shown in Figure 6-1 on page 100.

3. Enter option 1 (HA (High Availability)) from the QMGTOOLs main menu and press Enter to display the QHASTOOLS menu.

4. On the QHASTOOLS menu (Figure 6-4 on page 102), enter option 2 (Dump cluster data on local node only) and press Enter.

5. From the DMPCLUINF window (Figure 6-14), accept the defaults and press Enter. This option is typically needed by IBM Support. This option dumps the data to the QTILIB library.

```
                             (DMPCLUINF)

 Type choices, press Enter.

 Cluster Name . . . . . . . . . . > CLUSTER      Character value
 Local node name  . . . . . . . . > CTCIHA4A     Character value
 Save into save file? . . . . . .   Y            Y, N
    Data compression  . . . . . .   *YES         *YES, *NO, *LOW, *MEDIUM...
 Library to store data  . . . . .   QTILIB       Character value
 Collect system snapshot  . . . .   Y            Y, N


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

Figure 6-14   DMPCLUINF command prompt

6. Status messages are shown at the bottom of the window as the tool progresses with its data collection. After the data collection is complete, the message at the bottom of the window as shown in Figure 6-15 is displayed.

```
 Completed...save file CLUDOCS002 created.
```

Figure 6-15   QMGTOOLS collection completed for single node message

7. Repeat these steps for all nodes in the cluster.

8. After you repeat these steps for all nodes in the cluster, a CLUDOCS*xxx* save file exists in the QTILIB library on each node in the cluster. You must upload these save files to IBM and attach them to the problem management record (PMR) when you initially contact IBM about the issue you encountered. Including this data when you open a PMR helps to expedite the problem resolution for the issue that was captured.

Collecting diagnostic data for the PowerHA graphical user interface

This section describes how to use QMGTOOLS to collect data for problems that arise when you use the IBM Systems Director Navigator for i for PowerHA configuration and management. Perform the following steps after you re-create the PowerHA problem within IBM Systems Director Navigator for i:

1. From a 5250 session on your system, add the QMGTOOLS library to the current session library list by using the following Add Library List Entry (**ADDLIBLE**) command:

   ADDLIBLE QMGTOOLS

2. Display the main menu for QMGTOOLS by running the following command:

   GO MG

   The MustGather Data Collector menu is displayed as shown in Figure 6-1 on page 100.

3. Enter option 1 (HA (High Availability)) from the QMGTOOLs main menu and press Enter to display the QHASTOOLS menu.

4. On the QHASTOOLS menu (Figure 6-4 on page 102), enter option 8 (Collect GUI data) and press Enter.

5. The tool dumps the data to a compressed file in the integrated file system (IFS). After it completes, a message, such as the message that is shown in Figure 6-16, is displayed at the bottom of the window.

   This data typically contains the debug data that IBM Support needs to diagnose the problem. If a PMR is opened to IBM Support, this data must be collected and uploaded at that time.

```
Completed, file is in IFS directory /tmp/hasmlogs0916151625.zip
```

Figure 6-16   QMGTOOLS GUI data collection completed message

6.2.5 Development of QMGTOOLS

QMGTOOLS is continuously improved and new functions are added constantly. Therefore, it is important that the current version of the tool is installed on all nodes in all clusters as described in 6.2.2, "Verifying the build date of QMGTOOLS" on page 99.

If you have any suggestions that might help improve this tool, you can contact Benjamin Rabe (brabe@us.ibm.com) or David Granum (granum@us.ibm.com) so that your suggestions can be forwarded to the correct development team.

6.3 Methods for troubleshooting

This section provides general methods that help when you troubleshoot geographic mirroring and describes common issues that can occur with geographic mirroring and how to resolve those issues.

The following topics are described in this section:
- ▶ 6.3.1, "Ensuring that job descriptions have appropriate logging levels" on page 109
- ▶ 6.3.2, "Using the command-line interface when you re-create problems" on page 110

6.3.1 Ensuring that job descriptions have appropriate logging levels

Several user profiles are used to perform various cluster and PowerHA actions. It is important when you are troubleshooting to ensure that the job descriptions that are used for these user profiles are configured with the appropriate logging levels. The user profiles and associated job descriptions are listed in Table 6-1.

Table 6-1   PowerHA user profiles and associated job descriptions

User profile	Job description
QSYS	QGPL/QDFTJOBD
QCLUSTER	QSYS/QCSTJOBD
QHAUSRPRF	QHASM/QHAJOBD
Current interactive user	Can be identified by running `WRKJOB`, option 2.

To check each of these job descriptions, use the Display Job Description (`DSPJOBD`) command. The message logging value for each job description must be set in the following manner:

- ▶ Message logging level: `4`
- ▶ Message logging severity: `00`
- ▶ Message logging text: `*SECLVL`

If the LOG values of the job descriptions in Table 6-1 on page 109 are not set to these values, use the Change Job Description (`CHGJOBD`) command to change them.

> **Tip:** If the job description changes, the active jobs that use that job description do not pick up the new values until the job is restarted.

6.3.2 Using the command-line interface when you re-create problems

When you re-create problems, it is best to use command-line (CL) commands rather than menu options for better logging and easier analysis of the job log where the action is taken.

For example, if an issue occurs when you try to work with ASP copy descriptions, it is best to use the Work with Auxiliary Storage Pool Copy Descriptions (`WRKASPCPYD`) command rather than to use a Work with Cluster (`WRKCLU`) command and option 10 from the Work with Cluster menu. The `WRKASPCPYD` command provides better debug information to the user and to IBM Support, if necessary.

6.4 Troubleshooting common cluster architecture

This section describes common situations that you might encounter when you use geographic mirroring and how to resolve them. It is always best to stay at the current PTF levels to prevent the PowerHA environment from experiencing these issues. The information that is provided for each example in this section assumes that the nodes in the cluster are at the current PTF levels.

The following troubleshooting situations are described in this section:

- ▶ 6.4.1, "Cluster node does not start" on page 110
- ▶ 6.4.2, "Cluster node does not end" on page 111
- ▶ 6.4.3, "Cluster node in partition status" on page 112
- ▶ 6.4.4, "Cluster node in unknown status" on page 112
- ▶ 6.4.5, "IASP cannot be varied on on the current production system" on page 113
- ▶ 6.4.6, "Unexpectedly long vary-on time for IASP after the failover" on page 113
- ▶ 6.4.7, "MREs in cluster admin domain with an inconsistent global status" on page 114
- ▶ 6.4.8, "RMVCADMRE fails to remove monitored resource entry" on page 115
- ▶ 6.4.9, "User profile MRE in cluster admin domain with a failed status" on page 115
- ▶ 6.4.10, "Cannot access PowerHA graphical user interface" on page 116
- ▶ 6.4.11, "Checking the PowerHA status by using the PowerHA GUI" on page 118

6.4.1 Cluster node does not start

One of the most common issues in a PowerHA environment is a cluster node that does not activate. This situation can occur for several reasons. The most common cause is an inability to communicate between the nodes.

However, the following steps need to be run:

1. Ensure that the TCP/IP interfaces that are used for clustering and the loopback interface are active. The loopback interface is 127.0.0.1. These interfaces can be verified by using the following Work with Network (**NETSTAT**) status command:

   `NETSTAT *IFC`

   If any of these TCP/IP interfaces are inactive, they must be started for clustering to start and stay active. From the NETSTAT *IFC window, specify option 9 next to the correct interface. After the interface is activated, attempt to start the cluster node again.

2. If the TCP/IP interfaces are active, ensure that each cluster interface on each node can PING each cluster interface on the other nodes. Use the Packet INternet Groper (**PING**) command to test this connection from each node:

   `PING RMTSYS('B.B.B.B') LCLINTNETA('A.A.A.A')`

   B.B.B.B is another node in the cluster, and *A.A.A.A* is a local TCP/IP interface that is configured for the local cluster node.

3. Ensure that the InterNET Daemon server is active. To verify that the InterNET Daemon server is active, use the following Work with Active Jobs (**WRKACTJOB**) command:

   `WRKACTJOB JOB(QTOGINTD)`

   If the `QTOGINTD` job is not active, start it by using the following command:

   `STRTCPSVR *INETD`

   Verify that the QTOGINTD job is active. If it is now active, attempt to run the Start Cluster Node (**STRCLUNOD**) command again. If it starts, you do not need to proceed further through these instructions.

4. On the cluster node that is inactive, use the following Work with Active Jobs command (**WRKACTJOB**) command to ensure that the QCSTCTL (cluster control) job is inactive:

   `WRKACTJOB JOB(*SYS)`

5. If the local node is inactive, the QCSTCTL job on that node must not be active. If the QCSTCTL job is active, QMGTOOLS must be collected and uploaded to IBM when you create a PMR with IBM Support. (For more information about QMGTOOLS, see 6.2.4, "Collecting diagnostic data with QMGTOOLS" on page 103.) After that data is collected, you can perform more steps to alleviate this condition.

6. End the QCSTCTL job on this node by using the following application programming interface (API):

   `CALL PGM(QWCCTLSJ) PARM('*END'  'QCSTCTL' 'QSYS' 'nnnnnn')`

   The *nnnnnn* is the job number from the active QCSTCTL job.

7. End the other cluster nodes locally from each system by using the End Cluster Node (**ENDCLUNOD**) command. After you end each node, ensure that the QCSTCTL job is no longer active. If the job is still active, use the API from step 6 to end the job.

8. With all cluster nodes ended and no active QCSTCTL job, start each of the cluster nodes one by one, starting each node locally from itself by using the **STRCLUNOD** command.

6.4.2 Cluster node does not end

Due to timing issues, cluster nodes can get into a condition in which they do not end. This situation can result in an error message on an End Cluster Node (**ENDCLUNOD**) command or a situation in which the interactive session appears to hang.

In this case, the QMGTOOLS data collection that is described in 6.2, "Using QMGTOOLS for diagnostic data collection" on page 99 must be collected *while the system is in this condition*. IBM Support must be contacted and supplied with the QMGTOOLS output. After this data is collected, use the following steps to clear the condition:

1. From the node that does not end, use the following Work with Active Jobs (`WRKACTJOB`) command to obtain the details of the QCSTCTL (cluster control) job:

   `WRKACTJOB JOB(*SYS)`

2. End the QCSTCTL job on this node by using the following API call:

   `CALL PGM(QWCCTLSJ) PARM('*END' 'QCSTCTL' 'QSYS' 'nnnnnn')`

   The *nnnnnn* is the job number from the active QCSTCTL job.

3. End the other cluster nodes locally from each system by using the `ENDCLUNOD` command. After you end each node, ensure that the QCSTCTL job is no longer active. If the job is active, use the API from step 2.

With all cluster nodes ended and no QCSTCTL job active, start each of the cluster nodes one by one, starting each node locally by using the `STRCLUNOD` command.

6.4.3 Cluster node in partition status

Clustering uses heartbeat messaging, which uses the User Datagram Protocol (UDP) over port 5551 to ensure that each of the nodes can communicate with one another continually. A cluster node is marked in *partition* status if the local node is unable to communicate with that node. If this issue is a momentary drop in communications, troubleshooting this condition needs to focus on ensuring that each interface that is used for clustering on each of the nodes can communicate with each other.

One way to check for basic communications is to use the `PING` command:

`PING RMTSYS('B.B.B.B') LCLINTNETA('A.A.A.A')`

B.B.B.B is another node in the cluster, and *A.A.A.A* is a local TCP/IP interface that is configured for the local cluster node.

Ping uses the Internet Control Message Protocol (ICMP), and not UDP, so it is not a perfect test. Therefore, if `PING` is successful, and the cluster node remains in partition status, the underlying communications problem still exists.

The tool that can be used on an IBM i system to troubleshoot communications issues is the Trace Connection (`TRCCNN`) command. It is beyond the scope of this book to get into the details of tracing and analyzing communications problems. In addition to the IBM i tracing tools, it might be necessary to troubleshoot from a network perspective. This approach might include troubleshooting switches and routers and other network infrastructure, and analyzing network traffic over this infrastructure.

6.4.4 Cluster node in unknown status

A non-local cluster node in an *unknown* status occurs because the local cluster node is in an inactive state. Because it is inactive, the local cluster node cannot determine the status of any other nodes in the cluster. If the local cluster node is active, collect the QMGTOOLS data as described in 6.2.4, "Collecting diagnostic data with QMGTOOLS" on page 103. Otherwise, if the local cluster node is active as expected, you need to start the cluster node by using the `STRCLUNOD` command.

6.4.5 IASP cannot be varied on on the current production system

If a system failure occurred on your current production system and you recovered from it without failing over, you might run into a situation where you cannot seem to vary on the IASP on any of your systems in the cluster.

The reason for this situation is possibly that a failover message queue is defined either on a cluster or a cluster resource group (CRG) level and you did not answer that message yet. If you want to continue running on the formerly failed production system, answer that message with a `cancel`. If you want to fail over to the backup system, answer the message with `proceed`.

Ensure that you answer that message only when you decide how you want to proceed. Recovering from the wrong answer requires additional work.

6.4.6 Unexpectedly long vary-on time for IASP after the failover

This section provides information to check on your system if you are experiencing an unexpectedly long vary-on time of an IASP.

> **Note:** Ensure that you are on the current PTF levels. For more information, see 6.2.3, "Comparing system PowerHA PTFs with available PowerHA PTFs" on page 101.

The following Display ASP Status (`DSPASPSTS`) command supports looking at historical data:

```
DSPASPSTS ASPDEV(<iaspname>)
```

A value of ENTRIES(*ON) shows the results of all vary-on requests, beginning with the most recent and proceeding backwards in time. Compare how long the different steps take to identify where the most time is spent:

► Increase in UID/GID mismatch correction

   Ensure that user profiles are correctly synchronized between the cluster nodes.

► Increase in database cross-reference merge

   The time that is needed for database cross-reference merge mainly depends on the number of database objects in SYSBASE. If you experience a sudden increase, ensure that you understand why the number of database objects grew in the system ASP.

► Database access path recovery

   Check your system-managed access-path protection (SMAPP) settings for the IASP by using the Edit Recovery for Access Paths (`EDTRCYAP`) command. Check these values for estimated access path recovery in comparison to the target time. Also, check whether any files have uneligible access paths and understand why. Look at unprotected access paths and their estimated recovery time.

► Database/commit recovery

   Check whether any changes in your application led to longer commit cycles or whether specific batch programs with long running commit cycles were active at the point of the system failure.

6.4.7 MREs in cluster admin domain with an inconsistent global status

At times, managed resource entries (MREs) in an admin domain can show a status of *Inconsistent*. Typically, this status occurs because the admin domain is unable to update the MRE on all nodes in the admin domain. Use the Work with Monitored Resources (`WRKCADMRE`) command to display the status of the MREs. The panel is similar to Figure 6-17.

```
                     Work with Monitored Resources
                                                      System:   RCHESP2
 Administrative domain  . . . . . . . . . :    ESPADMN
 Consistent information in cluster  . . . :    Yes
 Domain status  . . . . . . . . . . . . . :    Active

 Type options, press Enter.
   1=Add    4=Remove    5=Display details    6=Print    7=Display attributes
                        Resource                   Global
 Opt     Resource       Type           Library     Status

         PWRHAUSR       *USRPRF        QSYS        Inconsistent
         GRANUM         *USRPRF        QSYS        Consistent
         QWEBQRYJOB     *JOBD          QWEBQRY     Consistent
         QWEBQRY21      *JOBD          QWEBQRY     Consistent
         TESTOUT        *PRTDEV        QSYS        Consistent
         QMAXSIGN       *SYSVAL                    Consistent
         QRETSVRSEC     *SYSVAL                    Consistent
                                                                More...
 Parameters for option 1 or command
 ===>
 F1=Help    F3=Exit    F4=Prompt    F5=Refresh    F9=Retrieve
 F11=Order by type and name        F12=Cancel     F24=More Keys
```

Figure 6-17   WRKCADMRE that shows an inconsistent MRE

Often, when an MRE is in an inconsistent state, it is a user profile. The example in Figure 6-17 shows this issue. To troubleshoot this situation, the best place to look is the job log of the admin domain job. This job has the same name as the cluster admin domain name.

By using the example in Figure 6-17, the job log for the actively running ESPADMN job contains messages that indicate why the MRE is inconsistent. This situation can occur for many reasons, but in this example, the issue was caused by user profile PWRHAUSR recently being added to the admin domain from RCHESP1. This user profile exists on both nodes, RCHESP1 and RCHESP2. However, the User ID number (UID) value for both user profiles is not the same on both nodes. The cluster admin domain attempts to change the UID value on RCHESP2 for PWRHAUSR to the same value that exists for the UID value on RCHESP1 for PWRHAUSR.

This attempt causes an error because the UID value that is used for PWRHAUSR on RCHESP1 is already used by another user profile on RCHESP2. Therefore, the change cannot be made by the cluster admin domain. The solution for this common situation is to manually change to an available UID value by using the Change User Profile (`CHGUSRPRF`) command.

6.4.8 RMVCADMRE fails to remove monitored resource entry

The removal of an MRE might fail when you issue a Remove Cluster Admin Domain
Monitored Resource Entry (`RMVCADMRE`) command. When you encounter this issue, it is a
defective situation. You need to open a PMR to the IBM i Support team.

> **Important:** Ensure that when you re-create this problem, the actual `RMVCADMRE` command
> is used and not the menu options. For more information, see 6.3.2, "Using the
> command-line interface when you re-create problems" on page 110.

After you re-create the problem, collect the problem determination data by using QMGTOOLS
as described in 6.2.4, "Collecting diagnostic data with QMGTOOLS" on page 103.

After you collect the problem determination data, follow these steps to circumvent this
situation:

1. End clustering on each node in the cluster.

2. Run the following program on one of the cluster nodes:

   ```
   CALL PGM(QSYS/QCSTADFIX)
   ```

3. Start clustering on the node where the program (step 2) is called.

4. Start clustering for the other cluster nodes.

The QSYS/QCSTADFIX program rectifies inconsistencies in the number of entries that are
used in an internal index by the admin domain. By collecting QMGTOOLS data before you run
the fix program, you help IBM Support to identify the root cause of the issue.

6.4.9 User profile MRE in cluster admin domain with a failed status

One common situation that is sometimes seen with the cluster admin domain is an issue with
user profiles or other objects that show a global status of *Failed* when you run a Work with
Monitored Resources (`WRKCADMRE`) command as shown in Figure 6-18 on page 116.

This common situation is almost always caused by an object that is removed from the system
without being removed from the cluster admin domain. However, it is always best to check the
admin domain job log for other potential issues that might cause the MRE to be in this global
status.

However, in this case, user ADMHA was an MRE in the admin domain with a consistent
global status that was deleted from the system. As part of the process of deleting that object
from the system, no cleanup is performed within the admin domain. Therefore, this MRE has
to be manually removed from the WRKCADMRE panel.

Figure 6-18 shows an MRE with a failed status.

```
                    Work with Monitored Resources
                                                        System:    RCHESP1
Administrative domain  . . . . . . . . . :    ESPADMN
Consistent information in cluster  . . . :    Yes
Domain status  . . . . . . . . . . . . . :    Active

Type options, press Enter.
  1=Add    4=Remove    5=Display details    6=Print    7=Display attributes
                        Resource                      Global
Opt      Resource       Type            Library       Status

         ADMHA          *USRPRF         QSYS          Failed
         GRANUM         *USRPRF         QSYS          Inconsistent
         QWEBQRYJOB     *JOBD           QWEBQRY       Consistent
         QWEBQRY21      *JOBD           QWEBQRY       Consistent
         TESTOUT        *PRTDEV         QSYS          Consistent
         QMAXSIGN       *SYSVAL                       Consistent
         QRETSVRSEC     *SYSVAL                       Consistent
                                                                 More...
Parameters for option 1 or command
===>
F1=Help   F3=Exit   F4=Prompt   F5=Refresh   F9=Retrieve
F11=Order by type and name      F12=Cancel   F24=More Keys
```

Figure 6-18   WRKCADMRE that shows a failed MRE

6.4.10 Cannot access PowerHA graphical user interface

To access the IBM Navigator for i PowerHA GUI, point your web browser to the following URL:

http://*<hostname_or_IP_address>*:2001

Several reasons exist why the PowerHA GUI cannot be accessed. One of the most common causes for this situation is a problem with communications between the workstation and the IBM i system that is being connected to. Check and resolve the following situations to access the PowerHA GUI:

► If you are connecting to the host name of the IBM i system, check the following areas:

  – Ensure that the workstation can resolve that host name by performing a **PING** from the workstation to that host name. If this **PING** is also unsuccessful, add the IBM i host name to the workstation host table or ensure that the Domain Name System (DNS) services are configured correctly on the workstation.

  – Attempt to connect to the IP address of the IBM i system rather than the host name. If this connection is successful, an issue exists with the workstation DNS or name resolution configuration.

► Attempt to connect to the IBM i from another workstation. If this effort is successful, a communications configuration problem likely exists on the initial workstation.

► If multiple workstations fail to connect to the IBM i system, the network infrastructure possibly is preventing the TCP/IP traffic from transmitting between the workstation and the IBM i system.

► On the IBM i system, verify that the Admin HyperText Transfer Protocol (HTTP) server is listening on port 2001, by running the following steps:

a. On the command line of the IBM i system, list the network connections by using the following Work with Network Connection Status (**NETSTAT**) command:

   NETSTAT *CNN

b. On the Work with IPv4 Connection Status display, press F14 (Display port numbers) to convert the port names to port numbers.

c. Ensure that port 2001 shows a status of *Listen* as shown in Figure 6-19.

```
                    Work with IPv4 Connection Status
                                                  System:    RCHESP2
 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
      *                 *      447  011:22:45  Listen
      *                 *      448  011:31:20  Listen
      *                 *      449  000:08:34  Listen
      *                 *      515  011:22:47  Listen
      *                 *      657  011:22:47  Listen
      *                 *      657  000:00:00  *UDP
      *                 *      992  011:31:50  Listen
      *                 *     2001  000:28:55  Listen
      *                 *     2002  006:03:42  Listen
      *                 *     2004  000:28:56  Listen
      *                 *     2005  000:00:03  Listen
      *                 *     2006  006:03:35  Listen


 More...
  F3=Exit    F5=Refresh   F9=Command line   F11=Display byte counts
 F12=Cancel F15=Subset    F20=Work with IPv6 connections   F24=More keys
```

*Figure 6-19   NETSTAT *CNN that shows port 2001 in Listen status*

d. If port 2001 is not in a Listen status, you must restart the HTTP server for the Admin instance by following these steps:

   i. Run the following command to end the HTTP Admin server:

      ENDTCPSVR SERVER(*HTTP) HTTPSVR(*ADMIN)

   ii. After you wait for a few minutes, ensure that all Admin jobs are ended:

      WRKACTJOB JOB(ADMIN)

   iii. If any Admin jobs are still not cleaned up, end them with the End Job (**ENDJOB**) command.

   iv. Start the HTTP Admin server:

      STRTCPSVR SERVER(*HTTP) HTTPSVR(*ADMIN)

e. After a few minutes, run the `NETSTAT *CNN` command again to verify that port 2001 is in a status of Listen.

f. After port 2001 is listening, reattempt to connect your workstation by pointing your workstation to the URL that is listed at the beginning of this section.

► If the issue still is not resolved, collect the QMGTOOLS data for the PowerHA GUI as described in "Collecting diagnostic data for the PowerHA graphical user interface" on page 108 and create a PMR with IBM Support.

6.4.11 Checking the PowerHA status by using the PowerHA GUI

When you click PowerHA in the IBM Navigator for i, a PowerHA status panel is displayed, which provides a high-level status of the following PowerHA and system components:

► Cluster nodes
► IASPs
► Cluster administrative domains
► CRGs
► TCP/IP interfaces

Figure 6-20 shows the PowerHA status panel with IASP and CRG components that are flagged as needing attention. Hovering the cursor over the warning icon provides a general description of the issue for that component.
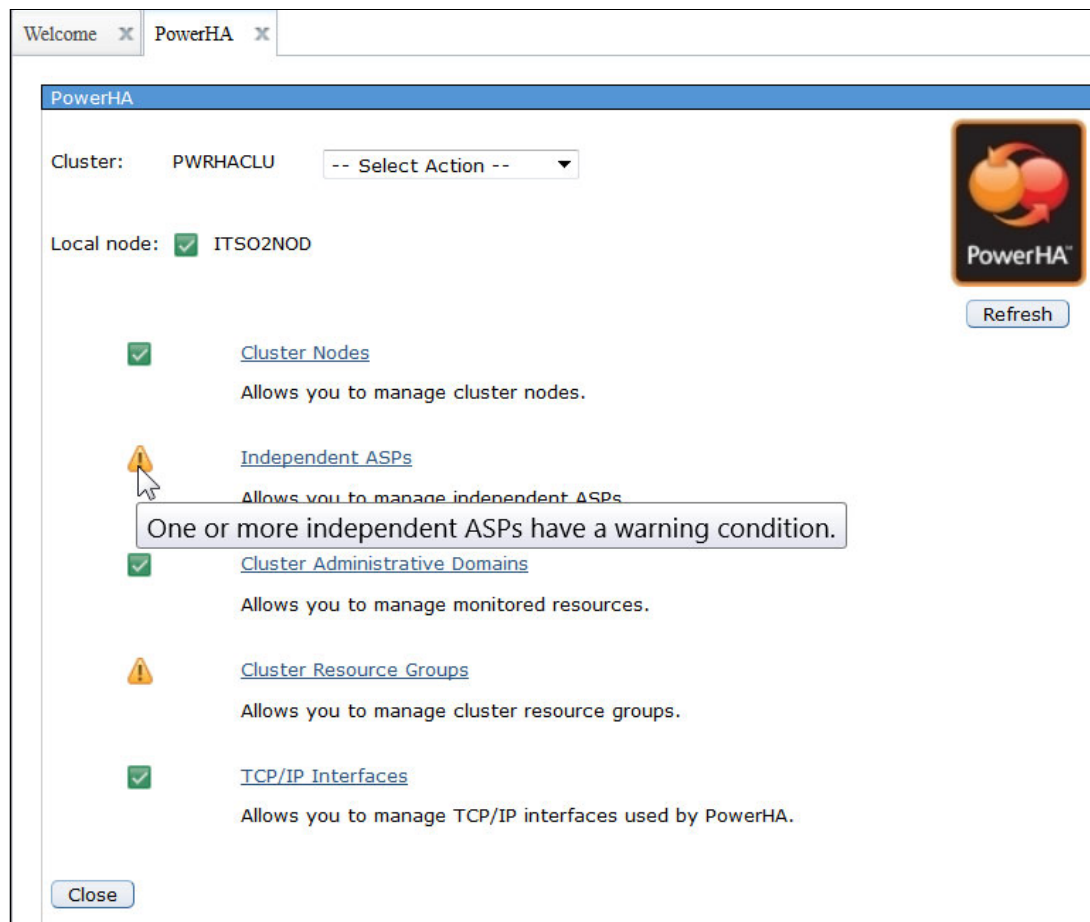
Figure 6-20   PowerHA GUI status window

For this example, the warning for IASPs was investigated. To get more details about this warning, click the IASPs component for a panel that is similar to the panel that is shown in Figure 6-21. Click the two-arrow icon (**>>**) for an option for details about this warning.

Figure 6-21   IASP status with a warning

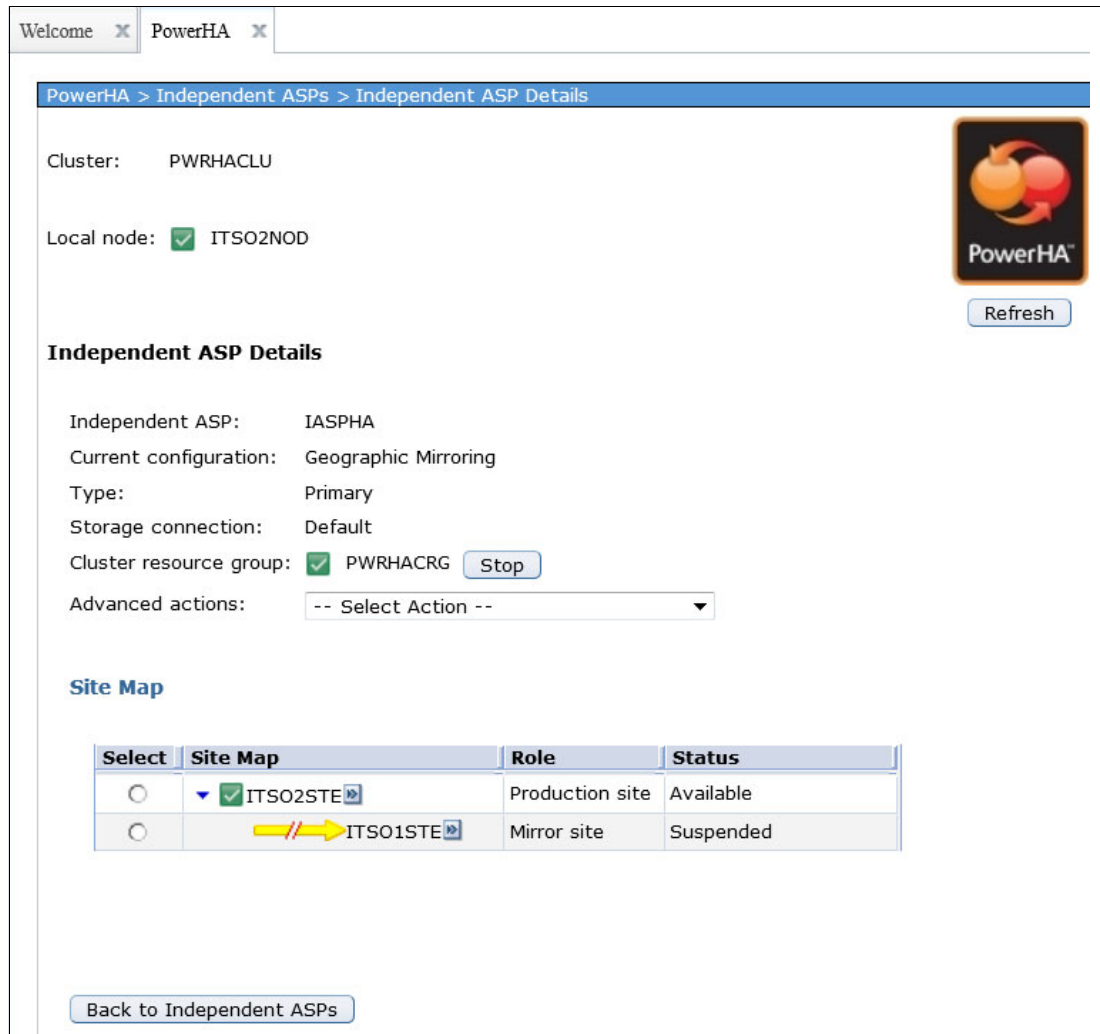The details for this issue are shown in Figure 6-22. From this panel, the session shows a status of *Suspended*. The broken yellow arrow indicates that the replication is not running.

Figure 6-22   IASP details that show a status of suspended

6.5 Troubleshooting PowerHA in a DS8000 environment

This section describes common situations that you might encounter and tools that are used in an PowerHA DS8000 environment.

The following topics are described in this section:

► 6.5.1, "Using DSCLI to access the DS8000 systems from IBM i" on page 121

► 6.5.2, "Diagnostic data for troubleshooting PowerHA in a DS8000 environment" on page 122

► 6.5.3, "Diagnostic data for troubleshooting PowerHA in a DS8000 environment by using ICSM or FSCSM" on page 122

► 6.5.4, "CMUC00201E: Authentication failure" on page 124

► 6.5.5, "CMUN00018E: Unable to connect to the management console server" on page 125

6.5.1 Using DSCLI to access the DS8000 systems from IBM i

When you troubleshoot the PowerHA in a DS8000 environment, you might need access to the management console of the DS8000 systems from the IBM i system.

To run commands within the IBM System Storage DS® Command-Line Interface (DSCLI), follow these steps:

> **Note:** These steps assume that DSCLI is installed on the IBM i system on which you are running the commands.

1. Run the following command to add the `QDSCLI` library to your library list:

   `ADDLIBLE QDSCLI`

2. On the IBM i command line, type **DSCLI** and press Enter.

3. On the Run Copy Services (DSCLI) command prompt (Figure 6-23), specify the connectivity options to connect to the management console of the DS8000.

   Specify the appropriate parameters for your DS8000 and press Enter:

   – Script = `*NONE`
   – HMC1 = IP address of HMC/secondary management console (SMC) of storage server
   – HMC2 = IP address of a second HMC/SMC, if used
   – User = DS8000 user ID
   – Password = Password for the DS8000 user ID that specified in the User field
   – DSCLI CMD = `*INT`

```
                      Run Copy Services (DSCLI)

 Type choices, press Enter.

 Script: *NONE or name  . . . . .

 Profile  . . . . . . . . . . . .    *DEFAULT

 HMC1 . . . . . . . . . . . . . .    *PROFILE
 HMC2 . . . . . . . . . . . . . .    *PROFILE
 User . . . . . . . . . . . . . .
 Password . . . . . . . . . . . .
 Install Path . . . . . . . . . .    '/ibm/dscli'
 DSCLI CMD  . . . . . . . . . . .
 Cmd Options  . . . . . . . . . .
               + for more values
                                                            More...
```
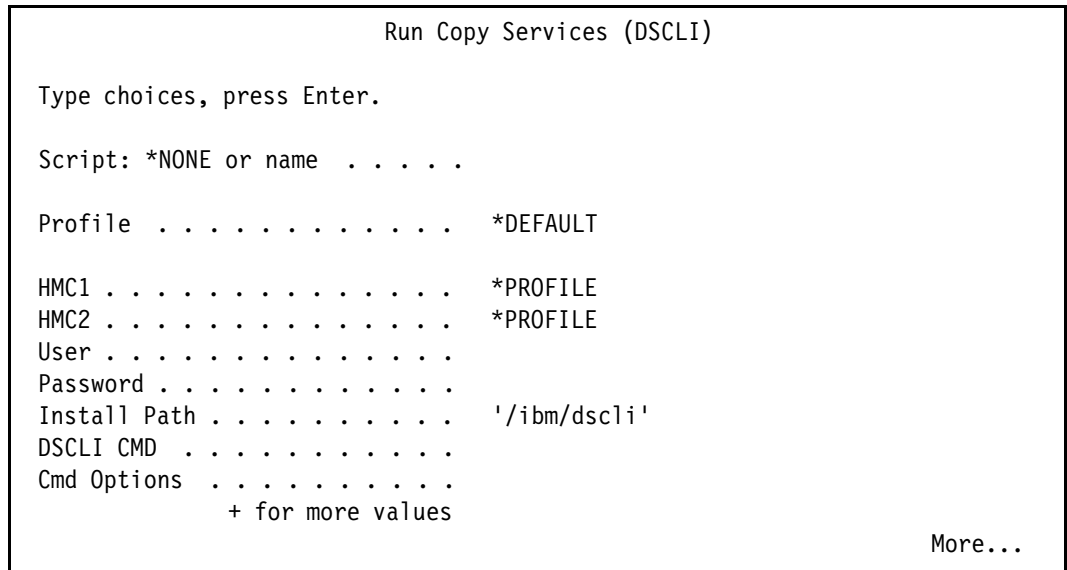
Figure 6-23   DSCLI command

4. The resulting DSCLI command entry panel is shown in Figure 6-24. From this DSCLI command entry shell, you can enter DS commands to help you troubleshoot your environment.

```
                            Java Shell Display


Date/Time: September 8, 2015 8:24:10 PM CEST IBM DSCLI Version: 7.7.5.14 DS:
IBM.2107-75APL71


dscli>




===>


F3=Exit  F6=Print F9=Retrieve F12=Exit
F13=Clear F17=Top F18=Bottom F21=CL command entry
```

Figure 6-24   DS8000 command-line interface

6.5.2 Diagnostic data for troubleshooting PowerHA in a DS8000 environment

If you use the IASP Copy Services Manager (ICSM) or Full System Copy Services Manager (FSCSM), collect the data that is specified in 6.5.3, "Diagnostic data for troubleshooting PowerHA in a DS8000 environment by using ICSM or FSCSM" on page 122.

You can troubleshoot most failures in a PowerHA environment by using external storage without ICSM or FSCSM by looking at the following diagnostic data:

► QHASVR job log
► Job log of the job that is running the failing command

6.5.3 Diagnostic data for troubleshooting PowerHA in a DS8000 environment by using ICSM or FSCSM

When you manage your environment with the IASP Copy Services Manager (ICSM) or by using the Full System Copy Services Manager (FSCSM), use the following Dump Information (**DMPINF**) command to gather troubleshooting data for IBM i support:

DMPINF ENV(*<environment_name>*) TYPE(*<environment_type>*)

This command dumps a file to the /QIBM/Qzrdhasm IFS directory. See Figure 6-25.

```
                        Command Entry                            GMENV
                                                    Request level:   1
 Previous commands and messages:
   > DMPINF ENV(DEMOGM) TYPE(*GMIR)
     ICSM information dumped to: /QIBM/Qzrdhasm/qzrdhasm_GMENV_151006_1049.txt



                                                                  Bottom
 Type command, press Enter.
 ===>
```

Figure 6-25   DMPINF results message in an ICSM environment

This file contains diagnostic data that includes the ICSM scripts, current results of those scripts, configurations for the environment, and debug messages for communications between the IBM i system and the DS8000 system. This data is the primary data that IBM Support requires to debug issues within an ICSM or FCSM environment.

When you run the `DMPINF` command in an FSCSM environment, you need to run the command on the Controller partition, which provides you and IBM Support all of the necessary debug data from the source, target, and controlling partitions.

The `DMPINF` command gathers the following diagnostic data in an FSCSM environment, plus additional debug and configuration information:

► Source partition:
  – `/tmp/qzrdiash.log` (non-**MKSYSCPY** actions)
  – `/tmp/fsfc/<config_name>/src.log` (source logs)
  – `FSFC` job log (user `QLPAR`)

► Target partition:
  – `/tmp/qzrdiash.log` (non-**MKSYSCPY** actions)
  – `/tmp/fsfc/<config_name>/tgt.log` (target logs)
  – `QSTRUPJD` job log
  – `ENDSYSCPY` job log
  – `/tmp/brms/flightrec` and `/tmp/brms/flightrec.bku` (Backup, Recovery, and Media Services (BRMS) logs)

► Controller partition:
  – `/tmp/qzrdiash.log` (non-**MKSYSCPY** actions)
  – `/tmp/fsfc/<config_name>/ctl.log` (controller logs)
  – `/tmp/fsfc/<config_name>/CTL_<job_name>.log` (job log of job that ran **MKSYSCPY**)

Also, the VIEWLOG tool can be run interactively to view the portion of the DMPINF data that shows the debug data while the IBM i interacts with the storage systems. An example of this VIEWLOG output is in Figure 6-26. VIEWLOG can be used to quickly view the status of recently executed scripts or FSCSM or ICSM commands.

```
 Browse : /QIBM/Qzrdhasm/qzrdhasm.log
 Record :    7654    of    7869 by  18                        Column :    1    130
 Control :

....+....1....+....2....+....3....+....4....+....5....+....6....+....7....+....
 288433 2015-10-07 08:47:37          This is a MMIR solution and a MMIR switch ty
 288433 2015-10-07 08:47:37          MMIR is reversed.
 288433 2015-10-07 08:47:37          Exit program ending. (qzrdiaexit : )

 288434 2015-10-07 08:50:04  Start  Exit Program for 330 starting from job 28843
 288434 2015-10-07 08:50:04          Node referrer is _mmir (qzrdiaexit : )
 288434 2015-10-07 08:50:04          Production node is DEMOHA
 288434 2015-10-07 08:50:04          Backup node is DEMOPROD
 288434 2015-10-07 08:50:04          MTIR node is
 288434 2015-10-07 08:50:04          This is a MMIR solution and a MMIR switch ty
 288434 2015-10-07 08:50:04          Exit program ending. (qzrdiaexit : )

 288387 2015-10-07 08:50:08  Start  STRFLASH for ACSMMIR starting from job 28838
                                     IBM i Copy Services Manager version 4.1.0 bu
 288387 2015-10-07 08:50:08          Using IASP CRG ACSMMIR (strflash : ACSMMIR)
 288387 2015-10-07 08:50:08          Flash target node is DEMOFC.
 288387 2015-10-07 08:50:08          ACSMMIR lock set by 288387/DPAINTER/QPADEV00
 288387 2015-10-07 08:50:08          Resolved the Node Reference to 1. (strflash
```

Figure 6-26   Example VIEWLOG output

6.5.4 CMUC00201E: Authentication failure

The CMUC00201E: Authentication failure error message commonly occurs when you run one of the following commands:

► Start Auxiliary Storage Pool Session (**STRASPSSN**) command
► End Auxiliary Storage Pool Session (**ENDASPSSN**) command
► Change Auxiliary Storage Pool Session (**CHGASPSSN**) command

The error message is in the job log of the job that runs the failing command.

This error indicates that the user successfully logged in to the management console for the DS8000. However, the password expired and authority to perform various functions on the DS8000 is limited.

Follow these steps to resolve this issue:

1. Issue the following command from a DSCLI interface on the DS management console:

   `chuser -pw <newpassword> <userID>`

   The *<userID>* is the user ID that was used to perform these DS functions and *<newpassword>* is the new password for the specified user ID.

   For more information about how to access the DS management console, see 6.5.1, "Using DSCLI to access the DS8000 systems from IBM i" on page 121.

2. If you use ICSM, the security file that the IBM i system uses to connect to the DS8000 must also be changed to reflect the new password by running the following command:

   `managepwfile -action change -pwfile /QIBM/QZRDHASM/sec.dat -mc1 <IPofMC1> -mc2 <IPofMC2> -name <userID> -pw <newpassword >`

   The variables are described:

   – *<IPofMC1>* is the IP address of the primary management console for the DS8000.

   – *<IPofMC2>* is the IP address of the secondary management console for the DS8000.

   – *<userID>* is the user ID that is used to perform these DS functions. This user ID is typically `QLPAR` when you use ICSM.

   – *<newpassword>* is the new password for the specified user ID.

6.5.5 CMUN00018E: Unable to connect to the management console server

The `CMUN00018E: Unable to connect to the management console server` error message from the DS8000 system can occur when you run one of the following commands:

► Start Auxiliary Storage Pool Session (**STRASPSSN**) command
► End Auxiliary Storage Pool Session (**ENDASPSSN**) command
► Change Auxiliary Storage Pool Session (**CHGASPSSN**) command

This message can be seen in the job log of the job that runs the failing command.

Follow these steps to resolve this issue:

1. Verify the basic IP connectivity between the IBM i system and the management console of the DS8000 by attempting to issue a **PING** command on the IBM i system to the management console of the DS8000.

2. If the **PING** connectivity is successful, attempt to establish a DSCLI connection to the management console of the DS8000 by using the instructions in 6.5.1, "Using DSCLI to access the DS8000 systems from IBM i" on page 121.

3. If the previous step is unsuccessful, it is possible that the level of the version of code on your DS8000 system and the version of Secure Sockets Layer (SSL) that is used by the IBM i DSCLI client might prevent the two systems from establishing a secure connection. This issue, along with the resolution, is described at the following website:

   https://ibm.biz/Bd4VKK

   If this article does not describe or resolve the issue, you must contact the IBM DS8000 storage team to determine why the DS8000 is posting this error to the IBM i. For more information about contacting IBM Support, see 6.5.8, "Contacting IBM Support" on page 128.

6.5.6 Check Peer-to-Peer Remote Copy (CHKPPRC) by using ICSM fails

Use the Check Peer-to-Peer Remote Copy (`CHKPPRC`) command in ICSM to check the status of Metro Mirror or Global Mirror. When `CHKPPRC` fails, you can review tools and diagnostic data to help you determine the cause of the issue.

Figure 6-27 shows an error message that you might receive when you attempt to run the `CHKPPRC` command.

```
Message ID . . . . . . :   IAS0070      Severity . . . . . . . :   60
Message type . . . . . :   Escape
Date sent  . . . . . . :   10/24/15     Time sent  . . . . . . :   11:09:30


Message . . . . : A PPRC check for IASP CRG <CRGname> has failed.
Cause . . . . . : One or more failures occurred while checking the status of
PPRC.  An SWPPRC command request will not operate.
Recovery  . . . : Refer to the prior diagnostic messages in the job log.  Also
use the VIEWLOG command to display additional details. Correct the problem and
retry the operation.
```

Figure 6-27   PPRC check failed

You can use the `VIEWLOG` command to look at relevant messages about this failed command. Figure 6-28 shows the initial message for this example. You need to look at the `lspprc_PS.script` file for possible additional error messages.

```
                              VIEWLOG

2015-09-01 12:09:07  Start  CHKPPRC for <CRGname> starting from job
175969/QUSER/QPADEV000B

2015-09-01 12:09:28  Processing file
/QIBM/Qzrdhasm/scripts/<CRGname>_MMIR/lspprc_PS.result. (checkThoseResults)

2015-09-01 12:09:28 Warning Strings Suspended | Duplex | Copy |  |  not found
in any records.

2015-09-01 12:09:28 Warning Expected lspprc_PS.script results not found.
```

Figure 6-28   Viewlog output for failed CHKPPRC

Use the `VIEWSCRIPT` command to display the file that contains the unexpected results. See Figure 6-29. In this case, an incorrect device ID was specified in the `lspprc_PS.script` file.

```
Browse :/qibm/qzrdhasm/scripts/<CRGname>_MMIR/lspprc_PS.result
 Record :      1   of       1 by  18                       Column :    1    129

....+....1....+....2....+....3....+....4....+....5....+....6....+....7....+....
 ***********Beginning of data**************
CMUN80027E lspprc: IBM.2107-75AX032: The DS Network Interface server is not
aware of the specified storage unit or storage image.
 ************End of Data******************
```

Figure 6-29   VIEWSCRIPT result

6.5.7 ICSM status codes

This section contains tables of status and request codes that can appear in the `DMPINF` or `VIEWLOG` output when you troubleshoot ICSM issues or check the status of your ICSM environment.

CRG PPRC status codes

Table 6-2 provides a list of several common PPRC status codes with their descriptions.

Table 6-2   PPRC status codes

Code number	Description
0 or *READY	PPRC ready for SWPPRC.
10	PPRC approved received reply from production node operator or user.
20	PPRC failover task is complete and a reply was received from the HA/DR node operator.
30	PPRC replication task is complete.
40	I/O resources were released.
50	I/O resources were reset.
100 or *INCOMPLETE	PPRC unscheduled switch is incomplete. An SWPPRC *COMPLETE is required if you are in a Metro Mirror environment. Global Mirror requires that completion is performed manually.

CRG FlashCopy status codes

Table 6-3 provides a list of common FlashCopy status codes with their descriptions.

Table 6-3   FlashCopy status codes

Code number	Description
0 or *NONE	FlashCopy ready for STRFLASH.
20	FlashCopy completed. Operator replied G to the IAS0001 message.
100 or *FLASHED	FlashCopy STRFLASH complete and ready for ENDFLASH.

CRG request codes

Table 6-4 provides a list of common CRG request codes with their descriptions.

Table 6-4   CRG request codes

Code number	Description
0 or *READY	No request. No action was taken on any node (used to set exit data only).
10	FlashCopy hardware check.
20	FlashCopy inquiry message to production system operator or user.
100	Not a request. It was used to define a FlashCopy/PPRC request boundary.
105	PPRC check CHKPPRC.
120	PPRC inquiry message to production system operator or user.
122	PPRC inquiry message to the backup system operator or user.

6.5.8 Contacting IBM Support

If you cannot resolve a problem, contact your country's IBM Support team. Use the following link to contact the IBM Storage support team:

https://ibm.biz/Bd4VK7

Use the following link to contact the IBM i support team:

https://ibm.biz/Bd4VaE

When you contact IBM i support, gather the following diagnostic data when you open a PMR:

► If you use PowerHA commands to manage your environment, gather the data that is described in 6.5.2, "Diagnostic data for troubleshooting PowerHA in a DS8000 environment" on page 122.

► If you use ICSM or FSCSM, gather the data that is described in 6.5.3, "Diagnostic data for troubleshooting PowerHA in a DS8000 environment by using ICSM or FSCSM" on page 122.

6.6 Troubleshooting PowerHA in a Storwize environment

This section describes common situations that you might encounter and common troubleshooting tools that are used in a PowerHA Storwize environment.

The following troubleshooting situations are described in this section:

6.6.1 Checking Storwize Copy Services status

When you use Storwize for your IASP storage and PowerHA for IBM System i®, you can log on to the Storwize system to check the status of the replication by following these steps:

1. To check the Metro or Global Mirror status, select **Remote Copy** from the Storwize GUI as shown in Figure 6-30.

Figure 6-30   Select Remote Copy to show the Remote Copy status

2. Figure 6-31 shows an example of several remote copy consistency groups and their states. The direction in which mirroring occurs is also shown.

Figure 6-31   Remote Copy status

3. For FlashCopy relationships and their status, select **Consistency Groups** from the Storwize GUI as shown in Figure 6-32.

Figure 6-32   Select Consistency Groups to show Consistency Group status

4. Figure 6-33 shows an example of the FlashCopy Consistency Group status window.

Figure 6-33   FlashCopy Consistency Group status

6.6.2 Diagnostic data for troubleshooting PowerHA in a Storwize environment

If you use the IASP Copy Services Manager (ICSM) or Full System Copy Services Manager (FSCSM), collect the data that is specified in 6.5.3, "Diagnostic data for troubleshooting PowerHA in a DS8000 environment by using ICSM or FSCSM" on page 122.

You can troubleshoot most failures in a PowerHA environment by using external storage without ICSM or FSCSM by looking at the diagnostic data in the QHASVR job log or the job log of the job that runs the failing command.

6.6.3 Gathering diagnostic data for an ICSM or FSCSM Storwize environment

When you manage your environment with the IASP Copy Services Manager (ICSM) or you use the Full System Copy Services Manager (FSCSM), use the following Dump Information (**DMPINF**) command to gather troubleshooting data:

DMPINF ENV(*<environment_name>*) TYPE(*<environment_type>*)

The **DMPINF** command dumps a file to the /QIBM/Qzrdhasm IFS directory. See Figure 6-34.

```
                            Command Entry                              DEMOFC
                                                              Request level:   1
 Previous commands and messages:
    > DMPINF ENV(ACSMMIR) TYPE(*MMIR)
      ICSM information dumped to:
/QIBM/Qzrdhasm/qzrdhasm_DEMOFC_151006_1014.txt


                                                                      Bottom
 Type command, press Enter.
 ===>
```

Figure 6-34   DMPINF in an ICSM environment

This file contains diagnostic data that is often required by IBM Support to debug issues within an ICSM or FCSM environment.

When you run a **DMPINF** command in an FSCSM environment, you need to run it on the Controller partition, which provides you and IBM Support with all of the necessary debug data from the source, target, and controlling partitions. The **DMPINF** command gathers the following diagnostic data in an FSCSM environment plus additional debug and configuration information:

► Source partition:
  – /tmp/qzrdiash.log (non-**MKSYSCPY** actions)
  – /tmp/fsfc/*<config_name>*/src.log (source logs)
  – FSFC job log (user QLPAR)
► Target partition:
  – /tmp/qzrdiash.log (non-**MKSYSCPY** actions)
  – /tmp/fsfc/*<config_name>*/tgt.log (target logs)
  – QSTRUPJD job log
  – ENDSYSCPY job log
  – /tmp/brms/flightrec and /tmp/brms/flightrec.bku (BRMS logs)
► Controller partition:
  – /tmp/qzrdiash.log (non-**MKSYSCPY** actions)
  – /tmp/fsfc/*<config_name>*/ctl.log (controller logs)
  – /tmp/fsfc/*<config_name>*/CTL_*<job_name>*.log (job log of job that ran **MKSYSCPY**)

6.6.4 Problems with DSPSVCSSN hanging or not responding

The Display SVC Session (`DSPSVCSSN`) command can hang for an unexpected amount of time. This command requires communications to occur between the nodes that are associated with the session. Often, this command appears to hang due to one of the cluster nodes that is in or intermittently going into and out of `Partition` status.

To check this situation, follow these steps:

1. Enter the Display Cluster Information (`DSPCLUINF`) command and press Enter.

2. Press Enter again to view the status of the cluster nodes.

3. If any of the cluster nodes are in `Partition` status, resolve this issue by following the steps in 6.4.3, "Cluster node in partition status" on page 112.

4. If the cluster nodes are all in `Active` status,collect QMGTOOLS data from all nodes while the command is in this hung condition. See 6.2.4, "Collecting diagnostic data with QMGTOOLS" on page 103 for steps to help you collect this data. Contact IBM Support at this time and provide them with this data.

6.6.5 CHGSVCSSN OPTION(*RESUME) fails

When you attempt to resume a suspended session, this process can fail. The common reason is that the CRG is not active. The job log for the interactive session contains messages that indicate a reason for the failure. If a message in the job log indicates that the CRG is not in the correct state, this situation can be checked and resolved by following these steps:

1. Run the Display CRG Information (`DSPCRGINF`) command to view the status of the *DEV CRG Type as shown in Figure 6-35.

```
                      Display CRG Information

Cluster  . . . . . . . . . . . . :    CLUSTER
Cluster resource group . . . . . :    *LIST
Consistent information in cluster:    Yes
Number of cluster resource groups:    2

                   Cluster Resource Group List


Cluster Resource Group   CRG Type      Status                    Primary Node
     ADMCRG               *PEER        Active                       *NONE
     CLUCRG               *DEV         Inactive                     CTCIHA4A



                                                                      Bottom
   F1=Help   F3=Exit   F5=Refresh   F12=Cancel   Enter=Continue
```

Figure 6-35   DSPCRGINF command that shows an inactive CRG

2. Start the CRG by using the Start Cluster Resource Group (`STRCRG`) command.

3. Attempt the Change Storage Area Network (SAN) Volume Controller Auxiliary Storage Pool Session (`CHGSVCSSN`) command again.

   If this command still fails, recheck the interactive session job log to see whether any new error messages indicate the reason for the failure of the `CHGSVCSSN` command. If you cannot determine any further cause, collect the QMGTOOLS data as shown in 6.2.4, "Collecting diagnostic data with QMGTOOLS" on page 103.

6.6.6 Contacting IBM Support

If you cannot resolve a problem, contact your country's IBM Support team. Use the following link to contact the IBM Storage support team:

https://ibm.biz/Bd4VK7

Use the following link to contact the IBM i support team:

https://ibm.biz/Bd4VaE

When you contact IBM i support, gather the following diagnostic data when you open a PMR:

► If you use PowerHA commands to manage your environment, gather the data that is described in 6.5.2, "Diagnostic data for troubleshooting PowerHA in a DS8000 environment" on page 122.

► If you use ICSM or FSCSM, gather the data that is described in 6.5.3, "Diagnostic data for troubleshooting PowerHA in a DS8000 environment by using ICSM or FSCSM" on page 122.

6.7 Troubleshooting PowerHA in a geographic mirroring environment

This section describes common situations that you might encounter and troubleshooting tools that are used in a PowerHA geographic mirroring environment.

The following troubleshooting situations are described in this section:

6.7.1 Problems with DSPASPSSN hanging or not responding

The Display Auxiliary Storage Pool Session (`DSPASPSSN`) command can hang for an unexpected amount of time. This command requires communications to occur between the nodes that are associated with the session. The common cause for this command to appear to hang is due to one of the cluster nodes that is in or intermittently going into and out of `Partition` status.

To check this situation, run the following commands:

1. Run the Display Cluster Information (`DSPCLUINF`) command and press Enter.

2. Press Enter again to view the status of the cluster nodes.

3. If any of the cluster nodes are in `Partition` status, resolve this issue by following the steps in 6.4.3, "Cluster node in partition status" on page 112.

4. If the cluster nodes are all in `Active` status, collect the QMGTOOLS data from all nodes while the command is in this hung condition. See 6.2.4, "Collecting diagnostic data with QMGTOOLS" on page 103 for steps to assist you to collect this data. Contact IBM Support and provide them with this data.

6.7.2 CHGCRG for RCYDMNACN(*CHGCUR) fails

You can use the Change CRG (`CHGCRG`) command to change the primary and backup node roles in the recovery domain. Use this command with the `RCYDMACN(*CHGCUR)` option during a switchover in a geographic mirroring environment when you want to promote the preferred backup system to the preferred primary system and vice versa.

The common reason for this command to fail is due to the CRG being in an `Active` state. You must end the CRG to perform this function. Run the End CRG (`ENDCRG`) command from any of the nodes in the recovery domain and try the `CHGCRG` command again.

6.7.3 Dual production role scenario

One situation that can arise in certain circumstances due to a switchover or failover process is a dual production role scenario. You can see this situation when you dump the advanced analysis macro `GEOSTAT` on both the preferred production and preferred backup copy systems.

An example of how this situation appears is shown in Figure 6-36 on page 136 and Figure 6-37 on page 137.

Figure 6-36 shows `GEOSTAT` on a preferred source system that shows the production role scenario.

```
     DISPLAY/ALTER/DUMP                              RCHESP1-GEOSTAT
RUNNING MACRO: GEOSTAT                          -ALL
PROCESSING CLUSTER NODE: RCHESP1
CRG ESPCRG     IS INACTIVE
PROCESSING IASP: 33 ESPIASP    AVAILABLE  PRIMARY ASP
IASP INFORMATION:
     PHYSICAL COPY ID:               OXE7
     COPY ROLE:                      OXD7  PRODUCTION
     REMOTEMIRRORERRORRECOVERYPOLICY:   2  SUSPEND
     REMOTEMIRRORENCRYPTION:            1  NO
     REMOTEMIRRORTRACKRESOURCES:   172277  PAGES
     CONNECTION ID:                    E8
     MIRRORCOPYSTATE:                   4  SUSPEND WITH TRACKING
     MIRRORCOPYDATASTATE:               2  USABLE
     REMOTEMIRRORDELIVERY:              1  ASYNC
     REMOTEMIRRORPERFORMANCEMODE:       2  ASYNC TARGET
     REMOTEMIRRORSYNCPRIORITY:         10  HIGH
     REMOTEMIRRORERRORRECOVERYTIMEOUT:  2  MINUTES
     REMOTEMIRRORAUTORESUME:            0
     SYNCSTATUS:                        2  SYNC IS REQUIRED
     SYNCTYPE:                          2  PARTIAL SYNC IS REQUIRED
          TOTAL DATA IN TRANSIT: 0 PAGES
     TRACKING SPACE:                   0% USED.
     DATA TRACKED CONN ID E8:      1792436 PAGES.
```

Figure 6-36   GEOSTAT on a preferred source system that shows the production role scenario

Figure 6-37 shows **GEOSTAT** on a preferred target system that also shows the production role scenario.

```
     DISPLAY/ALTER/DUMP                              RCHESP2-GEOSTAT
RUNNING MACRO: GEOSTAT                        -ALL
PROCESSING CLUSTER NODE: RCHESP2
CRG ESPCRG     IS INACTIVE
PROCESSING IASP: 33 ESPIASP    AVAILABLE  PRIMARY ASP
IASP INFORMATION:
     PHYSICAL COPY ID:              OXE8
     COPY ROLE:                     OXD7  PRODUCTION
     REMOTEMIRRORERRORRECOVERYPOLICY:    2  SUSPEND
     REMOTEMIRRORENCRYPTION:             1  NO
     REMOTEMIRRORTRACKRESOURCES:    172277  PAGES
     CONNECTION ID:                 E7
     MIRRORCOPYSTATE:                    4  SUSPEND
     MIRRORCOPYDATASTATE:                2  USABLE
     REMOTEMIRRORDELIVERY:               1  ASYNC
     REMOTEMIRRORPERFORMANCEMODE:        2  ASYNC TARGET
     REMOTEMIRRORSYNCPRIORITY:          10  HIGH
     REMOTEMIRRORERRORRECOVERYTIMEOUT:   2  MINUTES
     REMOTEMIRRORAUTORESUME:             0
     SYNCSTATUS:                         3  FULL SYNC IS REQUIRED
     SYNCTYPE:                           3  FULL SYNC IS REQUIRED
          TOTAL DATA IN TRANSIT: O PAGES
```

Figure 6-37   GEOSTAT on a preferred target system that also shows the production role scenario

This scenario can happen due to situations that cause a failover to occur making the preferred backup system the current production system, but timing and the lack of the systems to communicate with one another prevents the preferred production system from becoming the current backup system.

To resolve this situation, follow these steps:

1. Determine the system that you want to be the production system. You might want to consider which system has current data in the IASP. One system might not be accessible due to networking issues.

   This example assumes that you want the preferred production system to be the current production system again. For this example, the systems in the GEOSTAT dumps in Figure 6-36 on page 136 and Figure 6-37 are used. RCHESP1 is the preferred production system and RCHESP2 is the preferred backup system.

2. Ensure that the IASP is varied off on both RCHESP1 and RCHESP2 systems by using the Work with Configuration Status (**WRKCFGSTS**) command.

3. End the CRG by using the **ENDCRG** command. You must end the CRG to use the **CHGCRG** command to change node roles.

4. Issue the following Change CRG (**CHGCRG**) command to make RCHESP1 the *PRIMARY node and RCHESP2 the *BACKUP 1 node:

   CHGCRG CLUSTER(ESPCLU) CRG(ESPCRG) CRGTYPE(*DEV) RCYDMNACN(*CHGCUR)
   RCYDMN((RCHESP1 *PRIMARY *SAME SITE1 *SAME (*<ip_address_of_RCHESP1>*)) (RCHESP2
   *BACKUP 1 SITE2 *SAME (*<ip_addres_ of_RCHESP2>*')))

5. Start the CRG by using the **STRCRG** command.

6. Vary on the IASP on RCHESP1 by using the `WRKCFGSTS` command.

7. After the vary-on completes, a resynchronization of the IASP data starts. If it does not start, you can use the `CHGASPSSN OPTION(*RESUME)` command to start the resynchronization process. This process overwrites any changes to the IASP that occurred on RCHESP2 during this time.

6.7.4 CHGASPSSN OPTION(*RESUME) fails

When you attempt to resume a suspended session, this process can fail. The common reason for this failure is that the CRG is not active. The job log for the interactive session contains messages that indicate a reason for the failure. If a message in the job log indicates that the CRG is not in the correct state, this situation can be checked and resolved by following these steps:

1. Run the Display CRG Information (`DSPCRGINF`) command to view the status of the *DEV CRG Type as shown in Figure 6-38.

```
                          Display CRG Information

Cluster  . . . . . . . . . . . . :     ESPCLU
Cluster resource group . . . . . :     *LIST
Consistent information in cluster:     Yes
Number of cluster resource groups:     2

                      Cluster Resource Group List


Cluster Resource Group    CRG Type      Status                    Primary Node
      ESPADMN             *PEER         Active                        *NONE
      ESPCRG              *DEV          Inactive                     RCHESP1



                                                                      Bottom
F1=Help    F3=Exit    F5=Refresh    F12=Cancel    Enter=Continue
```

Figure 6-38   DSPCRGINF that shows an inactive CRG

2. Start the CRG by using the `STRCRG` command.

3. Attempt the `CHGASPSSN` command again. If this command still fails, check the interactive session job log again to see whether new error messages indicate the new reason for the failure of the `CHGASPSSN` command.

4. If you can determine nothing further, collect the QMGTOOLS data as described in 6.2.4, "Collecting diagnostic data with QMGTOOLS" on page 103.

6.7.5 Common return codes when you manage geographic mirroring

When you are running in a geographic mirroring environment, certain actions can cause various outcomes that result in a `CPFBA48` error message that is produced with a wide range of reason codes. This message can be posted when you manage the ASP session. Due to many reason codes that can occur for all known issues that result in `CPFBA48`, it is not possible to specify all of them within the message text.

The following document provides information about each of the possible return codes, their meanings, and possible recovery procedures:

http://www.ibm.com/support/docview.wss?uid=nas8N1020895

6.7.6 Contacting IBM Support

If you cannot resolve a problem, contact your country's IBM Support team. Use the following link to contact the IBM Storage support team:

https://ibm.biz/Bd4VK7

Use the following link to contact the IBM i support team:

https://ibm.biz/Bd4VaE

PowerHA Tools for IBM i

This appendix describes the PowerHA Tools for IBM i offerings and services that are available from IBM Systems Lab Services.

The PowerHA Tools for IBM i complement and extend the PowerHA and IBM storage capabilities for high availability (HA) and disaster recover (DR).

The PowerHA Tools for IBM i provide the following benefits:

► Help reduce business risks and improve resiliency for critical applications
► Simplify the setup and automation of HA, DR, and backup solutions
► Reduce the cost of maintaining and regularly testing an HA/DR environment
► Facilitate flexible deployment options for single or multiple site protection
► Assure consistent deployment by using preferred practices and experienced consultants

For more information about PowerHA Tools for IBM i, see the following IBM Systems Lab Services and Training website:

http://www.ibm.com/systems/services/labservices

PowerHA Tools for IBM i

Table A-1 lists the PowerHA Tools for IBM i that are available from IBM Systems Lab Services.

Table A-1   PowerHA Tools for IBM i

PowerHA Tools for IBM i	Capability	Benefit	DS8000	Storwize	Internal
Smart Assist for PowerHA on IBM i	Provides operator commands and scripts to supplement the PowerHA installation and ongoing operations for independent auxiliary storage pool (IASP)-enabled applications.	Simplifies deployment and ongoing management of high availability (HA) for critical IBM i applications.	X	X	X
IASP Copy Services Manager (Automated recovery with faster IASP-level vary-on, no system IPL)					
FlashCopy	Automates FlashCopy of IASP for daily offline backup with seamless Backup, Recovery, and Media Services (BRMS) integration.	Increases application availability by reducing or eliminating the backup window for routine daily backups.	X	X	
Logical unit number (LUN)-level Switching	Simplifies deployment and automates switching an IASP between IBM i cluster nodes in one data center.	Enables a business continuity manager to provide a simple, single-site HA solution.	X[1]	X[2]	
Metro Mirror or Global Mirror	Simplifies initial deployment and automates ongoing server and storage management of two-site Metro Mirror or Global Mirror HA or DR solutions. Requires IASP-enabled applications.	Enables a business continuity manager to provide seamless operation of integrated server and storage operations for two-site HA and DR.	X		
Metro Global Mirror (MGM)	Extends PowerHA functionality to provide a three-site server or storage replication solution that contains Metro Mirror for HA with Global Mirror for DR. Requires IASP-enabled applications and IBM Tivoli® Productivity Center for Replication.	Enables a business continuity manager to further lower business risks and maximize business resilience for critical applications that require three-site HA and DR protection.	X		
Full System Copy Services Manager (Automated recovery and requires full-system IPL)					
FlashCopy	Automates full system FlashCopy for daily offline backup with integrated support for BRMS without IASP-enabled applications.	Increases application availability by reducing or eliminating the backup window for online daily backups. Enables an entry solution while you plan for IASP enablement.	X	X	
Metro Mirror or Global Mirror	Simplifies initial deployment and automates ongoing server and storage management of two-site Metro Mirror or Global Mirror HA or DR solutions without IASP-enabled applications.	Enables a business continuity manager to provide seamless operation of integrated server and storage operations for HA and DR. Enables an entry solution while an IASP enablement is planned.	X	X[3]	

[1] DS8000 support available with PowerHA Tools for IBM i 6.1 or earlier; included in PowerHA SystemMirror 7.1
[2] V7000 support included with PowerHA 7.1 Technology Refresh (TR6)
[3] Storwize Full System Replication requires Full System Replication for PowerHA

IBM Lab Services Offerings for PowerHA for IBM i

Table A-2 lists the PowerHA for IBM i service offerings that are available from IBM Systems Lab Services.

Table A-2   PowerHA for IBM i service offerings

PowerHA for IBM i Service Offering	Description
IBM i High Availability Architecture and Design Workshop	An experienced IBM i consultant conducts a planning and design workshop to review solutions and alternatives to meet HA and DR and backup and recovery requirements. The consultant provides an architecture and implementation plan to meet these requirements.
PowerHA for IBM i Bandwidth Analysis	An experienced IBM i consultant reviews network bandwidth requirements for implementing storage data replication. IBM reviews I/O data patterns and provides a bandwidth estimate to build into the business and project plan for clients who are deploying PowerHA for IBM i.
IBM i Independent Auxiliary Storage Pool (IASP) Workshop	An experienced IBM i consultant provides jumpstart services for migrating applications into an IASP. Training includes enabling applications for IASPs, clustering techniques, and managing PowerHA and HA and DR solution options with IASPs.
PowerHA for IBM i Implementation Services	An experienced IBM consultant provides services to implement an HA and DR solution for IBM Power Systems servers with IBM Storage. Depending on specific business requirements, the end-to-end solution implementation can include a combination of PowerHA for IBM i and PowerHA Tools for IBM i, and appropriate storage software, such as Metro Mirror, Global Mirror, or FlashCopy.

B

IBM Live Partition Mobility

IBM Power System™ servers are designed to offer the highest stand-alone availability in the industry. Still, enterprises must occasionally restructure their infrastructure to meet new IT or business requirements. By moving your running production applications from one physical server to another, Live Partition Mobility (LPM) allows for nondisruptive maintenance or modification to a system, which mitigates the effects on partitions and applications that were caused by the occasional need to shut down a system.

> **Important:** Because LPM works only when the current production system is up and running, it is not a replacement for another high availability (HA) solution. Instead, it is another feature that you want to use with your HA solution.

LPM can help in reducing planned downtimes for hardware maintenance or replacement of systems and in moving logical partitions (LPARs) to another system to achieve a more balanced workload distribution. Because the complete LPAR is moved to a new system and nothing remains on the original system, LPM cannot help in scenarios where you need access to one environment when you are performing productive work on another environment, such as operating system maintenance or application maintenance.

This appendix describes the following topics:

Why LPM

The Live Partition Mobility for IBM i can be used for the following purposes:

► Migration. LPM is a good tool to use for migration between systems. Figure B-1 shows available migrations for IBM i.
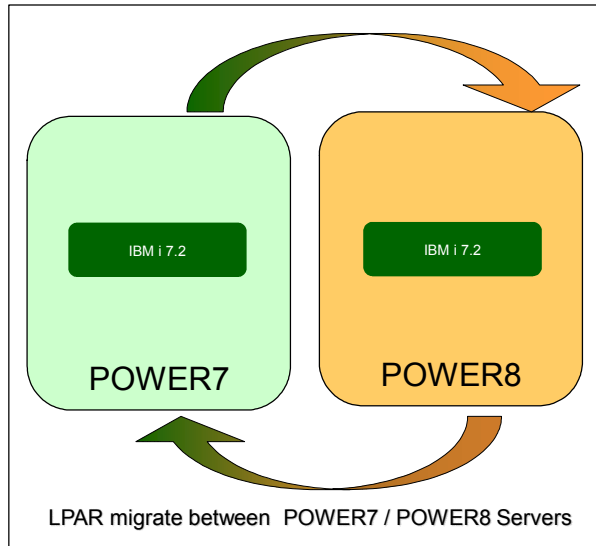
Figure B-1   LPM for IBM i

► Hardware maintenance activities. You can prepare concurrent maintenance for primary resources with no application outages.

► Workload balancing. LPM offers better distribution of server resources.

LPM requirements

This section describes the software and hardware requirements for LPM.

The minimum software requirements for LPM are listed:

► IBM i

► PowerVM Enterprise Edition on both source and target systems

► Integrated Virtualization Manager (IVM) is provided by the Virtual I/O Server (VIOS) at release level 1.5.1.1 or higher

► Virtual I/O Server at release level VIOS V2.2.2.1 Fix Pack (FP) 26 or higher

The minimum hardware requirements for LPM are listed:

► POWER7 with firmware release 740.40 or 730.51, or later

► IBM Hardware Management Console V7R7.5.0M0, or later

► All I/O must be virtual and supported by VIOS, which can be Virtual SCSI (VSCSI), Virtual Fibre Channel or N_Port ID Virtualization (NPIV), or Virtual Ethernet.

► Configuration must use only external storage and it must be accessible to both source and destination systems.

Note: Both source and destination systems must be on the same Ethernet network.

How LPM works

The following steps provide an overview of the process that is involved in using LPM to move a running LPAR to a new system:

1. As a starting point in Figure B-2, the IBM i client partition is running on System #1. On System #2, a Virtual I/O Server is running that connects to the storage system that System #1 is using. As shown in Figure B-2, no configuration is available on System #2 for the IBM i LPAR that we want to migrate to that system.

   The environment is checked for required resources, such as main memory or processor resources.
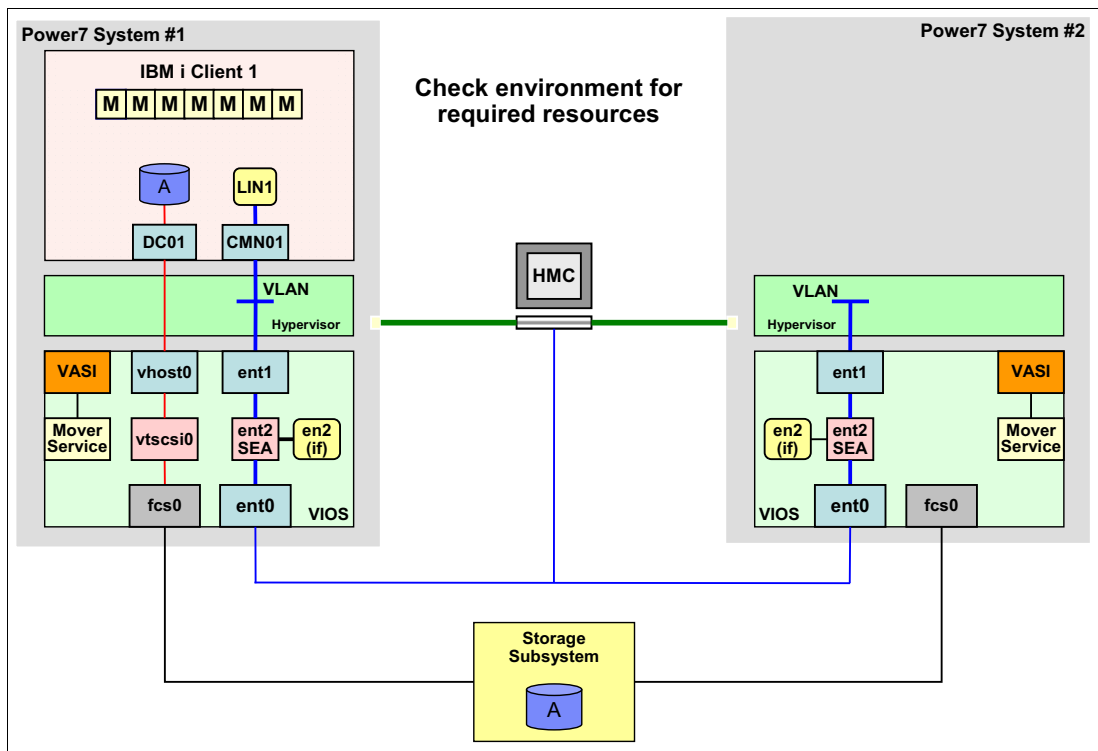
Figure B-2   LPM resource check

2. If this check is successful, a shell LPAR is created on System #2, as shown in Figure B-3.
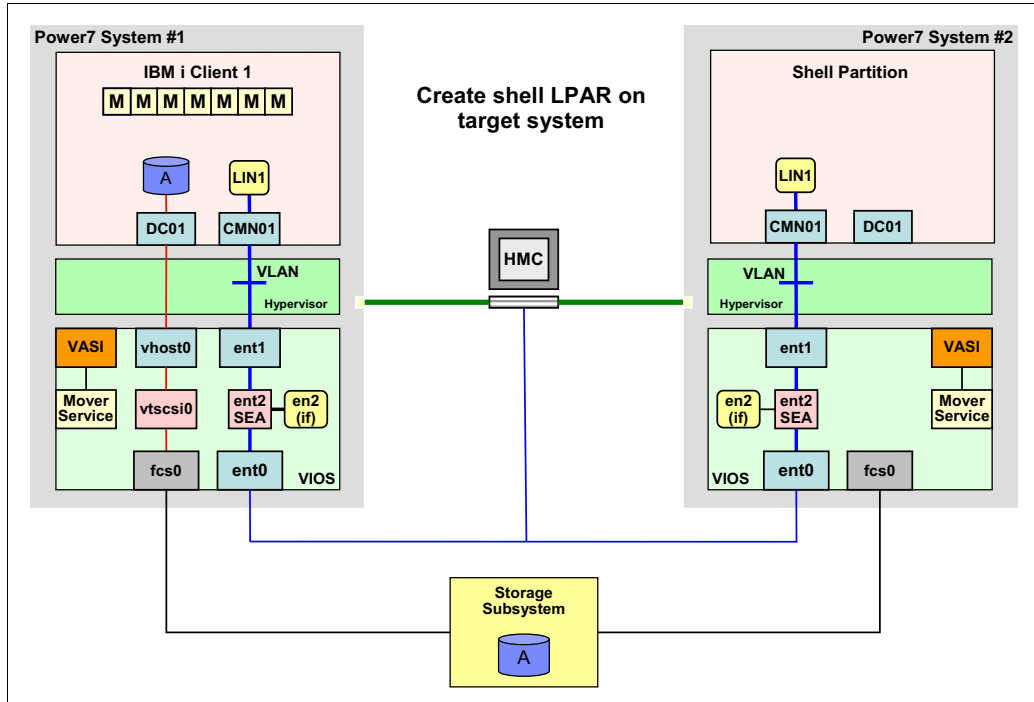
Figure B-3   LPM shell LPAR created

3. Virtual SCSI or NPIV devices are created in the virtual target VIOS server and in the shell partition to give the partition access to the storage subsystem that is used by System #1, as shown in Figure B-4.
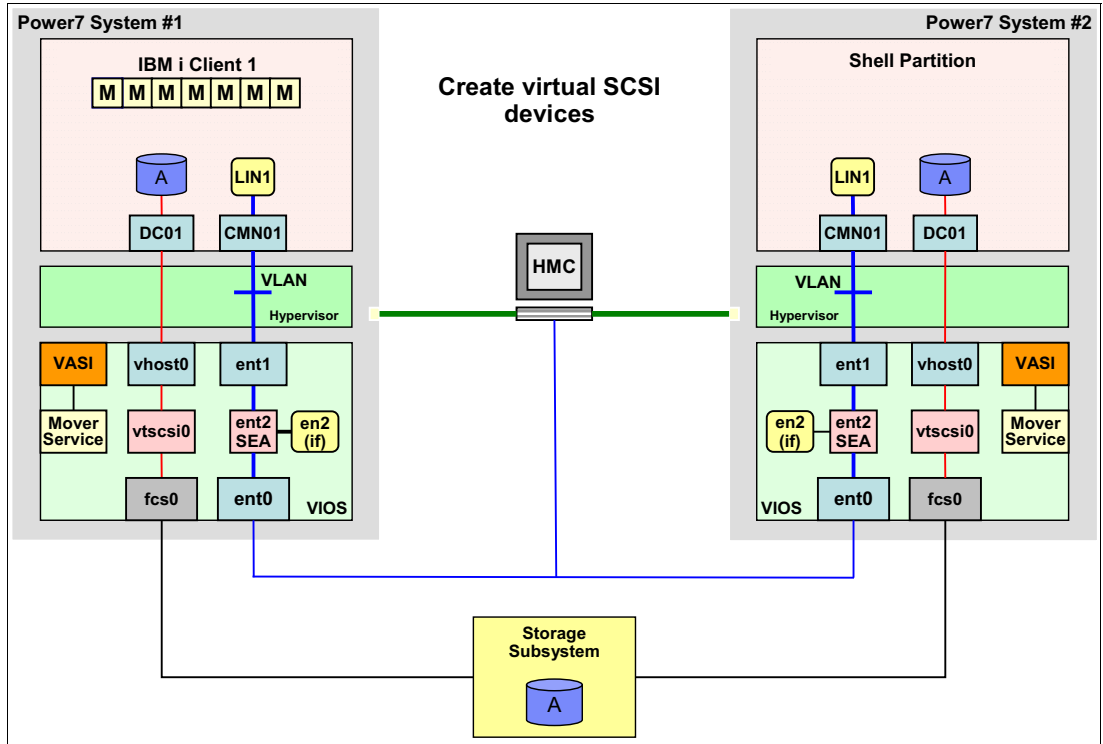
Figure B-4   LPM virtual SCSI devices created

4. Main memory is migrated from System #1 to System #2, as shown in Figure B-5. Users can still work on System #1 while this process occurs. Whenever a page from main memory is changed on System #1 after it was moved to System #2, that page must be moved from System #1 to System #2 again. Therefore, consider starting a partition migration during times of low activity on the system.
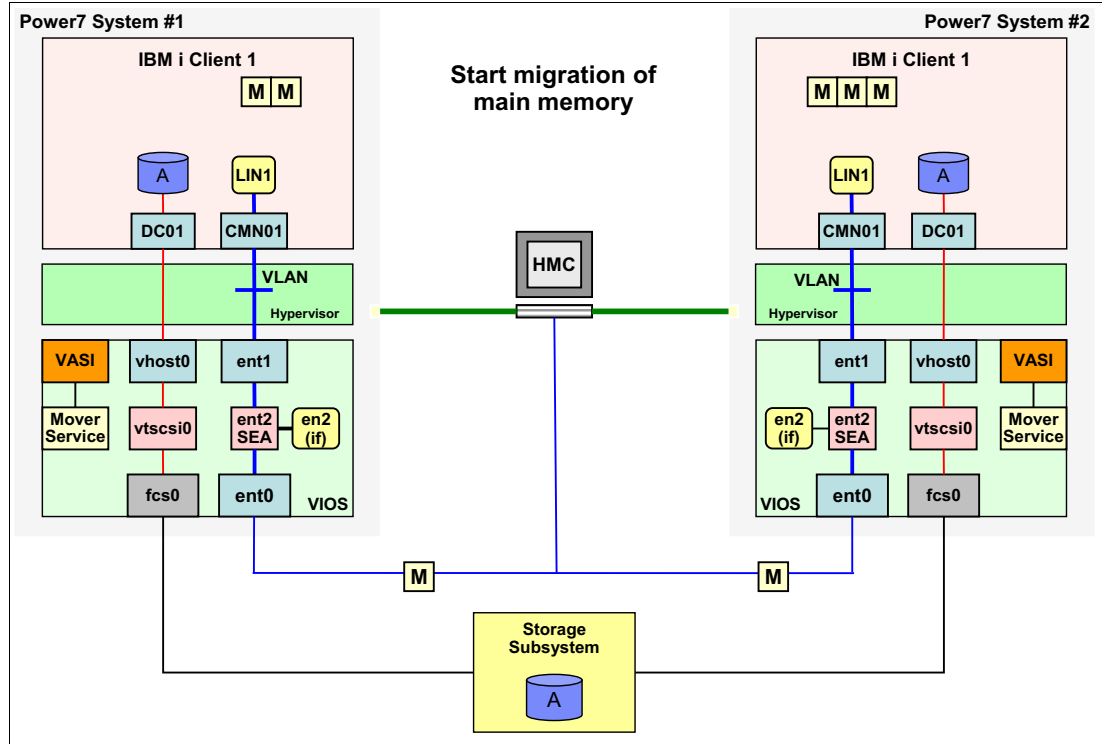
Figure B-5   LPM memory migration

5. When the system reaches a state where only a few memory pages were not migrated, System #1 is suspended, as shown in Figure B-6. From this point, no work is possible on System #1. Users notice that their transactions seem to freeze. The remaining memory pages are then moved from System #1 to System #2.
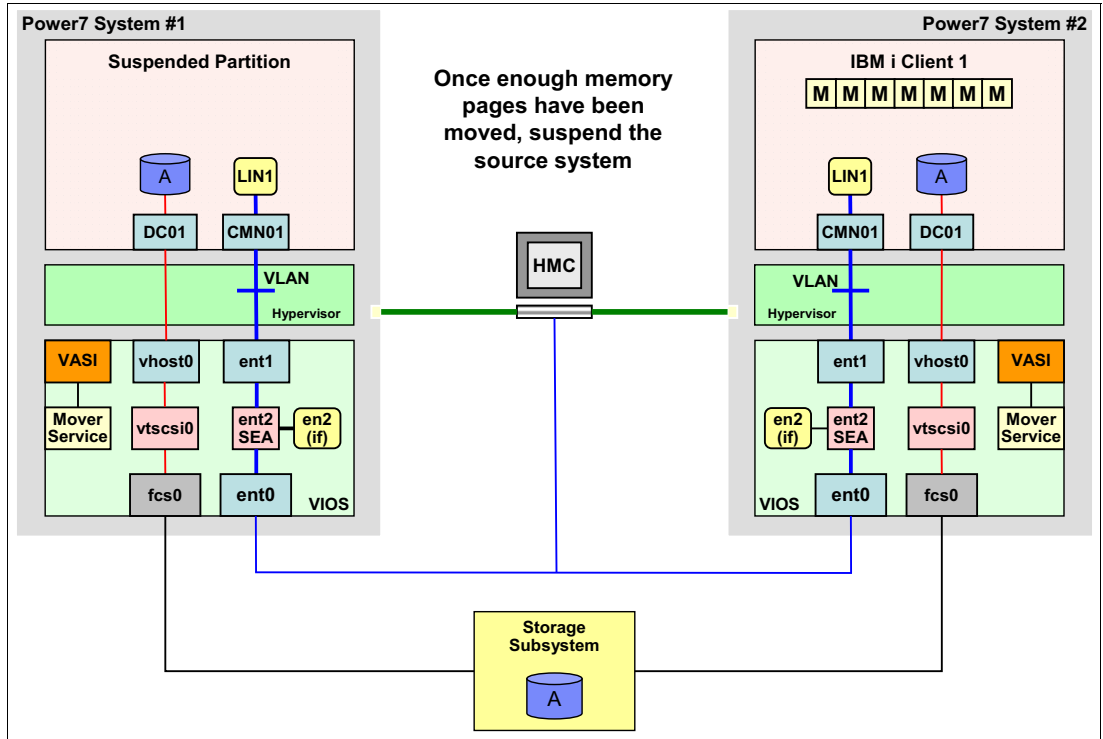
Figure B-6   LPM suspension of the source system

6. After the memory page migration completes, the original LPAR definition on System #1 is deleted with the virtual I/O (VIO) definitions in the VIOS, as shown in Figure B-7. The target system resumes and user transactions start to run again.

All IBM i jobs are moved over to the new systems. Users experience a delay in response time while the system is suspended. However, because jobs are still active on the system after the migration finishes, uses can continue to work without needing to log on to the system again.
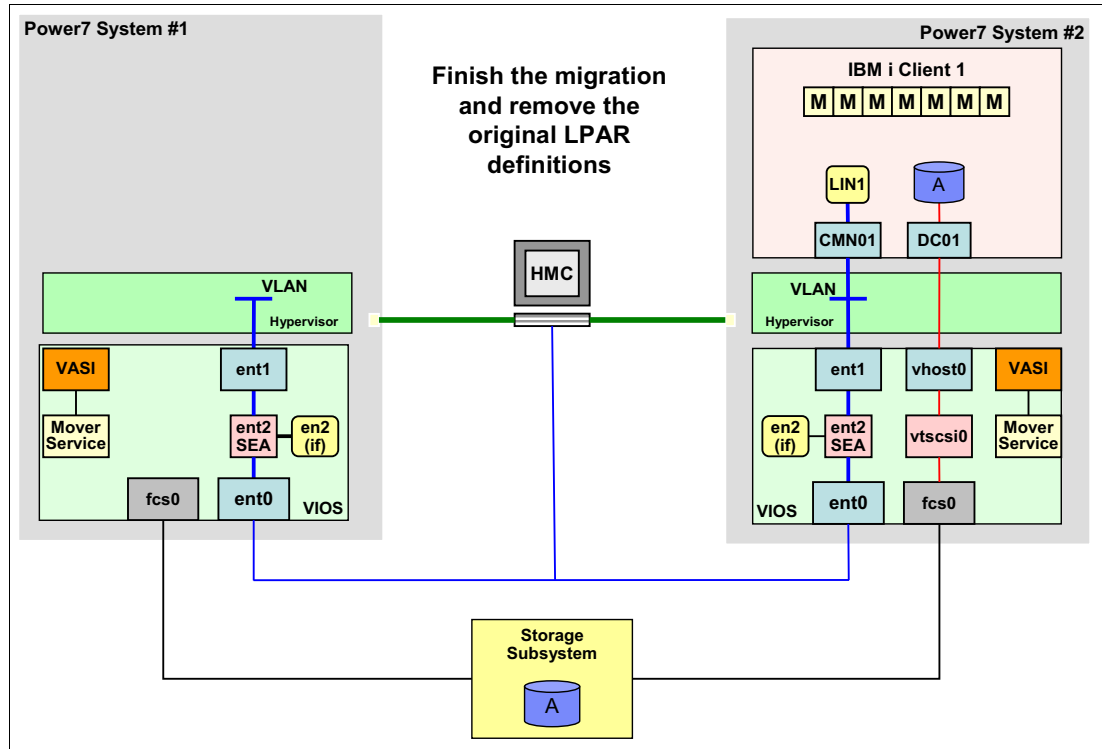
Figure B-7   LPM finished status

C
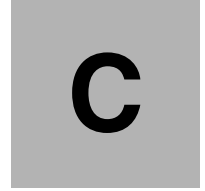
IBM Power Enterprise Pools

IBM Power Enterprise Pools is a technology for dynamically sharing mobile processor and memory activations among a group (or *pool*) of IBM Power Systems enterprise-class servers.

Power Enterprise Pools can support business goals in the following ways:

► Power Enterprise Pools can provide an organization with a dynamic infrastructure, reduced cost of performance management, improved service levels, and controlled risk management.

► The Power Enterprise Pools technology is ideal for further improving the flexibility, load balancing, and disaster recovery (DR) planning and operations of Power Systems.

► Instead of moving LPARs between systems to accommodate for varying resource demands, move resources to the servers that need them.

► The reliability, availability, and serviceability (RAS) of a Power Systems environment can be increased by using Power Enterprise Pools.

The following topics are described in this appendix:

IBM Power Enterprise Pools overview

IBM Power Enterprise Pools, together with mobile processor and memory activations, allow the reallocation of these mobile resources to any system within a defined pool. Systems in a pool can be of different processor generations and can run at different clock speeds.

Activation assignment and resource movement are controlled by a Hardware Management Console (HMC) and can be performed without any intervention from IBM. No limitation exists on the number of movements that can occur. No limit exists on the amount of time that a resource can be used on a system with a pool. The movement of resources is instant, dynamic, and nondisruptive.

Figure C-1 shows an example of a Power Enterprise Pool setup. The pool contains four systems with static core activations (shown in orange), mobile core activations that were previously assigned to the systems (shown in blue), and cores that are installed but not yet activated (shown as "dark" cores in black).
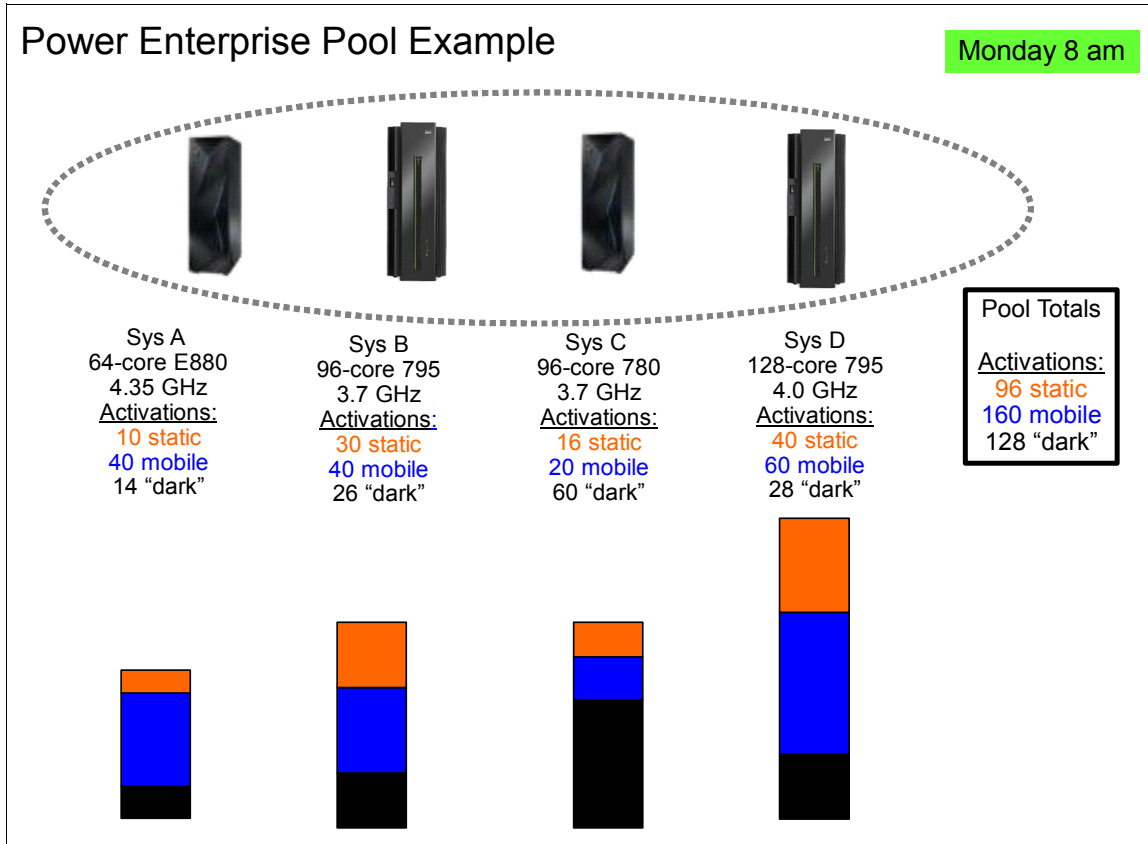

Figure C-1   Power Enterprise Pool starting point

If System A needs to be powered off to perform maintenance, you can move the mobile activations to other systems in the pool by simple operations on the HMC. You end up with a configuration as shown in Figure C-2.
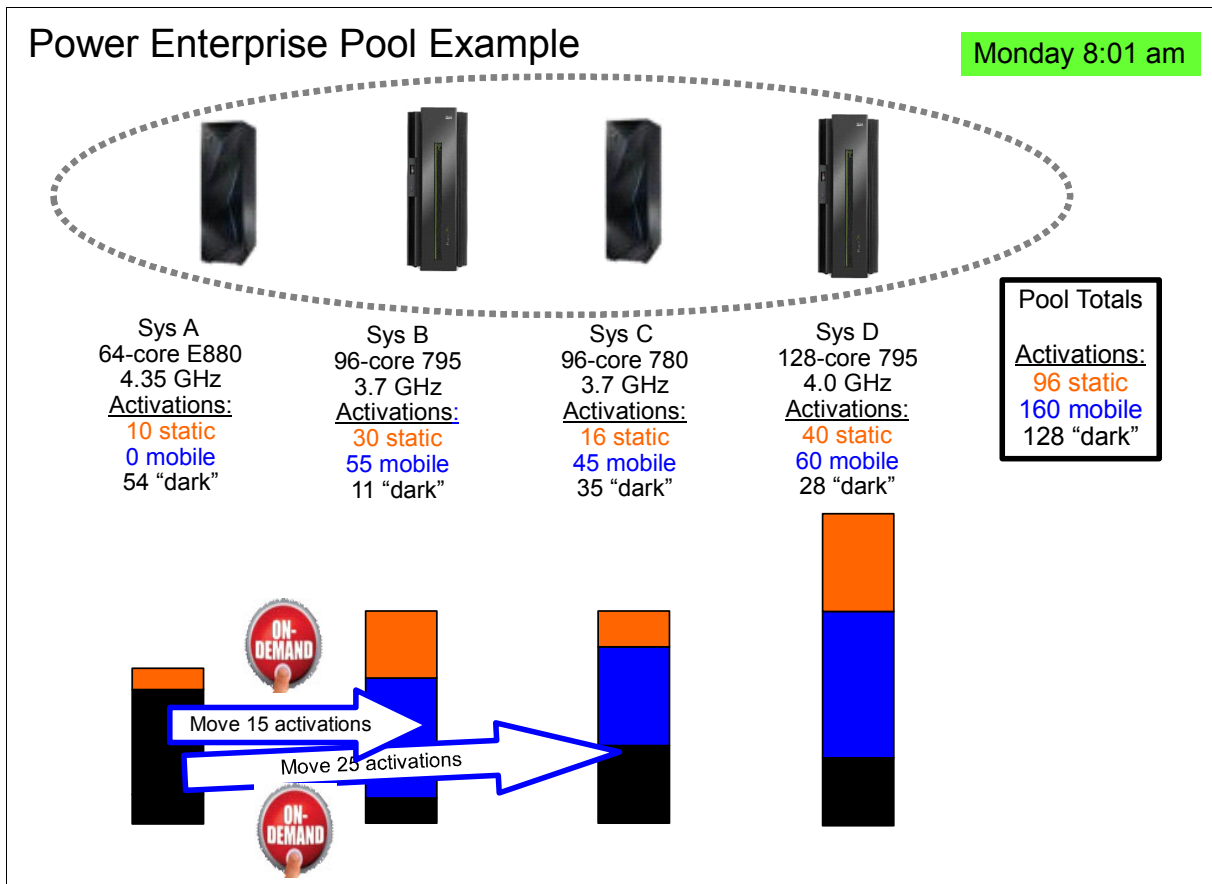

Figure C-2   Power Enterprise Pools moving core activations

In addition, you can combine Power Enterprise Pools with Elastic Capacity on Demand (CoD) as shown in Figure C-3 to accommodate for short performance peeks.
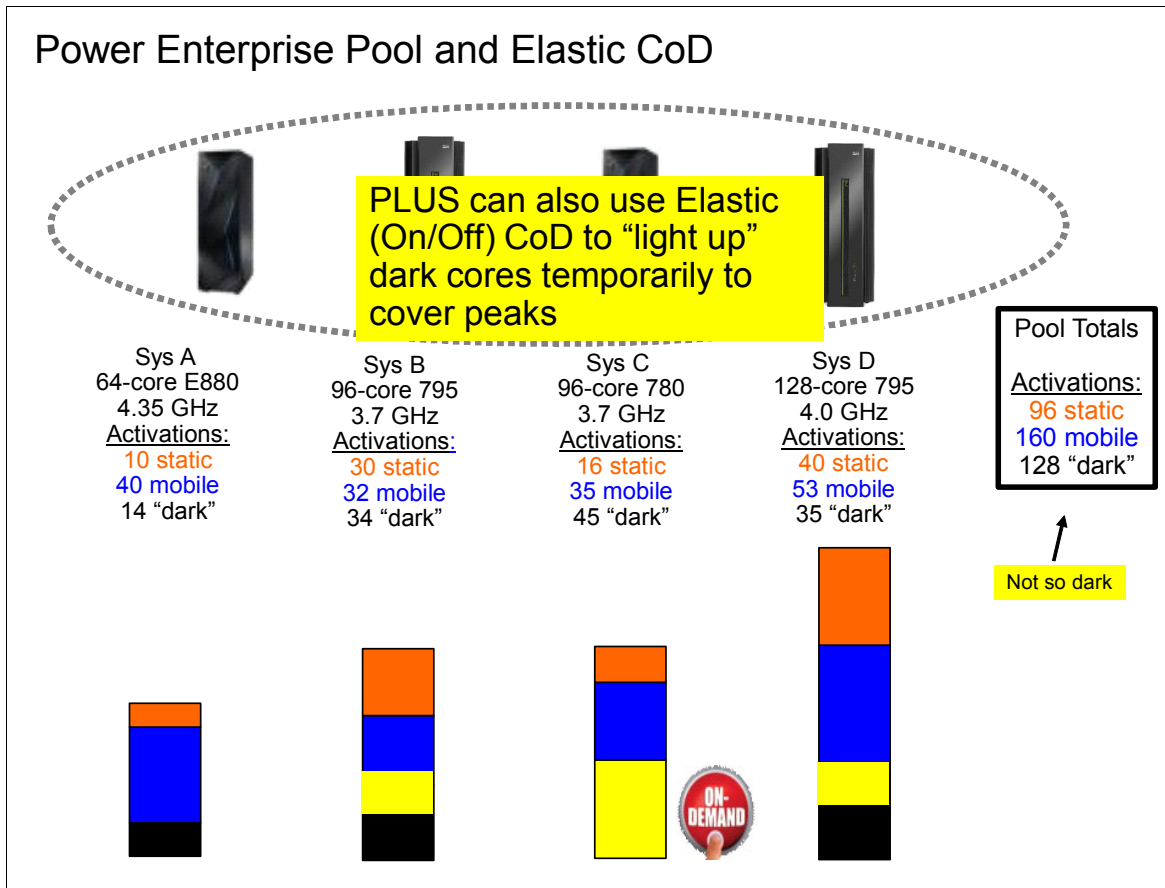

Figure C-3   *Power Enterprise Pool with elastic CoD*

Power Enterprise Pools prerequisites

Power Enterprise Pools are available for the following systems:

- ► 9117-MMD (770 D model)
- ► 9179-MDH (780 D model)
- ► 9119 FHB (795)
- ► 9119-MME (E870)
- ► 9119-MHE (E880)

Two kinds of pools exist:

- ► One pool type can contain Power 770 and Power 870 systems. The maximum number of systems in this pool is 48.

- ► The other pool type can contain Power 780, Power 795, and Power880 systems. The maximum number of systems in the second type of pool is 32.

A mixture of the two system types in one pool is not allowed. A specific system can belong to only one Enterprise Pool at any point in time.

You manage a Power Enterprise Pool by using an HMC. One master HMC, which needs to be attached to all systems within the pool, is required. One master HMC can manage multiple Enterprise Pools. The master HMC must have at least 2 GB of main memory to run this function. All servers and HMCs must be at firmware level 7.8 or higher.

Hardware and software maintenance must be consistent for all servers in the pool. Either all systems are running with warranty or maintenance or none of the systems are running with maintenance or warranty.

When you move processor activations within the pool, core-based software licenses can also move to another system together with the underlying core. For an IBM i environment, this capability is true for IBM i per core entitlements, PowerHA entitlements, and IBM i Enterprise enablements.

IBM i 5250 Enterprise enablements are temporarily transferable from a server in the pool that has more than one entitlement or that has a full entitlement. Each server in the pool that is going to use 5250 capability must have at least one 5250 Enterprise enablement feature. The base quantity of one 5250 enablement is not temporarily transferable. With a full Enterprise enablement feature, a maximum quantity of 5250 entitlements that is equal to the number of mobile processor activations can be temporarily transferred.

All systems in the pool require a certain number of static activations that cannot be transferred to another system in the pool. The minimum number of static activations are listed:

► Power 770: 4 cores
► Power 870: 8 cores
► Power 780: 4 cores
► Power 795: 24 cores or 25% of *installed* cores (larger one of both values)
► Power 880: 8 cores
► Memory: 25% of *installed* memory (Normal activation rules for memory still apply.)

Because mobile activations are still a hardware feature that belongs to a specific serial number, a system that leaves the pool (because it is no longer used in an environment) leaves the pool with all of the mobile activations that it brought into the pool.

For more information about how to order, implement, and manage Power Enterprise Pools, see the following IBM Redbooks publications:

► *Sharing Processor and Memory Activations Dynamically Among IBM Power Systems Enterprise Class Servers*, TIPS1169

► *Power Enterprise Pools on IBM Power Systems*, REDP-5101

D

Power Systems Capacity BackUp

The Power Systems Capacity BackUp (CBU) offerings are designed to support disaster recovery (DR) and high availability (HA) needs. The Capacity BackUp offerings recognize that true HA or DR solutions require at least two systems. If one system is not available, the other system "takes over". The CBU offering is designed to give flexible and economic options for deploying business continuity operations.

The CBU offering is designed to address that fact that IBM i licenses are not transferable between different system serial numbers. An *entitlement* is an agreement to allow the client use of the license keys on the CBU for production purposes if they are not used concurrently with the primary production system. *Transferring IBM i entitlements* means that the client agrees that they use the CBU processor activations that are purchased for HA operations nonconcurrently with its primary counterpart processor.

The following topics are described in this appendix:

CBU prerequisites

A CBU system can be ordered as an initial order or when you upgrade a model to a new system. It is not possible to change an existing system to a CBU model.

The primary system must be in an equal or higher processor group than the CBU system.

The primary system and the CBU system must belong to the same enterprise.

If a CBU is no longer affiliated with the original registering IBM client, it is not recognized as a CBU any more.

The CBU system must be registered to the corresponding primary system on the IBM Capacity Backup for Power Systems website:

https://ibm.biz/Bd4VgK

License considerations for CBU

For IBM i, a software key tells each system license manager how many IBM i processor license entitlements each system has. If the client ever runs or assigns more IBM i work than the system has entitlements, messages are periodically issued to the system operator that indicate that the system is out of license compliance. By the terms of the IBM licensing agreement, the client must then bring the system into compliance by either reducing IBM i workload or adding IBM i processor license entitlements.

The client must sign the additional license agreement for CBU temporary transfers. This agreement documents that IBM and the client understand when IBM i core or user entitlements and 5250 Enterprise enablements can be transferred from the primary system to the CBU system.

> **Important:** When entitlements are temporarily transferred to the CBU system, the software key information is not altered on either machine. The CBU system issues the out-of-compliance message that indicates that it is out of compliance. The additional license agreement terms give permission to the client under this situation to ignore these messages. The client agrees to monitor the IBM i usage on both the primary and CBU systems to ensure that the IBM i usage on the pair of machines does not exceed the total number of IBM i processor license entitlements.

In addition to the IBM i processor license entitlement, P05 and P10 processor group level systems have IBM i user entitlements. User entitlements are transferable from the primary to the CBU. If the primary system is an IBM i processor-based system (P20 processor group or higher), and the CBU is a processor-based and user-based system, you must purchase any required user entitlements on your CBU.

> **Important:** The special license agreement for CBU systems allows a temporary transfer of IBM i licenses from the primary system to the CBU system only. A transfer of IBM i licenses that belong to the CBU system to the primary system is not allowed.

To temporarily transfer either IBM i or 5250 Enterprise Enablements from the primary system, you need at least one IBM i and one Enterprise Enablement license on your CBU system. If the primary IBM license is an Enterprise Edition, the "base" 5250 Enterprise Enablement cannot be transferred, but optional enablements (above the base) can be transferred temporarily.

For implementations that are not based on storage-based clustering configurations, such as IBM PowerHA SystemMirror for i, which inherently manages the production workload to access only one node in the cluster configuration at a time, it is required that the primary and CBU partitions be managed so that the assigned processing units and assigned virtual processors for the virtual Capacity BackUp (vCBU) partition are set to the minimum entitled cores before the workload failover. The maximum processing units and virtual processors are set to the number of entitlements to which vCBU expands on a workload failover. The inverse operation is performed on the primary partition post failover. The primary is set to the minimum entitled cores.

Figure D-1 on page 162 shows an example of required licenses in an environment with three logical partitions (LPARs) on the primary system. No HA is needed for LPAR 1. Therefore, no counterpart exists for this LPAR on the CBU system. HA is required for LPAR 2 and LPAR 3. IBM i and PowerHA are licensed for all cores in these two LPARs on the primary system.

In this example, you receive one key for IBM i and one key for PowerHA for the primary system with eight core entitlements. On the CBU edition, only one license for IBM i and one license for PowerHA are required. You receive one key for IBM i with one entitlement (and are allowed to ignore the license error messages if you do not use licenses in parallel on both systems). For PowerHA, you can request a temporary key from the key center that is valid for 2 years for eight core entitlements on the CBU edition. For all other IBM i license programs (that are licensed for one server), you can also request temporary backup keys from the key center.

Important: You are not allowed to use those IBM i license programs in parallel on the primary system and the CBU system.

Figure D-1 shows a CBU licensing example of the required licenses in an environment with three LPARs on the primary system.
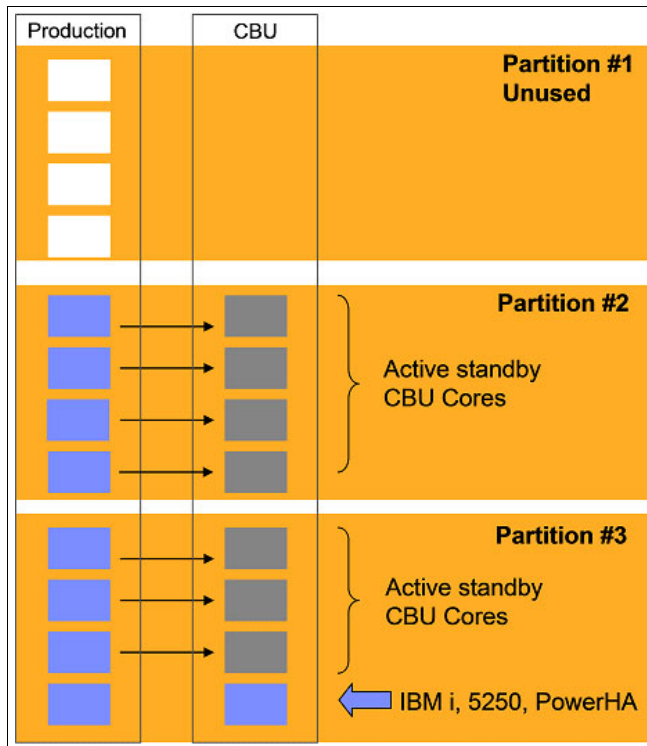

Figure D-1   CBU licensing example

If any workload that is running on the CBU system while the actual production is running on the primary system requires more than one core, the additional cores must be licensed for IBM i. The same rule is true if you want to run any other workload on the CBU system, for example, a test and development license program agreement (LPA). Ensure that you acquire all licenses on the CBU system that you need.

CBU setup options

Figure D-2 shows the basic setup of a primary system and a CBU system with the capability to transfer IBM i licenses from the primary to the CBU system temporarily.
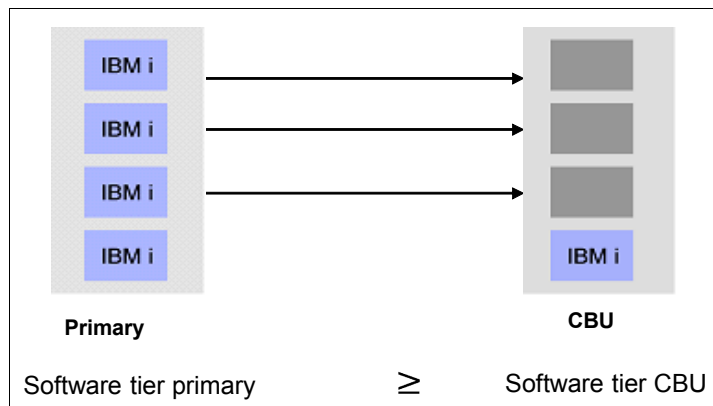

Figure D-2   CBU basic setup

Situations might occur where multiple primary systems are paired with multiple CBU systems as shown in Figure D-3.
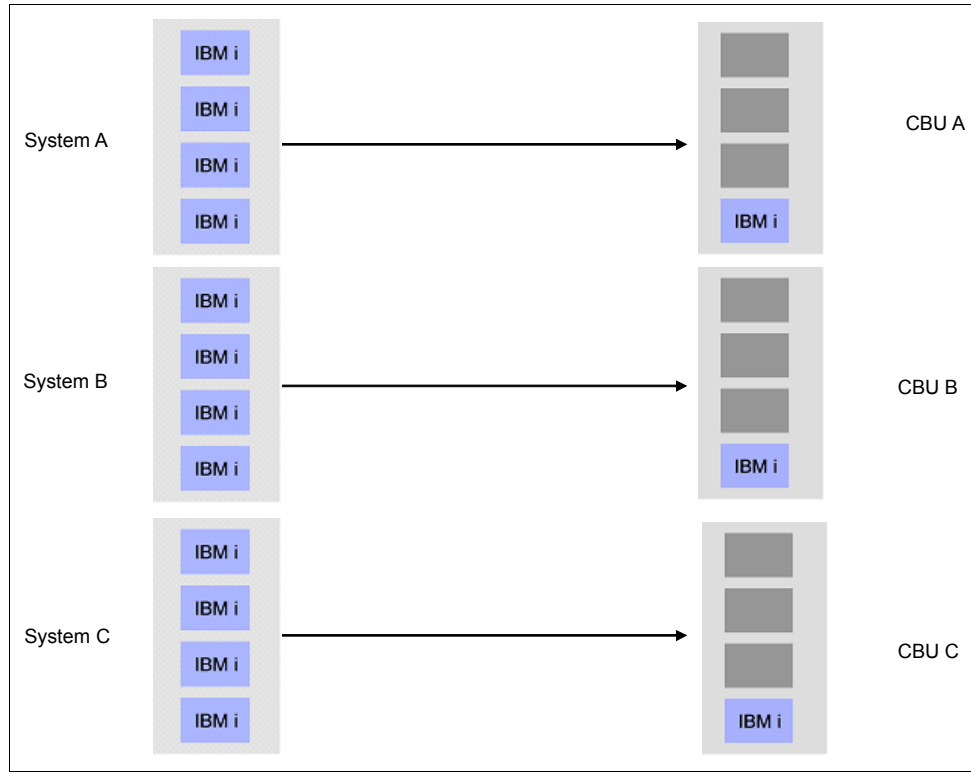

Figure D-3   Multiple system CBU setup

It might be feasible to combine the individual backup systems into one larger consolidated backup. This setup is also supported by a CBU edition as shown in Figure D-4 on page 164.

All systems in this setup must be owned by one company. Each of the primary systems must to be registered to the CBU edition.

Note: This setup requires the use of the concept of a virtual CBU (vCBU).

For each of the primary systems, a vCBU LPAR must be set up on the backup system. Each of these vCBU LPARs requires at least one per core license of IBM i. Each of the primary systems can fall over to the combined backup system individually. The IBM Electronic Service Agent™ must run on each of the primary systems and on the CBU system.

Figure D-4 shows an environment where the individual backup systems are combined into one larger consolidated backup, which is supported by a CBU edition.
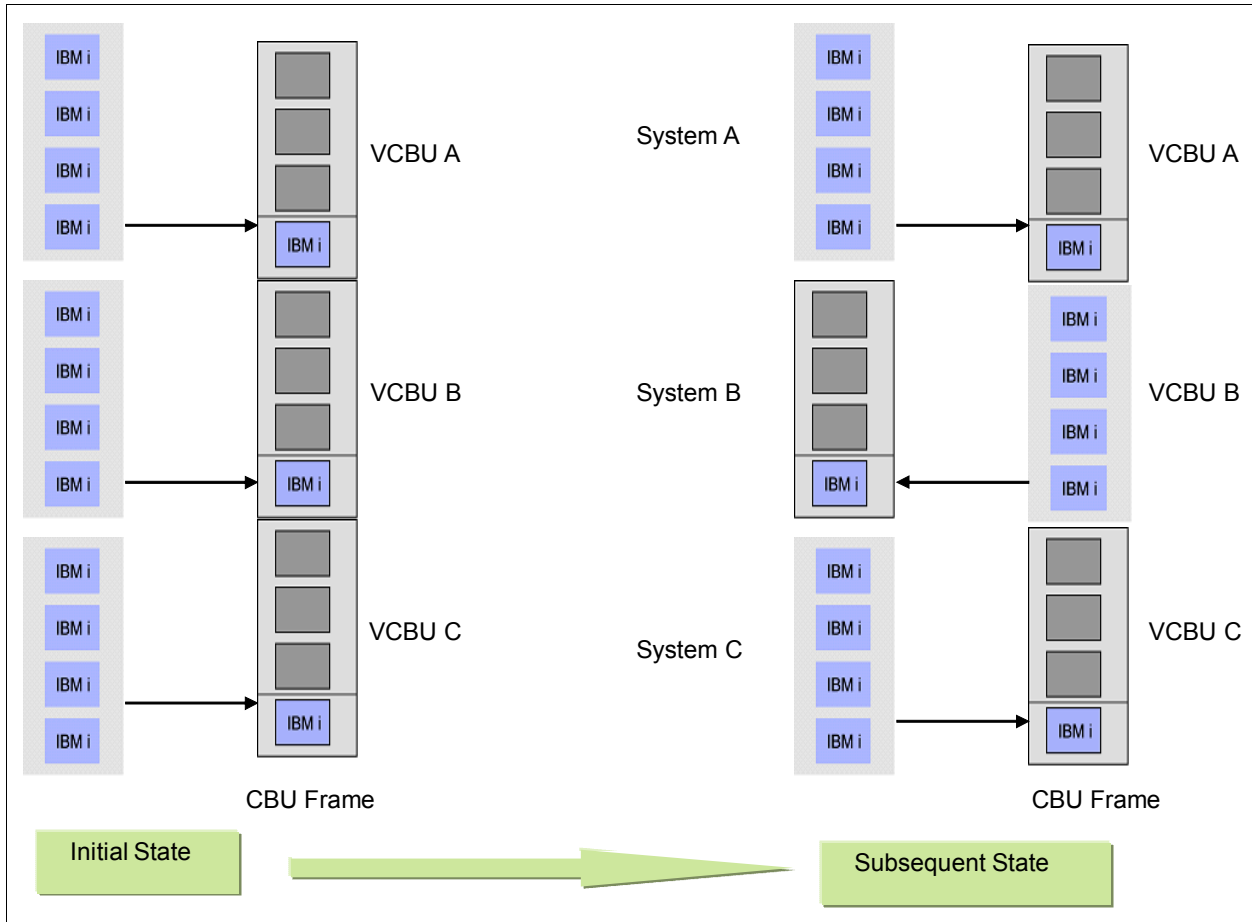

Figure D-4   Multiple primary systems in a single enterprise that uses virtual CBU

A Managed Service Provider (MSP) also can use the CBU concept for its customers. The basic setup as shown in Figure D-5 uses an individual primary system for each customer and individual CBU systems for each of the primary systems.

> **Important:** In this setup, all primary systems and all CBU systems must be owned by the MSP. You are not allowed to pair a primary system that is owned by an end customer with a CBU system that is owned by an MSP.

Figure D-5   MSP setup that uses basic CBU topology

It might be feasible for an MSP to combine the individual backup systems into one larger consolidated backup. This setup is also supported by a CBU edition as shown in Figure D-6.

Important: Again, all systems in this setup must be owned by one company. You are not allowed to have an environment where the primary systems are owned by the end customers and the backup system is owned by an MSP.

Each of the primary systems must be registered to the CBU edition. In this setup, you are required to use the concept of a virtual CBU (vCBU). For each of the primary systems, a vCBU LPAR must be set up on the backup system. Each of these vCBU LPARs requires at least one per core license of IBM i. Each of the primary systems can fail over to the combined backup system individually. However, only full IBM i per core licenses can be temporarily moved from the primary system to the CBU system. It is not possible to move, for example, one half of a license to the CBU edition and run the other half of the license on the primary system.

The Electronic Service Agent must run on each of the primary systems and on the CBU system.
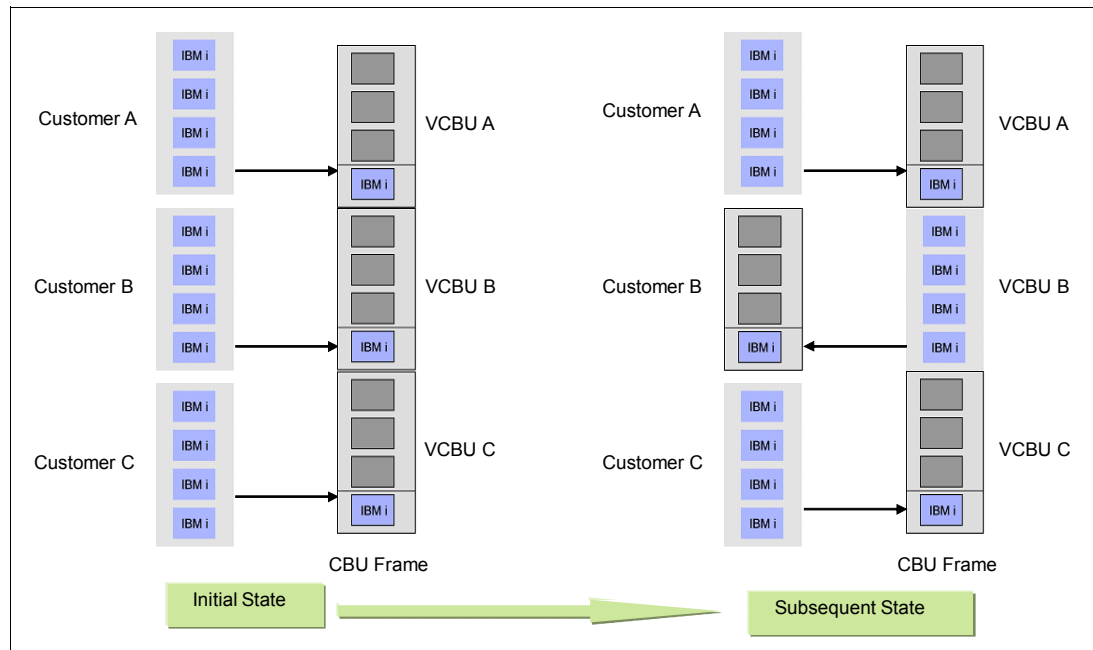

Figure D-6   MSP setup that uses virtual CBU

The use of virtual CBU is also possible in a setup where an MSP runs workloads from several customers in different LPARs of one large primary system. Each of the virtual CBU LPARs on the CBU system requires at least one IBM i per core license. See Figure D-7.

The Electronic Service Agent must run on each of the primary systems and on the CBU system.
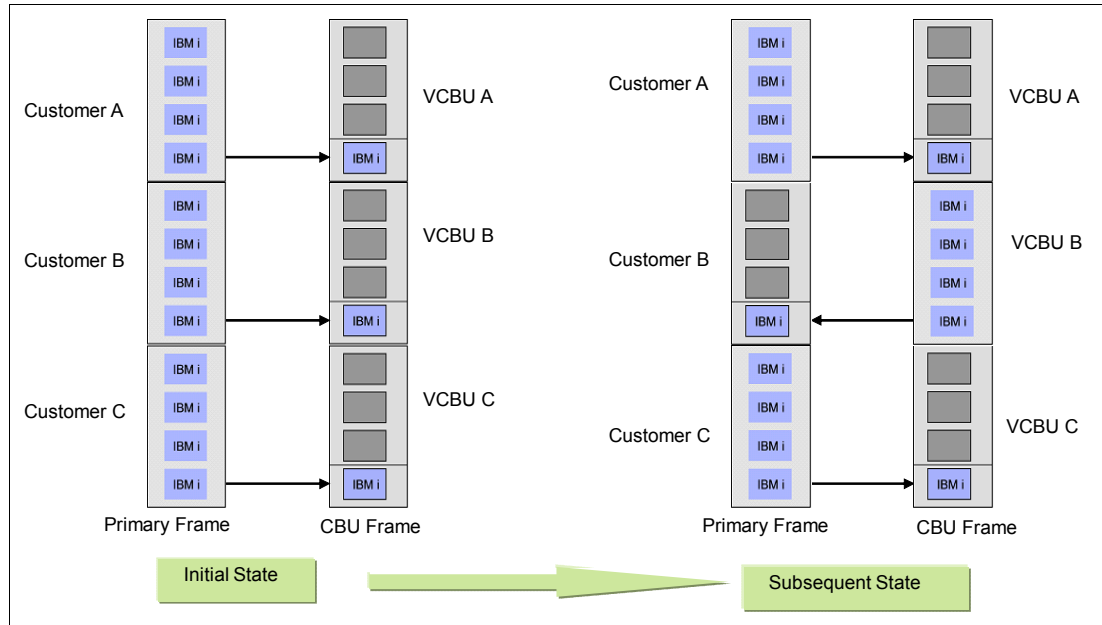

Figure D-7   MSP setup that uses a single primary system for multiple customers

It is also possible to run a three-site topology by using one primary system and attaching it to two separate CBU editions as shown in Figure D-8. This setup enables configurations where a local HA solution runs with a remote DR solution that replicates data to a third system. In this case, at least one base IBM i per core license is required on each of the CBU systems.
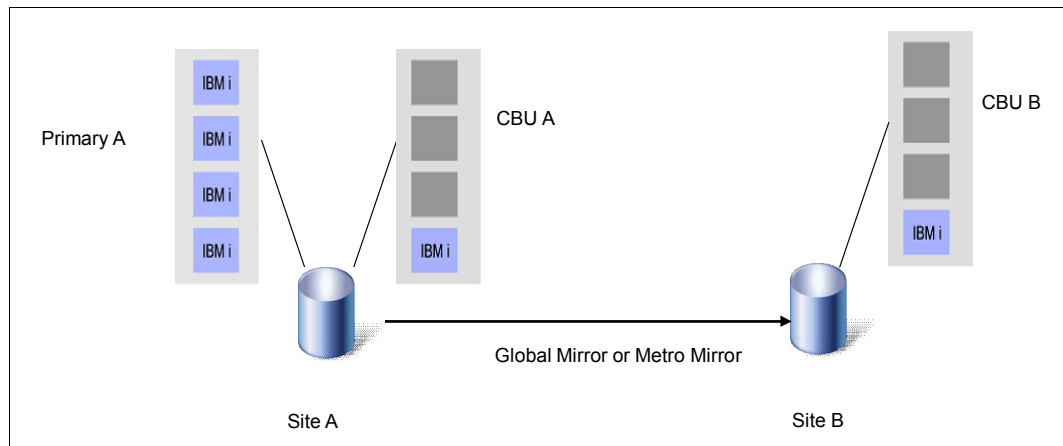

Figure D-8   Three-site topology with two CBU systems

Worksheet to configure the cluster framework

This appendix provides a worksheet to assist with cluster configuration as described in Chapter 4, "Creating the cluster framework" on page 69.

Worksheet for configuring the cluster framework

Use the worksheet in Table E-1 to configure your cluster. It is a handy reference when you complete the commands.

> **Tip:** We suggest that you do not use "primary" and "backup" as part of the naming conventions because these names are confusing after a switch is performed. After you configure the names, you cannot change the names unless you delete the entire configuration and start over.

For cluster communications, which are also known as the *heartbeat*, a minimum configuration of one IP address is required. However, for redundancy, it is considered a preferred practice to use a shared Ethernet adapter, a virtual IP address across two Ethernet adapters, or two IP addresses with two Ethernet adapters. In addition, all cluster IP addresses on a specific node must be able to communicate with all cluster IP addresses on all nodes in the cluster.

Also, in Table E-1, only the Configure (CFG), Create (CRT), and Add (ADD) type CL commands are shown.

Table E-1   Worksheet for configuring the cluster framework

Parameter	Keyword and type	Value	Description	Commands where the parameter is used
IASP	ASPDEV CHAR(10)		Independent auxiliary storage pool (IASP) name or device description.	CFGDEVASP
Cluster	CLUSTER CHAR(10)		Cluster name.	CRTCLU ADDCLUMON ADDDEVDMN CRTCAD ADDCADMRE
Node identifier (primary)	NODE CHAR(8)		Primary or source node name.	CRTCLU ADDCLUMON ADDDEVDMNE CRTCAD
Node identifier (backup)	NODE CHAR(8)		Backup node or target node name.	CRTCLU ADDCLUMON ADDDEVDMNE CRTCAD
Node identifier (additional)	NODE CHAR(8)		Additional node name if required.	CRTCLU ADDCLUMON ADDDEVDMNE CRTCAD
IP address (primary node)	Dotted IP		Cluster IP address for the primary node. It can have one or two addresses.	CRTCLU
IP address (backup node)	Dotted IP		Cluster IP address for the backup node. It can have one or two addresses.	CRTCLU ADDCLUNODE

Parameter	Keyword and type	Value	Description	Commands where the parameter is used
IP address (additional node)	Dotted IP		Cluster IP address for the additional node. It can have one or two addresses. Not applicable for all configurations.	CRTCLU ADDCLUNODE
Device domain	DEVDMN CHAR(10)		Device domain name.	ADDDEVDMNE
Cluster administrative domain	ADMDMN CHAR(10)		Cluster administrative domain name.	CRTCAD ADDCADMRE

The following information describes the parameters:

► IASP: The name of the IASP device that contains the initial source copy of the independent auxiliary storage pool (IASP). For more information about creating the IASP, see Chapter 2, "Independent auxiliary storage pools (IASPs)" on page 13.

► Cluster: A meaningful name that is up to 10 characters that identifies the unique cluster. The nodes will be added to the cluster.

► Node identifier: A meaningful name that is up to eight characters that represents the system or partition. A minimum of two nodes must be defined. The primary or source node normally contains the production or source IASP. The backup or target node normally contains the backup or target IASP.

► IP address: One or two IP addresses are configured on each node that is used for cluster communications. For a production environment, we suggest (but it is not required) that you use two dedicated, redundant Ethernet ports with the cluster IP addresses.

> **Reminder:** All cluster IP addresses on a specific node must be able to communicate with all cluster IP addresses on any other node in the cluster.

► Device domain: A meaningful name that is up to 10 characters that identifies the device domain within the cluster. The nodes and IASP will belong to the device domain.

► Cluster administrative domain: A meaningful name up to 10 characters that identifies the administrative domain for the nodes of the cluster.

F

Solutions that are hosted by IBM i

This appendix provides information and considerations when you use partitions that are hosted by IBM i in an IBM PowerHA SystemMirror for i environment.

Using IBM i hosting in a PowerHA environment to perform full system replication

IBM i hosting environments can be used to deploy IBM i, AIX®, or Linux partitions that are attached to IBM i by using virtual Small Computer System Interface (vSCSI) connections. In all of these cases, storage is provided to the guest (or hosted) environment by using network server storage spaces and a network server (NWS) description. Because these network server storage spaces from an IBM i perspective are simply objects that are located in the integrated file system (IFS), it is possible to move these storage space objects into an independent auxiliary storage pool (IASP) and use PowerHA to provide full system replication of these hosted environments.

Although IBM i hosting in a PowerHA environment is not advised for large production IBM i, AIX, or Linux partitions, it is a cost-effective option for smaller environments.

> **Important:** The use of a single hosting partition with this solution is not considered high availability (HA), but it is considered disaster recovery (DR). Limitations exist with this solution when it is compared to the use of PowerHA with an IASP implementation.

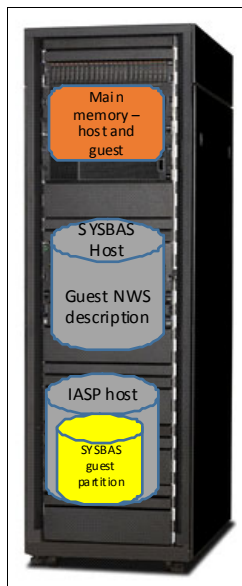Figure F-1 shows a graphical depiction of a base IBM i hosting configuration.

Figure F-1   Depiction of an IBM i hosting solution

Consider the following information when you use an IBM i hosting environment:

► By default, the hosting system will be a partitioned system. Therefore, a Hardware Management Console (HMC) or Virtual Partition Manager is required. See the following websites for information about creating hosted partitions:

  – https://ibm.biz/Bd4VhG

  – https://ibm.biz/Bd4Vhn

► The i hosting partition runs the PowerHA software and contains the IASP. We suggest that the hosting partition does not run any other production workloads that might affect disk performance. The guest partition is unaware of the PowerHA environment.

- ► Although the network server storage space can reside in an IASP, the network server description that is needed for virtual SCSI cannot. Therefore, ensure that you use the administrative domain to keep the network server descriptions synchronized between all nodes in your cluster.
- ► Virtual SCSI adapters must be defined on the production node host partition and on the backup node host partition to connect to the network storage spaces for the guest partition.
- ► The host partition requires more overall disk capacity and more bandwidth for replication when it is compared to a single partition that uses PowerHA.
- ► The guest partition does not require an IASP. Therefore, it requires little or no changes to the production system settings or applications. Because the backup copy is an exact image of the production copy, any changes that are made (for example, program temporary fixes (PTFs), updates, or application installations) to the production copy are mirrored to the backup copy.
- ► The network server description on the backup node must be varied off and the hosted partition on the backup node must be shut down while replication occurs.
- ► First, ensure that you end all activities in your guest environment before you end the host partition. Then, power down the guest partition and vary off the network server description on the host partition.
- ► Detaching the IASP from the host partition without bringing down the guest partition and bringing up the "backup" partition results in an abnormal IPL on that partition. This situation is similar to a single system that experiences a sudden power loss. Resuming replication after this situation occurs can take an extensive amount of time, depending on the jobs that were running on the production guest partition. For this reason, we do not advise that you use the backup partition copy for daily or weekly backup procedures.
- ► For a planned switch, end all activities in your hosted environment, power down the guest partition, and vary off the network server description before you perform the switch.
- ► After a planned switch or a failover, you must vary on the network server description on the backup host partition before you start the hosted environment.

> **Note:** Because the backup copy is an exact duplicate of the primary copy, it might be necessary to enter valid license keys for IBM and third-party products before they will function.

- ► We suggest that you perform a switch test before you move the environment to a production status to ensure that the setup works on the backup site (that is, that you performed the necessary configuration steps on the backup site correctly).

> **Important:** Journaling must be used in the hosted partition to help mitigate the occurrence of damaged objects if the host partition fails suddenly.

Figure F-2 shows a depiction of the replication of a hosted guest partition to a backup system.
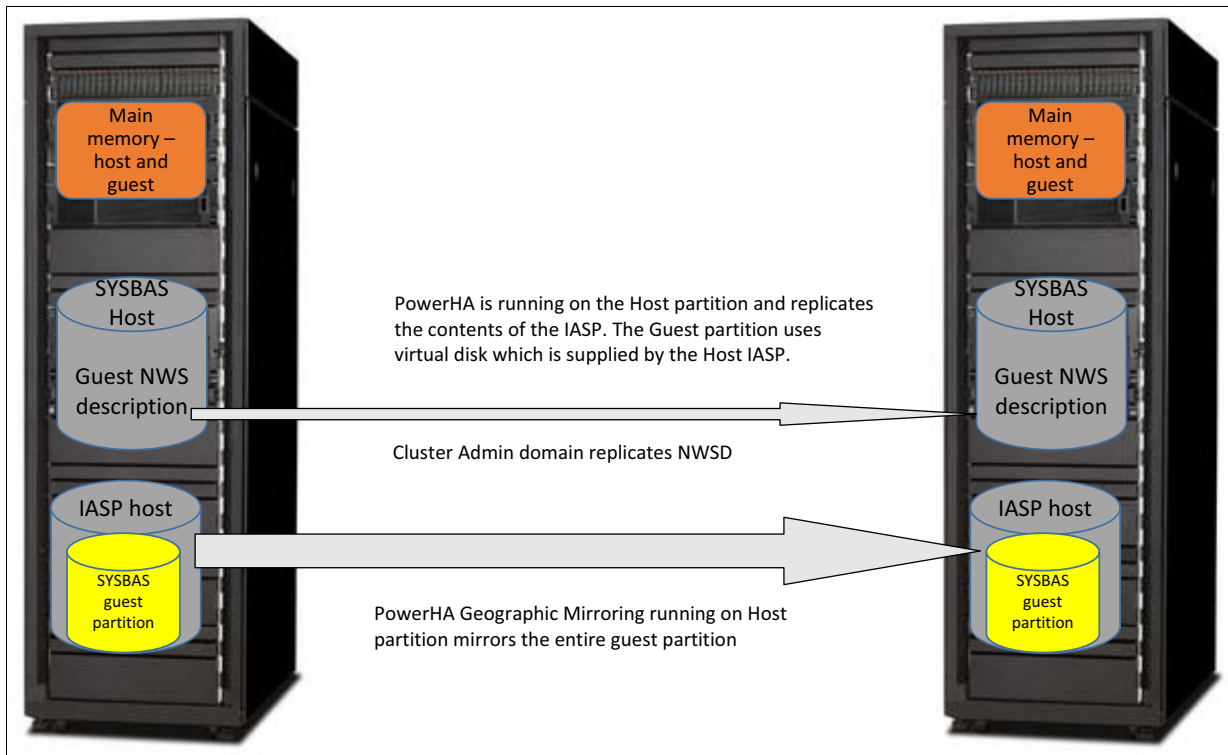

Figure F-2   IBM i hosting configuration with PowerHA geographic mirroring

The suggested configuration for HA of the hosting partition is to use dual hosts that connect to mirrored copies of the IASP. Then, the hosted or guest partition is configured with virtual SCSI connections to each of the dual hosts.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

► *IBM PowerHA SystemMirror for i: Using DS8000 (Volume 2 of 4)*, SG24-8403

► *IBM PowerHA SystemMirror for i: Using IBM Storwize (Volume 3 of 4)*, SG24-8402

► *IBM PowerHA SystemMirror for i: Using Geographic Mirroring (Volume 4 of 4)*, SG24-8401

► *Journaling - User ASPs Versus the System ASP*, TIPS0602

► *Soft Commit: Worth a Try on IBM i5/OS V5R4*, TIPS0623

► *The Journal Recovery Count: Making It Count on IBM i5/OS*, TIPS0625

► *Journal Caching: Understanding the Risk of Data Loss*, TIPS0627

► *Journaling - Configuring for your Fair Share of Write Cache*, TIPS0653

► *Journaling at Object Creation with i5/OS V6R1M0*, TIPS0662

► *Striving for Optimal Journal Performance on DB2 Universal Database for iSeries*, SG24-6286

► *Sharing Processor and Memory Activations Dynamically Among IBM Power Systems Enterprise Class Servers*, TIPS1169

► *Power Enterprise Pools on IBM Power Systems*, REDP-5101

You can search for, view, download or order these documents and other Redbooks, Redpapers™, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Other publications

These publications are also relevant as further information sources:

► Guide to fixes:

https://ibm.biz/Bd4VDS

Online resources

These websites are also relevant as further information sources:

- ► IBM Support Portal:

  https://ibm.biz/BdXqvs

- ► IBM i 7.2 Knowledge Center:

  https://ibm.biz/Bd4Vd8

- ► IBM PowerHA SystemMirror for i DeveloperWorks page:

  https://ibm.biz/Bd4VdJ

- ► IBM Full System Copy Services Manager (FSCSM) for PowerHA DeveloperWorks page:

  https://ibm.biz/Bd4vwr

- ► IBM Support Fix Central website:

  https://ibm.biz/Bd4L7N

- ► Latest group PTF for IBM PowerHA SystemMirror for i:

  https://ibm.biz/Bd4VDm

- ► IBM Recommended Fixes:

  https://ibm.biz/Bd4VDb

- ► IBM i Recommended Fixes - High Availability:

  https://ibm.biz/Bd4VDp

- ► MustGather: How To Obtain and Install QMGTOOLS:

  http://www.ibm.com/support/docview.wss?uid=nas8N1011297

- ► Geographic Mirroring: Listing of Most Common Return Codes for MSGCPFBA48:

  http://www.ibm.com/support/docview.wss?uid=nas8N1020895

- ► IBM DB2 for i: Code examples:

  https://ibm.biz/Bd4Vh8

- ► IBM Systems Lab Services:

  http://ibm.com/systems/services/labservices

- ► IBM Capacity Backup for Power Systems:

  https://ibm.biz/Bd4VgK

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Redbooks

IBM PowerHA SystemMirror for i: Preparation