Redbooks
ibm.com/redbooks

IBM PowerHA SystemMirror and IBM VM Recovery Manager Solutions Updates

Dino Quintero
Tim Simon
Felipe Bessa
Shawn Bodily
Carlos Jorge Cabanas Aguero
Vera Cruz
Sachin P. Deshmukh
Dishant Doriwala
Alexander Ducut
Karim El Barkouky
Ash Giddings
Santhosh S Joshi
Youssef Largou
Jean-Manuel Lenez
Juan Prada Diez
Vivek Shukla
Kulwinder Singh
Antony Steel
Yadong Yang

Power Systems

In partnership with
IBM **Academy of Technology**

IBM

Redpaper

IBM Redbooks

IBM PowerHA SystemMirror and IBM VM Recovery Manager Solutions Updates

May 2023

REDP-5694-00

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

First Edition (May 2023)

This edition applies to the following products:
IBM i V7R3
IBM i V7R4
IBM i V7R5
IBM AIX 7.2.5.1
IBM AIX 7.2.7
IBM AIX 7.3.0.1
IBM AIX 7.3 SP2
IBM AIX 7.2 TL5 SP4
IBM PowerHA EE 7.2.7 and 7.2.6
VM Recovery Manager 1.7

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®
Cognos®
Db2®
DS8000®
FlashCopy®
GDPS®
HyperSwap®
IBM®
IBM Cloud®
IBM Cloud Pak®
IBM FlashSystem®
IBM Security®
IBM Spectrum®
Parallel Sysplex®
POWER8®
POWER9™
PowerHA®
PowerVM®
Redbooks®
Redbooks (logo)  ®
Storwize®
SystemMirror®
WebSphere®
XIV®

The following terms are trademarks of other companies:

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

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

Red Hat, OpenShift, RHCSA, are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

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

Preface

This IBM® Redpaper publication provides an overview of the business continuity solutions for applications running on IBM Power. IBM Power servers are designed for availability, reducing single points of failure (SPOFs), and providing redundant hardware components to minimize hardware failures that interrupt critical business functions. However, a highly available (HA) platform is not a guarantee of 100% availability, so IBM also developed software and strategies to help you to continue to run your business in the rare event of hardware, software, or site failures that impact your production environment.

This publication provides an overview of the high availability and disaster recovery (HADR) solutions that you can use to keep your business running with minimal interruption if an outage occurs. It describes the features that were introduced with the latest version of those products.

This paper is written for users, even those without strong technical skills, who are looking for HADR solutions for their applications running in an IBM Power environment and want to understand the current IBM offerings.

Authors

This paper was produced by a team of specialists from around the world working at IBM Redbooks, Austin Center.

Dino Quintero is a Systems Technology Architect with IBM Redbooks®. He has 28 years of experience with IBM Power technologies and solutions. Dino shares his technical computing passion and expertise by leading teams developing technical content in the areas of enterprise continuous availability, enterprise systems management, high-performance computing (HPC), cloud computing, artificial intelligence (AI) (including machine and deep learning), and cognitive solutions. He is a Certified Open Group Distinguished Technical Specialist. Dino is formerly from the province of Chiriqui in Panama. Dino holds a Master of Computing Information Systems degree and a Bachelor of Science degree in Computer Science from Marist College.

Tim Simon is an IBM Redbooks Project Leader in Tulsa, Oklahoma, US. He has over 40 years of experience with IBM, primarily in a technical sales role working with customers to help them create IBM solutions to solve their business problems. He holds a BS degree in Math from Towson University in Maryland. He has worked with many IBM products, and has extensive experience creating customer solutions by using IBM Power, IBM Storage, and IBM zSystems throughout his career.

Felipe Bessa is an IBM Brand Technical Specialist and Partner Technical Advocate on IBM Power. He works for IBM Technology in Brazil and has over 25 years of experience in the areas of research, planning, implementation, and administration of IT infrastructure solutions. Before joining IBM, he was recognized as a Reference Client for IBM Power Technologies for SAP and SAP HANA, IBM PowerVC, IBM PowerSC, Monitoring and Security, IBM Storage, and the Run SAP Like a Factory (SAP Solution Manager) Methodology. He was chosen as an IBM Champion for IBM Power for 2018 - 2021.

Shawn Bodily is seven-time IBM Champion for IBM Power. He is a Senior IT Consultant for Clear Technologies in Dallas, Texas. He has 29 years of IBM AIX® experience, with the last 25 years specializing in HADR that is primarily focused on IBM PowerHA®. He has written and presented extensively about high availability (HA) and storage at technical conferences, webinars, and onsite to customers. He is an IBM Redbooks Platinum Author who has co-authored over a dozen IBM Redbooks publications and IBM Redpaper publications.

Carlos Jorge Cabanas Aguero has been a consultant with IBM Technology Lifecycle Services in Argentina for the last 11 years. Before this role, he worked in various roles in the IT industry supporting specializing in IBM AIX and other UNIX solutions. He has extensive experience supporting IBM AIX and Linux on IBM Power, including HADR solutions and performance tuning.

Vera Cruz is a consultant for IBM Power in IBM ASEAN Technology Lifecycle Services. She has 28 years of IT experience doing implementation, performance management, HA and risk assessment, and security assessment for IBM AIX and IBM Power across diverse industries, including banking, manufacturing, retail, and government institutions. She has been with IBM for 8 years. Before joining IBM, she worked for various IBM Business Partners in the Philippines and Singapore working as Tech Support Specialist and Systems Engineer for IBM AIX and IBM Power. She holds a degree in Computer Engineering from the Cebu Institute of Technology University in Cebu, Philippines.

Sachin P. Deshmukh is the Global Power and AIX Platform Lead for Kyndryl, and is based in the US. His areas of expertise include IBM AIX operating system (OS) provisioning and support, IBM PowerHA, virtualization, and cloud. He provides guidance, oversight, and assistance to global delivery teams that support Kyndryl accounts. As a member of the Critical Response Team, he works on major incidents and high severity issues. He participates in proactive Technical Health Checks and Service Management Reviews. He interacts with automation, design, procurement, architecture, and support teams for setting delivery standards and creating various best practices documentation. He creates and maintains the IBM AIX Security Technical Specifications for Kyndryl. He is certified on various other platforms, such as AWS Solutions Architect (Associate), AWS Cloud Practitioner, and Red Hat Certified System Administrator (RHCSA). Before moving to Kyndryl in 2021, he was with IBM since 1999. He is closely associated with IBM AIX and the IBM Power platform for close to 30 years.

Dishant Doriwala is a Senior Staff Software Engineer, and a Test Lead for the VM Recovery Manager (VMRM) for High Availability and Disaster Recovery product. He works at IBM Systems Development Labs, Hyderabad, India, and has 10 years of experience in th industry with expertise in testing HADR products. He has experience working with various HADR solutions, such as IBM PowerHA SystemMirror®, Reliable Scalable Cluster Technology (RSCT), VMRM, and Geographic Logical Volume Manager (GLVM). He has expertise with enterprise storage, such as IBM SAN Volume Controller (SVC) and IBM Storwize®, Hitachi, Dell EMC Symmetrix Remote Data Facility (SRDF), and IBM XIV®. He has authored white papers, IBM Redpaper publications, and technical articles for the HADR domain. He holds a Master's in Software Technologies degree from International Institute of Information Technology, Pune, and a Bachelor’s in Electronics and Communications degree from VTU, Bangalore.

Alexander Ducut has been with IBM for 26 years in different roles, such as Brand Technical Specialist, Technical Support, Client Technical Architect, Services Delivery Manager, and Client Technical Manager. He designed and implemented several complex projects that involve IBM servers and storage with HADR across different industries. He co-authored three IBM Redbooks publications. He helps clients address their digital transformation initiatives with infrastructure, cloud, and application modernization through IBM Power based solutions.

Karim El Barkouky is a Senior IT Management Consultant. He works in MEA - Technology Services - Lab Services. He worked at IBM Systems as L2 remote support as a global IBM PowerHA SME in Cairo, Egypt. He has 8 years of experience in the industry with expertise in several implementations, and consultancy tasks for various HA solutions, such as IBM PowerHA SystemMirror, IBM Spectrum® Scale, Cluster Aware AIX (CAA), RSCT, GLVM, VMRM, SUSE Linux Enterprise Server - SUSE/HA extension, and container orchestrators like Red Hat OpenShift. He is a recognized trainer that delivered various IBM AIX and HA training sessions across the MEA region. He has experience with IBM Power servers and the IBM Power software family, such as IBM PowerVM®, IBM PowerVC, IBM PowerSC, and Linux on IBM Power.

Ash Giddings is based in the UK and is a Product Manager for Maxava. With over 30 years of experience, his areas of expertise include IBM i HA, systems management, and performance analysis and tuning.

Santosh S Joshi is a Senior Staff Software Engineer at IBM India Systems Development Lab, IBM India. He has over 18 years of experience in the information technology field. He works for the IBM PowerVM Live Partition Mobility (LPM) development team, and before this role, he worked for IBM VM Recovery Manager High Availability and Disaster Recovery for IBM Power solution development. He holds a Bachelor of Engineering degree in Electronics and Communication from Visvesvaraya Technological University, Belagavi India. His areas of expertise include PowerVM virtualization, HADR solutions, clustering technologies, and SAN**.**

Youssef Largou is the founding director of PowerM, a platinum IBM Business Partner in Morocco. He has 21 years of experience in systems, HPC, middleware, and hybrid cloud, including IBM Power, IBM Storage, IBM Spectrum, IBM WebSphere®, IBM Db2®, IBM Cognos®, IBM WebSphere Portal, IBM MQ, Enterprise Service Bus (ESB), IBM Cloud Pak®, and Red Hat OpenShift. He has worked within numerous industries with many technologies. Youssef is an IBM Champion for 2020 - 2022, and an IBM Redbooks Gold Author. He designed many reference architectures. He is a five-time IBM Beacon Award Finalist in Storage, Software-Defined Storage, and LinuxONE. He holds an Engineer’s degree in Computer Science from the Ecole Nationale Supérieure des Mines de Rabat and Excecutif MBA from EMLyon.

Jean-Manuel Lenez has been a pre-sales engineer since 1999 with IBM Switzerland. He specializes in UNIX, IBM Power, IBM AIX, and IBM i server technologies, and associated products such as PowerVM, PowerHA, PowerSC, Linux on IBM Power, and IBM Cloud®. He is heavily involved in his pre-sales mission, where he leads projects with major customers around various subjects, such as AI, deep learning, SAP HANA, server consolidation, and HADR.

Juan Prada is an IBM i Senior System Administrator in Madrid, Spain. He has over 18 years of experience with IBM i administration, and experience with PowerVM solutions and IBM Storage (the IBM DS8000® family, Storwize, SVC, and IBM FlashSystem®). He has worked for IBM Business Partners and for customers in the financial and retail sectors.

Vivek Shukla is a Presales Consultant for cloud, AI, and cognitive offerings in India. He is an IBM Certified L2 (Expert) Brand Technical Specialist. He has over 20 years of IT experience in infrastructure consulting, IBM AIX, and IBM Power servers and IBM Storage implementations. He has hands-on experience with IBM Power servers, IBM AIX, system software installation, Requests for Proposals, Statements of Work, sizing, performance tuning, root cause analysis, DR, and mitigation planning. He has written several FAQs for IBM Power, and is a Worldwide Focus for Techline Flash FAQs. He holds a master's degree in Information Technology from IASE University, and a bachelor's degree (BTech) in Electronics & Telecommunication Engineering from IETE, New Delhi. His areas of expertise include Power Enterprise Pools, Red Hat OpenShift, IBM Cloud Pak®, and hybrid cloud.

Kulwinder Singh is a Technical Support Professional with IBM India Systems Development Lab, IBM India. He has over 25 years of experience in the IT infrastructure management. He supports customers as IBM AIX L2 development support for IBM AIX, PowerHA, and IBM VM Recovery Manager High Availability and Disaster Recovery on Power. He holds a Bachelor of Computer Application degree from St. Peter's University. His areas of expertise include IBM AIX, HADR solutions, IBM Spectrum Protect, and SAN.

Antony Steel is a senior technical staff member working with IBM Australia. A research chemist by training, he brings a unique experience and perspective with over 30 years of experience in the IT industry as a programmer, customer, and IBM Business Partner. For over 20 years, he was with IBM Australia and Singapore as Senior Managing Consultant / Advanced Technical Support. Antony’s customers include users, senior management, and other key stakeholders in a range of industries, including some of the largest financial and business institutions and government departments in Australia, New Zealand, and the Asia Pacific region. His areas of interest are IBM AIX, HADR, and clustering. He is an IBM Champion who has assisted with preparing HA and IBM AIX certification exams.

Yadong Yang is an IBM IT Management Consultant on IBM Power. He works for IBM Technology Services (formerly IBM STG Lab Services) in the US. He has worked for IBM for about 20 years. His expertise includes IBM PowerVM, PowerHA SystemMirror, VMRM, IBM AIX System Administration, IBM AIX Performance, Linux on Power, IT infrastructure architecture, and SAN storage. He has more than 26 years of experience with IBM AIX. He has a Ph. D. degree in Mathematics from North Carolina State University.

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

Jeff Boleman
Principal Product Manager, IBM PowerVS IaaS
IBM Systems, Vera Cruz, CA

Ramya Bommineni
IBM - VMRM Development Partner
IBM Systems, IBM India

Uma Maheswara Rao Chandolu
Director - PowerHA IBM AIX and VMRecovery Manager HADR
IBM Systems, IBM India

Jes Kiran Chittigala
HADR Architect for IBM Power VMRM, Master Inventor
IBM India System Development Labs

Steven Finnes
IBM Power HA Product Offering Manager
IBM Systems, Rochester, MN

Abhilash Kadivendi
IBM - VMRM Development Partner
India

Adhish Kapoor
VMRM Developer
IBM India System Development Labs

Brian Nordland
IBM i High Availability
IBM Systems, US

Pandi Jai Sree
IBM - VMRM Development Partner
India

Srikanth Thannerru
Advisory Software Engineer, VMRM
IBM India System Development Labs

Douglas Yakesch
IBM Power User Technologies Build Tools Team Lead
IBM Systems, Austin, TX

Vijay Yalamuri
IBM India

Tom Weaver
PowerHA, CAA - IBM AIX Development Support
IBM Systems Technology Lifecycle Services, Austin, TX

Scot Stansell
IBM AIX Technical Specialist - PowerHA Team
IBM Systems Technology Lifecycle Services, Coppell, TX

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 IBM Redbooks 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 papers to be as helpful as possible. Send us your comments about this paper 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, IBM Redbooks
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find 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

Introducing high availability and disaster recovery

This chapter defines common concepts and their associated terms that are used when planning and implementing high availability and disaster recovery (HADR) solutions with IBM Power, either on-premises or in the cloud.

Although many of these concepts are generic and applicable across most IT Infrastructures, we focus on the IBM HADR products for the IBM Power platform.

This chapter describes the following topics:

- Introduction to high availability and disaster recovery
- Key recovery objectives
- High availability and Disaster recovery
- Cloud and hybrid cloud disaster recovery
- Assessing and designing for continuous operations

1.1 Introduction to high availability and disaster recovery

Today's enterprises cannot afford system outages. Even a few minutes of application downtime can result in considerable financial loss, erosion of customer confidence, damage to your brand image, and possible public relations problems.

To better control and manage their IT infrastructure, many enterprises concentrate their IT operations into large (and on-demand) data centers. These data centers must be flexible and resilient enough to handle the ups and downs of continuously available workloads. They also must manage changes and respond to threats with consistent availability, security, and privacy around the clock. To effectively create and manage solutions to address this challenge requires an integration of operating system (OS), clustering software, storage, and networking.

How a system, server, or environment handles failures is characterized as its reliability, availability, and serviceability (RAS). In today's world, the RAS of both the OS and the hardware on which it runs has assumed great importance.

Today's businesses require that IT systems be self-monitoring, self-healing, maintainable without outages, and support continuous operations. Customer's IT systems are meeting their required levels of RAS by using varying degrees of redundancy and error correction.

The RAS characteristics of a server platform can be a market differentiator, and one where IBM Power, whether running AIX, IBM i, or Linux, excels. The RAS characteristics of IBM Power are a key focus in the design and architecture of the server, and now IBM Power servers are attaining RAS levels close to the ones that are experienced only previously with mainframe systems.

RAS is often measured in terms of *nines* of availability, which is an expression of what percentage of time that the solution is available. The amount of downtime that is allowed while still meeting that percentage of uptime is shown in Table 1-1. For example, to achieve five nines of availability requires that the solution not be down more than 5 minutes and 35 seconds over a year's time.

Table 1-1 Six levels of nines and their availability times

Number of nines	Uptime%	Maximum annual downtime
Six	99.9999	31.56 seconds
Five	99.999	5 minutes 35 seconds
Four	99.99	52 minutes 33 seconds
Three	99.9	8 hours 46 minutes
Two	99.0	87 hours 36 minutes
One	90.0	36.5 days

Overview of availability solutions

There are many solutions for increasing the availability of your applications, and they provide a wide range of availability options. Table 1-2 lists various types of availability solutions and their characteristics.

Table 1-2 Types of availability solutions

Solution	Downtime	Data availability	Observations
Stand-alone	Days	From last backup	Basic hardware and software
Enhanced stand-alone	Hours	Until last transaction	Double most hardware components
High availability or highly available (HA) clustering	Seconds	Until last transaction	Double hardware and extra software costs
Fault-tolerant	Zero	No loss of data	Specialized hardware and software, and expensive

HA solutions, in general, offer the following benefits:

- Standard hardware and networking components that can be used with the existing hardware.
- Works with nearly all applications.
- Works with a wide range of storage and network types.
- Excellent availability at a reasonable cost.

The HA solution for IBM Power servers offers distinct benefits:

- Proven solution with over 29 years of product development.
- Using *off-the-shelf* hardware components.
- Proven commitment for supporting customers.
- IP version 6 (IPv6) support for both internal and external cluster communication.
- Smart Assist technology enabling HA support for many prominent applications.
- Flexibility, that is, virtually any application running on a stand-alone AIX system can be protected with PowerHA.

When implementing a HA solution, consider the following aspects:

- Implementation requires a carefully created design and detailed planning from end to end.
- Focus on eliminating of single points of failure (SPOFs).
- Carefully select the appropriate hardware.
- Use proper implementation practices. Do not take shortcuts.
- Requires disciplined system administration practices and change control.
- Documented operational procedures.
- Comprehensive test plan and periodic scheduled tests.

1.1.1 Fault tolerance

Although many organizations are blurring the meaning of “fault-tolerant systems”, we define a fault-tolerant environment as one that has no service interruption and a higher cost. A HA environment is one that has only a minimal service interruption.

Fault tolerance relies on specialized hardware to detect a hardware fault and instantaneously switch to a redundant component, whether the failed component is a processor, memory board, power supply, I/O subsystem, or storage subsystem. This cutover is seamless and not detectable by the user, and thus offers non-stop service. A high premium is paid in both hardware cost and performance, with limited software options. HA solutions are designed to work with a far greater number of applications. The fault-tolerant model does not address software or human failures, the latter of which has consistently been reported as contributing to most downtime experience by organizations.[1][2]

The HA model views availability not as a series of replicated physical components, but rather as a set of system-wide, shared resources that cooperate to ensure access to essential services. IBM HA combines software with industry-leading hardware to minimize downtime by quickly restoring essential services when a system, component, or application fails. Although not instantaneous, services can be restored rapidly.

Many customers are willing to absorb a small amount of downtime with HA rather than pay the higher cost of providing true fault tolerance. Also, in most HA configurations, the standby resources are available for use during normal operation, which ensures that redundant components operate correctly when required and gives organizations the option to use them to run less critical workloads, such as development or test.

HA systems fill the niche for organizations whose applications can withstand a short interruption and can be recovered automatically from an unexpected halt. This option is not recommended for the few industries that are life-critical and have applications that cannot afford any downtime at all.

1.1.2 Downtime

Downtime is any period during which an application or service is unavailable to serve its clients. Downtime can be classified into two categories:

- Planned:
  - Hardware upgrades.
  - Hardware or software repair or replacement.
  - Software (OS and application) updates or upgrades.
  - Backups (offline).
  - Periodic testing is required for cluster validation.
  - Development.
- Unplanned:
  - Administrator or human errors.
  - Application failures.

[1] ITIC 2021 Global Server Hardware, Server OS Reliability Report noted that 64% of the responding organizations said human error caused outages.

[2] Uptime Institutes's 2022 Outage Analysis reported that nearly 40% of organizations surveyed suffered a major outage that was caused by human error over the past 3 years.

- Hardware failures.
- OS errors.
- Environmental disasters.

Downtime is often associated with unplanned outages, but some downtime is the result of a planned outage. Planned outages are necessary to help maintain systems and minimize the risk of an unplanned outage.

Uptime is a percentage of the amount of time that a system's services are available, so anything less than 100% means that some downtime occurred. Any downtime, planned or unplanned, counts against total uptime. A planned and implemented HA solution can help minimize, mask, or prevent outages for planned maintenance.

Typically, organizations view their applications in terms of *recovery time objective (RTO)*, which is the time until service resumes, and *recovery point objective (RPO)*, which is the amount of data that is lost to set the application's service-level agreement (SLA). This step is an important one in an organization's HADR planning because RPO and RTO can be used to group their applications and determine the correct HADR solution for that group.

Figure 1-1 shows the combination of events that make up RPO and RTO.

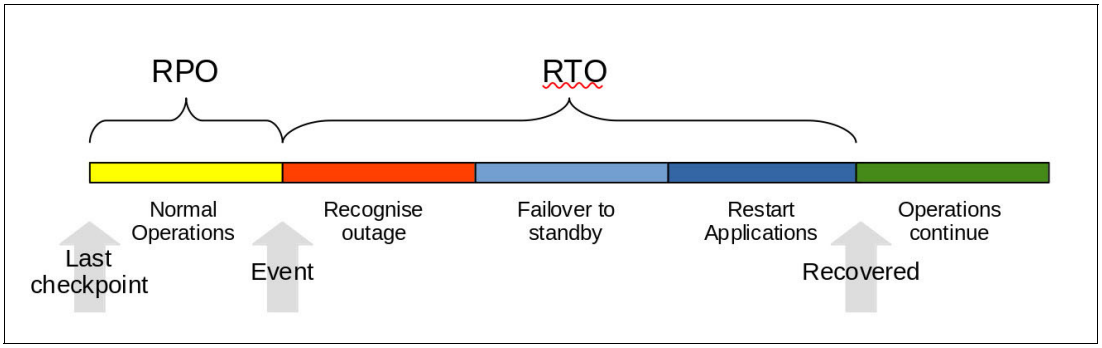

Figure 1-1 RTO and RPO

1.1.3 Single points of failure

A *SPOF* is any individual component that is integrated into a system that, if it fails, renders the application unavailable for users.

Good design can remove SPOFs in the cluster: nodes, storage, and networks. Solutions like PowerHA manage these components and also the resources that are required by the application. These resource components also include scripts to start or stop the application, which are known as application controllers.

Ultimately, the goal of any IT solution in a critical environment is to provide continuous application availability and data protection. HA is one of the building blocks that is needed to achieve the continuous operation goal. HA is based on the availability of the hardware, software (OS and its components), application, and network components.

To avoid SPOFs, use the following items:

- Redundant servers
- Redundant network paths
- Redundant storage (data) paths

- Redundant (mirrored and RAID) storage
- Monitoring of components
- Failure detection and diagnosis
- Automated application fallover
- Automated resource reintegration

A good design avoids SPOFs, and tools like PowerHA and VM Recovery Manager (VMRM) can manage the availability of the application through the individual component failures. Table 1-3 lists each object, which, if it fails, can result in loss of availability of the application. The object can be a physical or logical component.

Table 1-3 Single points of failure

Cluster object	SPOF eliminated by
Node (servers)	Multiple nodes.
Power or power supply	Multiple circuits, power supplies, or uninterruptible power supply (UPS).
Network	Multiple networks that are connected to each node, and redundant network paths with independent hardware between each node and the clients.
Network adapters	Redundant adapters, and use other HA type features, such as Etherchannel or Shared Ethernet Adapters (SEAs) through a Virtual I/O Server (VIOS).
I/O adapters	Redundant I/O adapters and multipathing software.
Controllers	Redundant controllers.
Storage	Redundant hardware, enclosures, disk mirroring or RAID technology, and redundant data paths.
Application	Configuring application monitoring and backup nodes to acquire the application engine and data.
Sites	Use more than one site for DR.
Resource groups (RGs)	An RG is a container of resources that are required to run the application. The SPOF is removed by moving the RG around the cluster to avoid failed components.

PowerHA and VMRM also optimize availability by enabling dynamic reconfiguration of running clusters. Maintenance tasks such as adding or removing nodes can be performed without stopping and restarting the cluster.

1.2 Key recovery objectives

There are a few key objectives to recoverability that must be considered when balancing the cost to the organization of the outage against the cost of the solution:

- Network Recovery Objective (NRO)

  How long it takes to switch network access or name resolution.

- Recovery scope

  Recovery scope defines which resources are part of a backup. The scope is defined according to the business goals and criticality of the business service.

- RTO

  What is an acceptable amount of time to be without system access?

  - If it is minutes to a few hours, the usage of automated recovery is imperative.
  - If it is hours to days, you may use manual recovery steps.

- RPO

  After an outage occurs, how much, if any, data is acceptable to either re-create or do without?

  - If zero, then synchronous replication is required.
  - If greater than zero, then asynchronous replication might be acceptable and can extend the distance between production and DR data centers.

- Consistency

  After a successful recovery from backup, the data must be checked for consistency. There are two major consistency concepts to consider:

  - Crash consistency

    The restored data bytes match the ones in the primary system at the time of the crash. Data in cache at the primary was not sent to the secondary system, so some log recovery might be required before the data can be used.

  - Application consistency

    Applications can access data from the time of the backup without failure.

- SLAs

  There is an agreement between the service provider and client that defines the disaster recovery (DR) strategy and design for stated business continuity and service resiliency requirements.

To answer these questions accurately, a risk and requirements analysis is required with a downtime cost analysis for each service. Organizations must go beyond stating that their DR objectives are zero across the board because this goal is often unachievable and does not recognize the different value of each application to the organization and the relative outage costs. The following sections describe the concepts of continuous availability in more detail.

1.3 Continuous operations, continuous availability, and business continuity

Enterprises are looking for an application infrastructure that is HA to continue running their business if there are problems occurring in the infrastructure. There are many terms that are used to define the ability to maintain business operations when failures occur in the infrastructure, either planned or unplanned failures, and whether they are the result of a natural disaster such as hurricane, flood, or tornado or if they are other types of failures. This section defines some of the terms that are used and describes the differences between those solutions.

1.3.1 Continuous operations

Continuous operations apply to IT environments and systems that can continuously operate and mask planned outages from users. Continuous operations employ non-disruptive hardware, software, configuration, and administrative changes.

Hardware component failure represents a proportion of overall system downtime, but planned downtime is the most common reason for downtime. All HADR solutions should help reduce planed downtime, but it is vital that downtime is planned for and scheduled in advance so that unplanned downtime can be minimized.

1.3.2 Continuous availability

Continuous availability means that customers services and applications are available whenever customers want them regardless of the hour of the day or the day of the week. IT management no longer can use only evenings and weekends to perform maintenance and backup tasks. The traditional view is that continuous availability (the elimination of downtime) is the sum of continuous operations (the masking or elimination of planned downtime) and HA (the masking or elimination of unplanned downtime).

Continuous availability entails being “always on”. From the standpoint of users, the application (such as mobile banking, digital government, and airline applications) is always available. Conversations about DR, SLAs, or planned downtime are not relevant to business application owners; instead, their view is how to make this business application always available to their clients.

Most of today’s HA solutions are based on the integration of the OS with clustering software, storage, and networking. When a failure is detected, the integrated solution triggers an event that performs a predefined set of tasks that are required to reactivate the application with access to data and network on another set of servers or storage. This function is defined as *IT continuous availability*. Scaled-out solutions that use multiple instances of the application also can provide continuous availability because the failure of a single instance does not impact the overall availability of the application.

Continuous availability is a collective term for the following characteristics of a product:

- Capable of performing its intended functions under stated conditions for a stated period (reliability).
- Ready to perform its function whenever requested (availability).
- Able to quickly determine the cause of an error and provide a solution to eliminate the effects of the error (serviceability).
- Encompasses techniques for the following goals:
  - Reducing the number of faults
  - Minimizing the effects of faults when they occur
  - Reducing the time for repair
  - Enabling the customer to resolve problems as quickly and seamlessly as possible

1.3.3 Business continuity

The terms *business continuity* and DR are sometimes used interchangeably (as are business resumption and contingency planning). Here, *business continuity* is defined as the ability to adapt and respond to risks and opportunities to maintain continuous business operations. However, business continuity solutions that are applied in one industry might not be applicable to a different industry because they can have different sets of business continuity requirements and strategies.

Business continuity is implemented by using a plan that follows a strategy that is defined according to the needs of the business. A total business continuity plan has a broader focus and includes items such as a crisis management plan, business impact analysis, human resources management, business recovery plan procedure, test plan, and documentation.

1.4 High availability

HA is an attribute of a system that provides service during defined periods at acceptable or agreed-on levels and masks both planned and unplanned outages from users. HA is possible by using redundant hardware components, automated failure detection and recovery, bypass reconfiguration, testing, problem determination, and change management procedures.

In addition, HA and its associated processes provide access to applications regardless of hardware, software, or system management issues by greatly reducing or masking planned downtime.

HA solutions help eliminate SPOFs by using appropriate design, planning, selection of hardware, configuration of software, and carefully controlled change management discipline. HA does not mean *zero* interruption to the application. HA is more synonymous with fault-*resilient* instead of fault-*tolerant.*

An HA environment often involves more demanding RTOs (seconds to minutes) and more demanding RPOs than a DR scenario. HA solutions provide fully automated failover to an alternative system so that users and applications can continue working with minimum disruption. HA solutions must provide an immediate recovery point while providing a recovery time capability that is better than the recovery time that you experience in a non-HA solution.

1.4.1 Virtual machine or LPAR restart

LPAR or virtual machine (VM) restart options, such as Simplified Remote Restart (SRR) on IBM Power, are considered an entry point HA solution that is designed for applications that can withstand a slightly longer outage. With a restart solution, the LPAR or LPARs are restarted on other servers, so the restart takes slightly longer because it includes the restart of the OS and the application. Restart solutions are also typically less complex and easier to manage than clustered solutions.

1.4.2 Clustered solutions

A *clustered solution* consists of a group of servers with an application that monitors and coordinates activities across the cluster. If there are any failures or application issues, the clustering software attempts to restart the application on an operational node with network and storage access.

1.4.3 Application or DB replication

These solutions are heavily application-dependent and rely on the application or database transferring all updates to a different server. There, the updates are played against the local copy to bring it up to a more current state. Often, these solutions can be combined with a clustered solution to provide an alternative copy of the data with the ability to delay playback to avoid operator errors or corruption.

1.4.4 Scale-out solutions

Scale-out solutions, or multiple instances of the application, can assist with both redundancy (a single instance failing will not affect the other consumers) and performance (through increasing or decreasing the number of instances as required).

1.5 Disaster recovery

For our purpose, *DR* is defined as the ability to recover a data center at a different site if a disaster destroys the primary site or otherwise renders it inoperable. The characteristics of a DR solution are that IT processing resumes at an alternative site.

DR is a coordinated activity to enable the recovery of IT and business systems in a disaster. A DR plan covers both the hardware and software that is required to run critical business applications and the associated processes that are required to recover the functions running at a site. DR for IT operations employs extra equipment, which is in a physically different location, and the use of automatic or manual actions and methods to recover all the critical business processes.

Every location, although different, has some type of disaster to worry about. Fire, tornadoes, floods, earthquakes, and hurricanes can have far-reaching geographical impacts, which drive remote disaster sites to be further apart. Some industries have regulations that set the minimum distance between sites.

Here are some important questions about designing for disasters:

- What is the monetary impact to the business in a disaster?
- How soon can the business be back in production (RTO)?
- Will data be lost? If so, how much (RPO)?
- What communication bandwidth is required and is it acceptable?
- What DR solutions are viable based on the inter-site distance requirements?
- What DR solutions are viable based on the application requirements?

DR strategies cover a range from not having a recovery plan to automatic recovery with high data integrity. Data recovery strategies must address the following issues:

- Data readiness levels:
  - Level 0

    None. No provision for DR or off-site data storage.
  - Level 1

    Periodic backup. Data that is required for recovery up to a certain date is backed up and sent to another location.
  - Level 2

    Ready to roll forward. In addition to periodic backups, data update logs are periodically sent to another location either by using physical media or electronically. The recovery point is up to the latest update log at the recovery site.
  - Level 3

    Roll forward or forward recover. A shadow copy of the data is maintained on disks at the recovery site. Data update logs are received and periodically applied to the shadow copy by using recovery utilities.
  - Level 4

    Real-time roll forward. Like rollforward, except updates are transmitted and applied while they are being logged at the original site. This real-time transmission and application of log data does not impact transaction response time at the original site.
  - Level 5

    Real-time remote update. Both the original and the recovery copies of data are updated before sending the transaction response or completing a task.
- Site interconnection options:
  - Level 0

    None. There is no interconnection or transport of data between sites.
  - Level 1

    Manual transport. There is no interconnection. For transport of data between sites, dispatch, tracking, and receipt of data is managed manually.
  - Level 2

    Remote tape. Data is transported electronically to a remote tape. Dispatch and receipt are automatic. Tracking can be either automatic or manual.
  - Level 3

    Remote disk. Data is transported electronically to a remote disk. Dispatch, receipt, and tracking are all automatic.

- Recovery site readiness:
  - Cold

    A *cold site* is an environment with the proper infrastructure, but little or no data processing equipment. This equipment must be installed as the first step in the data recovery process. Both periodic backup and ready to roll forward data can be shipped from a storage location to this site when a disaster occurs.
  - Warm

    A *warm site* has data processing equipment that is installed and operational. This equipment may be used for other data processing tasks until a disaster occurs. Data processing resources can be used to store data, such as logs. Recovery begins after the regular work of the site is shut down and backed up. Both periodic backup and ready to roll forward data can be stored at this site to expedite DR.
  - Hot

    A *hot site* has data processing equipment that is installed and operational, and the data can be restored either continually or regularly to reduce recovery time.
  - Active-active

    A subset of the applications is active in both sites concurrently.

When these components are combined, you get the seven tiers of DR, as shown in Figure 1-2 on page 13:

- Tier 0

  There is no off-site or off-site data. Recovery must be local.
- Tier 1

  Backups are only on tape, and they should be offsite. However, they are not kept at any site where hardware might be used to perform the recovery. The site can be cold, but often is a storage data vault.
- Tier 2

  Offsite backups are on tape and stored offsite at least at a warm site, but should be stored at a hot site.
- Tier 3

  Data is transmitted electronically, at least critical data, to the hot recovery site. Provides shorter recovery time of critical data and services.
- Tier 4

  Point-in-time copies, like IBM FlashCopy®, to a hot site. The copying can go both directions.
- Tier 5

  Data is continuously copied to the remote hot site by using a two-phase or two-site commit. This tier can be storage, host, or application-based replication.
- Tier 6

  From a data perspective, there is zero or near-zero data loss with instantaneous recovery. This tier is often storage-based replication.
- Tier 7

  In addition to Tier 6, automation of recovery procedures to restore the services is included. This tier is the highest level of protection that is available.

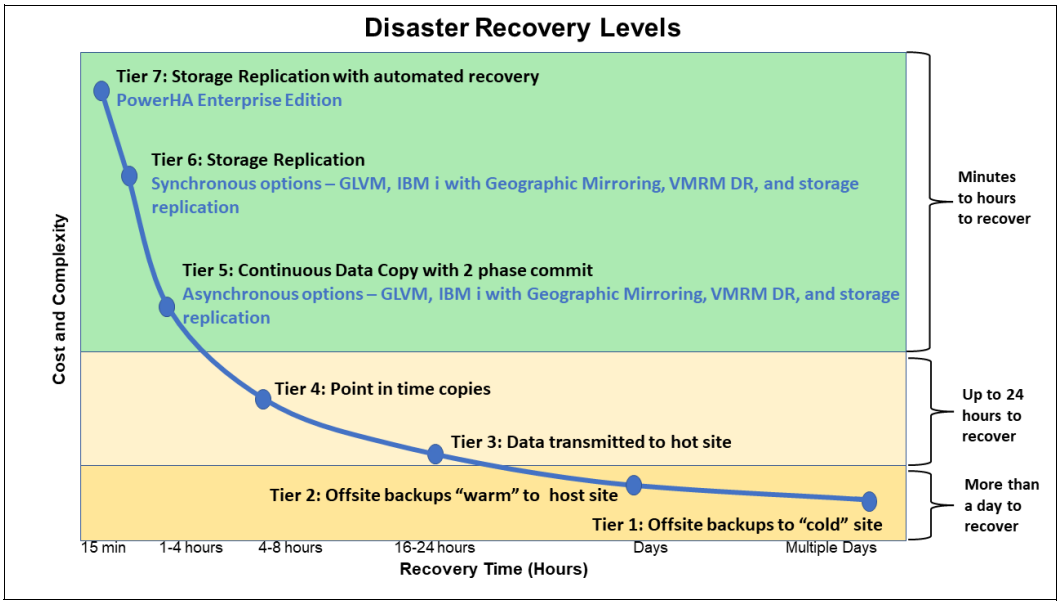

Figure 1-2 Tiers of high availability and disaster recovery

1.5.1 Factors to consider when extending HA to DR

The following general issues must be considered when planning for DR.

Data replication and throughput challenges

This section describes data replication latency and throughput challenges.

Network latency

Network latency is the time that it takes for messages to go across the network. Even when there is plenty of network bandwidth, it still takes a finite amount of time for the bits to travel over the inter-site link. The speed of the network is limited by the quality of the switches and the laws of physics, and the network latency is proportional to the distance between the sites. Even if a network can transmit data at a rate of 120 kilometers per millisecond, it still adds up over a long distance.

For example, if the sites are 60 km apart, all I/O must travel 60 km from the application to the remote storage. After the remote storage is updated, the result of the I/O request must travel 60 km back to the application. This 120 km round trip adds about 1 millisecond to each I/O request, and this time can be greater depending on the number and quality of routers or gateways that are traversed. Suppose that the sites are 4000 km apart, so each I/O request requires an 8000 km round trip, adding approximately 67 milliseconds to each I/O. The resulting application response time is usually unacceptable.

Depending on the application, synchronous mirroring might be required to meet the RPO. For metro distances, which is usually 100 km or less, most applications can tolerate the extra latency. Greater distances typically need asynchronous replication.

The critical issue is that if distance and application performance require an asynchronous solution, then the organization must be prepared to lose data because in a disaster any updates in cache that were not transmitted to the DR site are lost.

Network throughput

Another limitation on the operation of a DR site is the network bandwidth, which you can think of as the diameter of the pipe. The bigger the diameter, the more data that can be sent. But if the diameter is insufficient, then the data backs up until the flow is reduced. This reduced flow adds to the latency in the I/O or fills the cache faster if you use asynchronous replication.

Providing the bandwidth to meet the peak I/O workloads might involve an expensive network that sits idle for most of the time (assuming that peaks are rare). However, if the bandwidth is insufficient for peak I/O, then the application performance suffers.

Planning both bandwidth and latency

Planning for latency is relatively simple. After the sites are selected, the minimum latency that is incurred by the distance between the sites is largely set. But, the latency can be affected by the quality of the network hardware. The application performance and user acceptance are the final arbiters of what is tolerable. The I/O peak must not exceed the bandwidth of the network.

Planning bandwidth is more difficult because the bandwidth must be sufficient for both normal operations *and* recovery requirements. If there is a disaster, after recovery happens, the networks, depending on the topology, might have to support the extra activity as users catch up with lost processing and the system refreshes stale data at the recovery site.

Data divergence and recovery planning

Typically, data divergence and recovery planning is the result of loss of access to the active site when asynchronous or time-interval shipping of data is used. Then, the organization must decide whether they should move production to the alternative data center with the aged data or wait to resume at the primary site after fixing the failed components.

If operations were moved to the alternative site, one of the following decisions must be made when the failed site is recovered and if the *lost* data can be recovered:

- Move operations back to the recovered site and not recover the data that is cached there.
- Move operations back to the recovered site by using the data there and discard the data that was created while running on the alternative site.
- Attempt to recover the cached data while using the recent data from the alternative site.

To make this decision, the organization must understand the following items:

- The amount of data that can be lost and its potential value.
- Alternative (manual) methods to recover the data.
- Site recovery time.
- Whether the failure is localized or does it apply to the whole data center, and if localized, what is the cost in moving all operations to the alternative site?

A good test plan, which is performed regularly, helps with planning, and training staff in the procedures.

DR site

The following points must be considered to ensure that operations continue smoothly at the DR data center. Although many of these points are obvious and easy to remedy, you do not need the extra stress during DR recovery.

- Does the remote data center have access to the following up-to-date information?
  - Authorizations in place to access the remote data center?
  - Is documentation sufficient for any skill level to recover and resume operations as trained staff might be unavailable or unable to reach data center?
  - Backups and backup mechanisms?
  - Contracts and support agreements. Is the hardware support up to date and the data center staff authorized to contact support?
  - Licenses and application software are supported on data center serial numbers?
- Is all required infrastructure outside the cluster replicated in the remote data center?

1.5.2 VM Restart Manager with DR

In addition to general considerations about latency and throughput, there are a few extra items that must be considered when planning for VMRM DR:

- Until Version 1.7, there was no redundancy option for the KSYS manager. Version 1.7 introduced HADR for KSYS.
- Ensuring that the storage layer is supported. The VMRM DR solution supports the following storage devices:
  - EMC Storage Systems

    VMRM DR supports storage devices for the EMC VMAX family (VMAX1, VMAX2, and VMAX3). The EMC storage devices must be Symmetrix Remote Data Facility (SRDF)-capable and must have Solutions Enabler SRDF Family 8.1.0.0 or later installed. Both SRDF/S (Synchronous) and SRDF/A (Asynchronous) replication modes are supported. The SYMCLI interface on the KSYS node must be the same or later version with the SYMCLI interface on the storage agent.
  - EMC Unity Storage System

    VMRM DR supports EMC Unity Storage System 5.0.6.0.6.252 or later. Both synchronous and asynchronous modes of data replication are supported across sites.
  - IBM SAN Volume Controller (SVC) and IBM Storwize Storage Systems

    VMRM DR supports IBM SVC 6.1.0 or later and IBM Storwize V7000 7.1.0 or later. Both Metro Mirror (synchronous) and Global Mirror (asynchronous) modes of data replication are supported across sites.
  - IBM System Storage DS8000 series

    VMRM DR supports DS8700 or later and DS8000 Storage Systems with DS CLI 7.7.51.48 or later. Only the Global Mirror (asynchronous) mode of data replication is supported across sites.

- IBM XIV Storage System and IBM FlashSystem A9000

  VMRM DR supports IBM XIV Storage System and IBM FlashSystem A9000. Both Metro Mirror (synchronous) and Global Mirror (asynchronous) modes of data replication are supported across sites.

- Hitachi Storage Systems

  VMRM DR supports the Hitachi Virtual Storage Platform (VSP) G1000 and Hitachi VSP G400 with CCI Version 01-39-03/04 and model RAID-Manager or AIX. Both synchronous and asynchronous modes of data replication are supported across sites.

- The VIOS must have a Shared Ethernet Adapter (SEA) configuration to bridge to the same Ethernet network between the hosts at the same site.
- The same virtual local area network (VLAN) must be configured across the site. If a different VLAN is required at the target or backup site, the KSYS configuration must be updated for the different VLAN ID at the target or backup site.
- For HA, ensure that there is redundant connection from the KSYS LPAR to the Hardware Management Console (HMC) and from the HMC to the VIOS LPARs. Any connectivity issues between the KSYS LPAR, HMC, and VIOS LPARs can lead to disruption in the regular data collection activity and DR operations.

1.5.3 Clustering with DR

Another key consideration for a clustered solution is the requirement for a quorum site if you use automated failover.

An automated DR solution must avoid creating a split-brain scenario. A *split-brain scenario* occurs when the connection between nodes is interrupted and multiple nodes attempt to become the primary node. This scenario leads to the real possibility of lost or corrupted data. To avoid this problem, when the nodes in the two data centers lose contact, then the clustering software uses the quorum site (often called the third site or “laptop solution”) to determine which site should continue to operate. Without this check, each site assumes that the other site failed, with the currently running site continuing operations and the DR site starting each application.

The quorum site often has a disk device (Fibre Channel (FC) or iSCSI) or file in a network shared file system. The usage of locks on the “quorum” device controls which site is active.

1.6 Cloud and hybrid cloud disaster recovery

A *hybrid cloud application* is a mix of on-premises, private, or public cloud platforms with orchestration between these distributed platforms and workloads to perform as a single business service. The flexibility, agility, scalability, and interoperability of a hybrid cloud environment creates a platform to run business-critical applications. The hybrid cloud applications that are built on IBM Power servers are known for their high performance and reliability.

Although public cloud service providers ensure HA through data center redundancy, it is not always sufficient to protect from human or system errors or natural disasters hitting the services on a hybrid cloud application. Recent public cloud outages also point toward the need for a robust DR solution for your critical applications. Many protected environments, even IBM Power servers, can fail due to a single or multiple failures. To prepare for those scenarios, proactively define your DR strategy and design.

As with HA, there are two approaches to DR: The application-driven approach, and the approach that is driven by the underlying technology. In application-driven DR, the application handles the replication, and with the technology-driven approach, the infrastructure or OS handles the replication and orchestrates the recovery.

1.6.1 IBM Power in IBM Cloud

IBM Power Systems Virtual Servers (PowerVS) is a IBM Cloud solution that uses IBM Power servers within IBM Cloud. PowerVS is available in multiple geographically dispersed data centers or regions. Within an IBM Cloud region, IBM uses placement policies and SRR to ensure local availability. For the next level of availability, organizations may mirror data across two storage solutions or pathways.

Figure 1-3 shows the types of HA solution that are available with PowerVS, either based on VM restart or a clustered solution that uses placement groups and different storage controllers and pools.

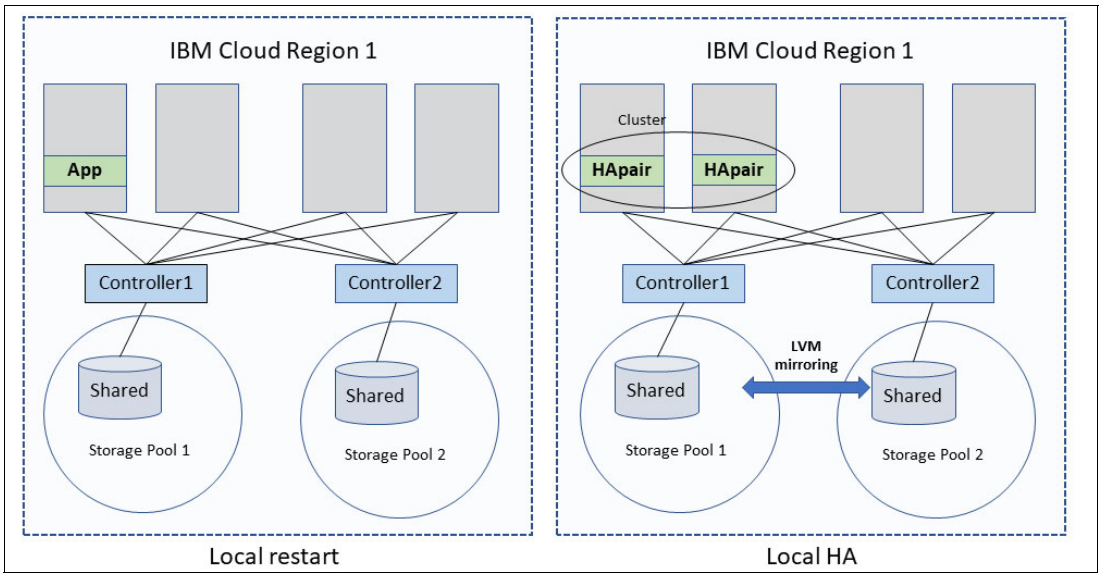

Figure 1-3 Options with PowerVS for HA

Until recently, the only option for DR was replication over IP, that is, using Geographic Logical Volume Manager (GLVM) for AIX and IBM i Geographic Mirror. Figure 1-4 shows the usage of mirroring data over IP between PowerVS data centers and IBM PowerHA SystemMirror to manage the application, IP addresses, and access to the storage with the associated replication direction changes.

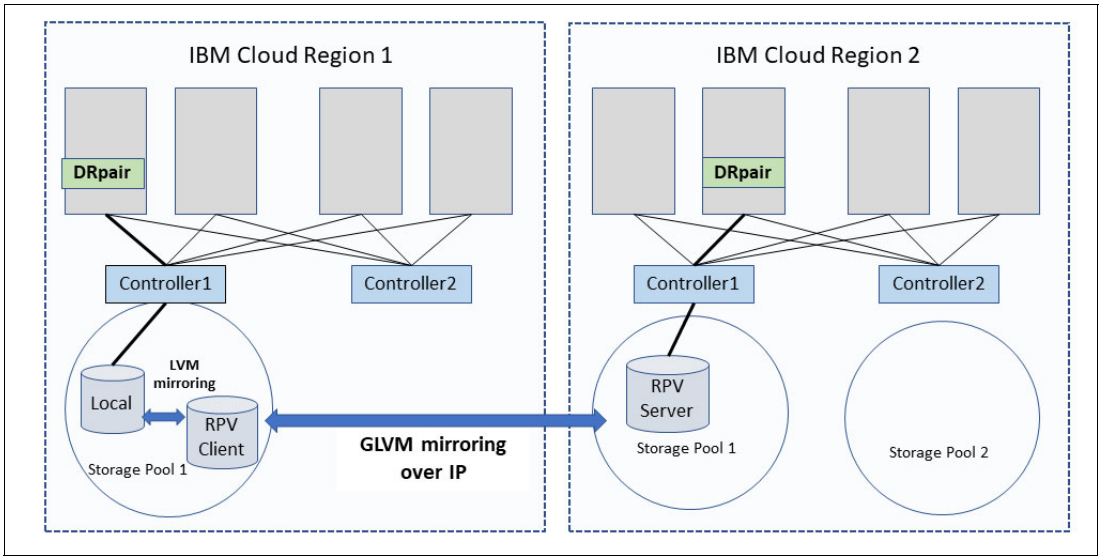

Figure 1-4 IP mirroring for DR

However, in September 2022, IBM announced a new service in the IBM Cloud that is called *Global Replication Service* (GRS), which operates between selected IBM Cloud Data Centers. GRS is an implementation of Storwize Global Mirror Change Volume Replication, which is coupled with several options that are optimized for automation and multi-tenancy by using Consistency Groups (CGs). Global Mirror is asynchronous replication that operates between two sites even when the distance between them is over 300 km. Figure 1-5 shows how GRS uses Storwize replication to provide DR.

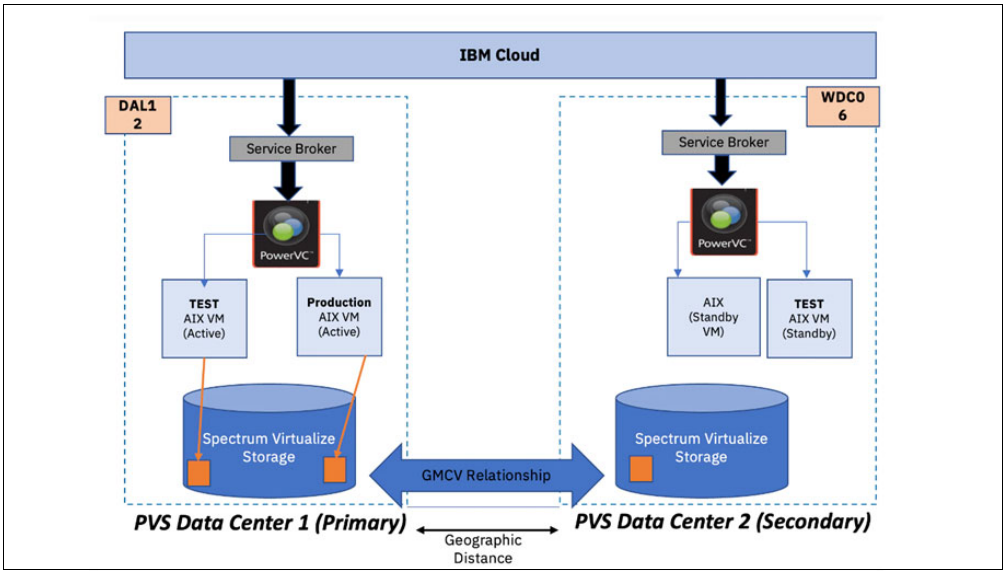

Figure 1-5 IBM PowerVS Global Replication Service

Here are some key features of GRS:

- Volume-based asynchronous storage replication by using CGs.
- CG services to manage the CGs, such as create or delete, add or remove volumes, and stop or start.
- Volume services, such as create, delete, and retype to support mirrored volumes.
- VM services, such as deploy, delete, attach, detach, clone, and snapshot for mirrored volumes.

The plan for GRS is that it is an enabler for a range of DR solutions to be provided by others:

- Client-based solutions based on their environment (that is, apps) and their own custom DR scripting.
- Client-based solutions that are based on their manual DR processes.
- IBM Technology Services solutions that are based on IBM i or AIX DR toolkits.[3]
- Independent software vendor (ISV)-based solutions, such as PowerHA for IBM i.

It is planned that more PowerVS data centers will be set up with the required connectivity for Storage Replication:

- Physical design (connectivity, switches, and so on)
- IBM Spectrum Virtualize Global Mirroring Change Volume capabilities
- Device and data center relationships
- Operational setup and dynamics

This approach is consistent with how many IBM Power clients are doing DR today on-premises. For more information, see the GRS documentation on IBM Cloud.

1.7 Assessing and designing for continuous operations

To assess and design continuous operations, follow these general guidelines:

- Conduct a business impact analysis where you identify and prioritize system components by correlating them to the mission and business process that the system supports, and use this information to characterize the impact on the process if the system was unavailable. Unavailability of maintenance and industrial engineering for an aircraft application can, for example, prevent aircraft from taking off, which can have a huge compliance, financial, operational, and reputational impact.
- Document your recovery time from an unplanned outage event. This task entails estimations that are based on which period and applications are running and the associated current staffing requirements.
- Establish estimates for the financial cost of planned or unplanned downtime. This task varies by when the outages occur. Get agreement and approval on the estimates from the CFO's office.
- Establish an approved time budget for planned and unplanned outage downtime.
- Establish an agreement with your customers on RTO and RPO.
- Be able to test your current HA capabilities by conducting regular failover operations. If you do not do this task, you do not have a HA solution; you have a theory.

[3] At the time of writing, an AIX took kit is in development.

- For HA, you want to build in failover or switching between systems for HA solutions that satisfy your availability goals. Even in an 8:00 AM - 6:00 PM system, achieving “four nines” almost surely entails eliminating all SPOFs and automating your failover.
- For DR, ensure that your systems can withstand a disaster, which typically entails constructing a backup system in a separate place from the primary to ensure that local phenomena such as weather or earthquakes do not destroy both systems. Because of the distances between the two systems, a DR failover differs from a HA failover.
- Prepare for the worst-case scenario. Your solution must be able to operate under extreme situations, run production from a backup site for an extended period, and operate in a degraded or damaged environment. Your IT depends on many elements, which can be harmed in an emergency: availability of key persons, network services that are provided by third parties, or dependencies on key suppliers and contractors.
- Conduct annual or biannual DR tests to validate that your approach is effective, repeatable, and relatively simple to run with minimal IT operations and application support involvement.
- Test your DR capabilities. To verify your DR capabilities, you can run many types of tests: dry tests, DR simulations, and switchovers. Like with HA, if you do not do this task, you do not have a DR solution; you have a theory.
- Ensure that your documentation is complete and current. After any failover (test or real), review and update the procedures.
- Ensure that trained staff is available and can access either site during a disaster.
- Your choice of solution must consider complexity, how dependent it is on human intervention, and what is its impact on production resources and performance.
- Continuous availability is more than hardware and software. The facility prerequisite, the process, and resource requirements must be considered.

1.7.1 Issues in managing an HADR environment

Here are the critical components of successful HADR environments:

- Planning
- Monitoring
- Maintaining
- Documenting
- Testing

Planning

An important component of planning an overall HADR plan is to regularly review the organizations applications against the required RTO and RPO and ensure that the HADR solutions deliver those requirements. Typically, an organization deploys many of the HADR solutions, and chooses the ones that meet the requirements of each class of application.

Chapter 2, “IBM HADR solutions for IBM Power Systems” on page 25provides an overview of the options that are available and what they provide.

The other components are covered elsewhere, but include the following items:

- Risk analysis and review of the types of possible disasters.
- Planning network throughput and latency while reducing the risks of both data centers being impacted by the same disaster.
- Ensuring that customers receive their expected quality of service when using the DR site.

- For automated systems, there is a third site to protect against a loss of inter-site communications.
- Testing the DR plan without interrupting operations or availability.
- Ensuring that your plan supports the extra load that is required in recovering back to the primary site after a DR.
- Planning staff availability and training.
- Planning resources for normal operations and during recovery from disaster.

The following list is built on the experience with issues that customers faced building their HADR solutions. As part of the planning process, review this list.

- Systems that are provisioned for DR are often of a different type, size, and capacity than production.
- Maintain consistency across the cluster in the following areas:
  - Application binary files
  - Configuration files
  - Application scripts
  - Users, groups, and passwords
- Application licenses may be tied to hardware serial numbers.
- Some local HA options, such as multiple instances of an application, no longer exist at the DR site if the services are combined on the same server.
- Production applications that are tied to a specific network address or network name during installation.
- The network at the DR site is on a different subnet or VLAN, and DNS updates might take time.
- Node name and hostname conflicts between existing systems in the DR site and the new systems being implemented under the DR plan.
- Multiple implementation standards for various functional system types, such as stand-alone, HA, and DR.

With planning, many of these issues can be avoided:

- Do not hardcode hostnames or IP addresses.
- Use unique names across the enterprise.
- Use scripts that check the host or site, rather than having different scripts for each host.
- Use local name resolution within the cluster to prevent DNS-related issues. Often, organizations have site-specific host files for cluster nodes to handle the subnet changes.
- Establish good change control to ensure that any administration changes are reflected across the cluster.
- Use PowerHA provided tools to keep users (such as C-SPOC or `clpasswd`) and files (file collections) consistent across the cluster.

Other considerations for planning DR vary for each application environment. The connectivity options and the distance between sites also dictate what type of data replication options are available. There is a careful balance that is required between the bandwidth that is required and the latency that is encountered when traversing greater distance. Although technologies might support “unlimited” distance, there is the cost, which is how much data is lost in a disaster.

Monitoring

Monitoring the entire environment is important to find and fix problems before they lead to an outage. For example, when a redundant component fails, the component is fixed or replaced to continue to provide the original level of redundancy. Undetected or unresolved problems can accumulate over time, which removes redundancy and can ultimately lead to an outage.

Maintaining

Although problems that are found by monitoring often lead to maintenance, it is not the only component of maintaining an environment. Normal maintenance often includes the following items:

- Backups
- Installing OS updates
- Installing application updates
- User access and password management
- Old data and files cleanup
- Current documentation
- Problem detection and fixes
- Security scans

Documenting

Documenting can be a time-consuming task, but it is important during normal operations and especially in an emergency. Documentation can be done in various ways, and as a best practice, keep documentation on the company's intranet when possible. Documentation must be constantly maintained, and there are often scripts and automated tasks that can help you keep system documentation current.

Another critical component of documentation is the post-outage review and update of lessons learned. After every incident, you can improve your organization's HADR by learning from the experience, improving your monitoring so that similar events are captured in future, and updating your documentation, training, and testing.

Testing

All plans and solutions are worthless if they are never tested. *All* change and management procedures must be tested in a non-production environment before they are implemented in production. All HADR solutions must be methodically tested regularly. It is better to find a problem during planned testing than during an unplanned outage.

1.7.2 Comparing the options

At a high level, the solutions that are presented in this publication apply equally to the following scenarios:

- HA within a data center
- HADR between two data centers

There are some differences and limitations for each solution, as shown in Table 1-4.

Table 1-4 Availability solution options for different data center configurations

Option	Within one data center	Between data centers
Live Partition Mobility (LPM)	Yes	Yes
SRR	Yes	Yes
VMRM HA	Only on-premises	N/A
VMRM DR	N/A	Only on-premises
GRS	N/A	Only PowerVS
PowerHA Standard	Yes	N/A
PowerHA Standard cross-site	N/A	Yes
PowerHA SystemMirror Enterprise Edition	N/A	Yes
PowerHA SystemMirror Enterprise Edition with IBM i Geographical Mirror		Yes

In Chapter 2, “IBM HADR solutions for IBM Power Systems” on page 25, we look at the IBM HADR products and how they help organizations meet their continuous availability requirements.

2

IBM HADR solutions for IBM Power Systems

This chapter explores the high availability and disaster recovery (HADR) options that are available in IBM Power and how they meet the range of availability requirements typically that are found in an organization.

This chapter describes the following topics:

- Introduction
- LPAR and virtual machine restart options
- Clustering options
- Comparing the clustering options (RTO, RPO, and cost and complexity)

2.1 Introduction

Typically, an organization has a wide range of environments:

- Environments and applications with near-zero recovery point objective (RPO) and recovery time objective (RTO) requirements.

  Typically, these applications are the ones where an outage of them incurs the greatest cost to the organization or they have the highest regulatory requirements.

- Environments or applications where the time that the organization can afford to have them unavailable is in the order of hours.

  These applications are important, but short interruptions of service are acceptable and do not create undue financial impacts.

- Environments or applications that can be unavailable for days.

  These functions must be performed, but they are usually not time-sensitive.

For availability considerations, applications can be typically grouped as follows:

- Single-instance applications with no built in availability.
- Single-instance applications with some availability. For example, databases with built-in replication or log shipping.
- Applications that can scale out or run multiple instances. This design is for both availability and to manage performance by increasing or decreasing the number of instances.

Because there is a range of requirements across the set of customers and within a typical organization, IBM offers a range of solutions, which are designed to meet different RTO or RPO requirements while recognizing that the relative cost of an outage also varies across applications. These solutions are listed here and then further described in the following sections and chapters.

Live Partition Mobility

Live Partition Mobility (LPM) is a tool that lets you move LPARs (virtual machines (VMs)) from one server to another while the VM remains active. This feature is especially useful for planned outages where you can move VMs between servers for maintenance or upgrades. LPM can still be used sometimes when a VM is not running due to an unplanned event. LPM is OS-independent and can be used for VMs running IBM AIX, IBM i, or Linux.

Simplified Remote Restart

Simplified Remote Restart (SRR) is an automation solution that is configured on and operates through the Hardware Management Console (HMC) or a PowerVC instance.

If an LPAR is virtualized to meet the LPM requirements, then a server failure initiates a restart of the specified LPARs on other servers in the environment. PowerVC can add greater flexibility in controlling the destination server for each LPAR. SRR is also OS-independent and can be used on IBM AIX, IBM i, or Linux.

VM Recovery Manager HA

The VM Recovery Manager (VMRM) HA solution implements recovery of the VMs by using VM restart technology. The VM restart technology relies on an out-of-band monitoring and management component that restarts the VMs on another server when the host infrastructure fails. The VM restart technology should be differentiated from a conventional clustering technology that deploys redundant hardware and software components for a near real-time failover operation when a component fails.

The VMRM HA solution is ideal to ensure high availability (HA) for many VMs, and if it meets your RPO requirements, it is easier to manage than cluster environments because it does not have the complexities that are introduced by clustering. VMRM HA is based on SRR technology, so it is OS-independent and can be used with IBM AIX, IBM i, or Linux.

VM Recovery Manager DR

The disaster recovery (DR) capability to recover applications and services to a second location in a major site outage is a key component in providing continuous business services. The IBM VM Recovery Manager DR for IBM Power solution is a DR solution that is easy to deploy and provides automated operations to recover the production site. The VMRM DR solution is based on IBM Geographically Dispersed Parallel Sysplex® (IBM GDPS®), which optimizes the usage of resources and manages the data replication that is required to start the systems in the DR location. This solution does not require you to deploy the backup VMs in the DR location. Thus, the VMRM DR solution reduces the software license and administrative costs. VMRM DR is based on SRR technology, so it is OS-independent and can be used with IBM AIX, IBM i, or Linux.

PowerHA SystemMirror Standard Edition for AIX

PowerHA SystemMirror for AIX is a clustered solution of AIX servers with shared storage. As the storage is shared, this solution requires that all the servers be within a single campus or within metro distances to avoid application impact due to latency. PowerHA monitors the environment, including the application themselves, and restarts the application on other nodes in the cluster to work around local failures (server, storage, or network).

PowerHA SystemMirror Enterprise Edition for AIX

The Enterprise Edition extends the Standard Edition by managing replication of the storage to a remote or DR site. By using storage replication, you can extend the distances between the servers to essentially unlimited distances.

PowerHA SystemMirror for IBM i

PowerHA SystemMirror for IBM i offers a complete end-to-end integrated clustering solution for HADR. It provides a data and application resiliency solution that is an integrated extension of IBM i capabilities for system and storage management. It is built to provide application availability through either planned or unplanned outages.

2.2 LPAR and virtual machine restart options

In this section, we describe methods of moving VMs between servers. Moving a VM between servers in a planned outage can be performed without an outage by using partition mobility. For unplanned outages where the hosting server is down, then partition mobility cannot be used. Instead, you must use a partition restart option to move the partition and restore its operation.

Partition mobility and LPAR restart are simpler to manage and configure compared to other HA options. Also, because they manage at the LPAR level, all IBM Power operating systems (OSs) (AIX, IBM i, and Linux) are supported.

2.2.1 Partition mobility (Live Partition Mobility)

Partition mobility is a component of the PowerVM Enterprise Edition hardware feature. It can migrate AIX, IBM i, and Linux LPARs from one system to another one. The mobility process transfers the system environment, which includes the processor state, memory, attached virtual devices, and connected users. For more information, see Partition Mobility in IBM Documentation.

Live Partition Mobility

By using active partition migration (LPM), you can migrate AIX, IBM i, and Linux LPARs that are running, including the OS and applications, from one system to another one. The LPAR and the applications that are running do not need to be shut down. The HMC is used to initiate the partition mobility move. For more information, see Partition Mobility in IBM Documentation.

Inactive partition mobility

By using inactive partition migration, also known as cold partition mobility, you can migrate a powered-off AIX, IBM i, or Linux LPAR from one system to another one. You use the HMC to migrate the inactive LPAR from one server to another one.

Because the HMC always migrates the last activated profile, an inactive LPAR that has never been activated *cannot* be migrated. For inactive partition mobility, you can either select the partition state that is defined in the hypervisor or the configuration data that is defined in the last activated profile on the source server.

Suspended partition mobility

For IBM AIX and IBM i partitions that meet the requirements, a partition can be suspended before doing a migration and then resumed when the migration is completed. For more information about suspending a partition, see Configuration requirements in IBM Documentation.

IBM i provides more support for partition mobility by providing exit programs that are run either before the partition is suspended or resumed to perform certain actions to make the partition eligible for suspended migration. For more information, see IBM i Live Partition Mobility.

Figure 2-1 on page 29 shows the benefits of using LPM.

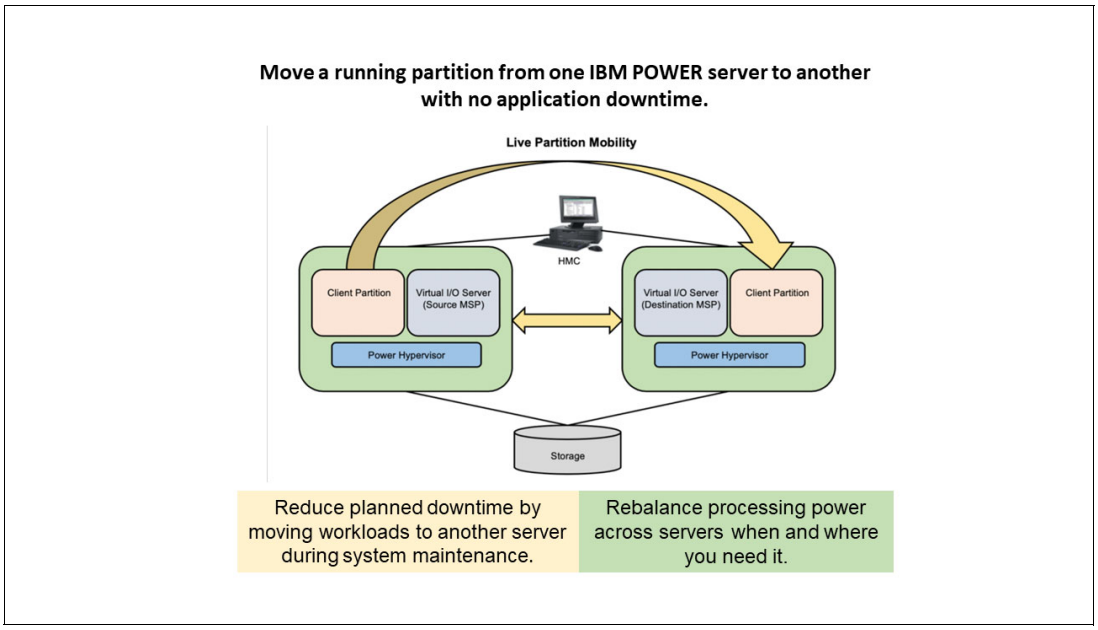

Figure 2-1 Benefits of LPM

With LPM, you can move partitions between servers to accomplish the following goals:

- Balance workloads across servers in your environment. Moving partitions to servers with less utilization can improve the performance of your application.
- Move workloads off a system to perform firmware maintenance or other planned system outages such as planned power outages in your site.
- Move workloads to new systems as you upgrade your server environment.
- Move workloads off a server where there are errors occurring to avoid an unplanned outage.

LPM itself is not an HA solution because it cannot react to the system outage by itself. However, it can be a tool to eliminate some planned outages and add to the overall availability of your systems.

Live Partition Mobility prerequisites

The minimum requirements for LPM are as follows:

- IBM Power servers with compatible firmware levels.
- An IBM HMC that is connected to both servers.
- When the source server and destination server are managed by different Hardware Management Servers, the Secure Shell (SSH) authentication keys between both consoles must be set up correctly.
- PowerVM Enterprise Edition on both source and destination servers.
- The LPAR is configured as I/O restricted, and all the I/O must be virtualized (VSCSI, NPIV, Virtual Ethernet, and VNIC).
- There must be at least one Virtual I/O Server (VIOS) LPAR on the source server and one VIOS LPAR on the destination server.

- The LPAR uses external storage only, and that storage system must be accessible to both source and destination systems.
- The destination server must have sufficient available resources to host the migrating LPAR.

How LPM works

Two servers participate in the LPM process and are managed by an HMC. The source server is the server that is running the LPAR that you want to migrate, and the destination server is the server that you want to run the LPAR on when it is migrated. The process is shown in Figure 2-2.

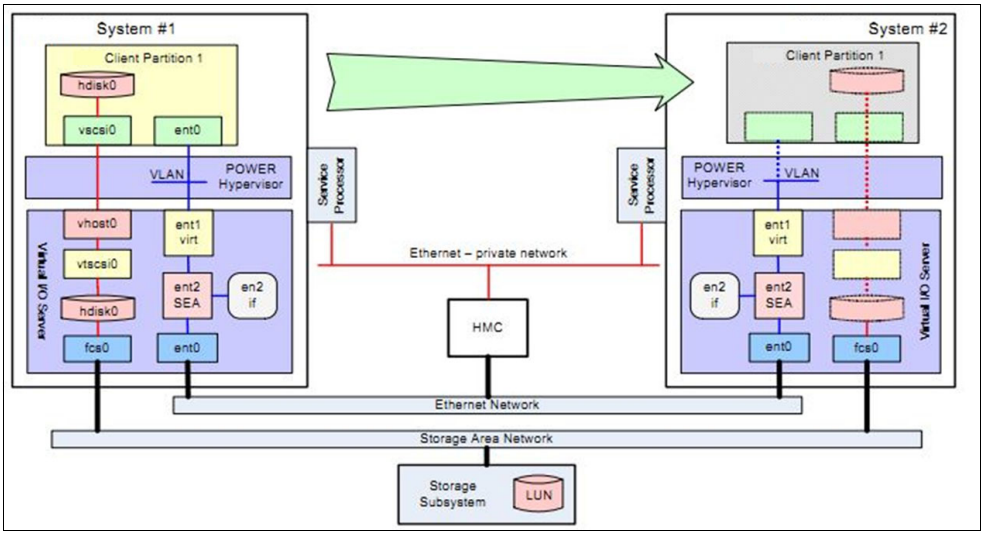

Figure 2-2 Illustrating the LPM environment

You can validate the existing configuration of both servers before starting the migration process. The migration is started by using either the **`migrlpar`** command on the HMC CLI or through the HMC GUI.

When the process of LPM starts, the HMC creates a migration profile for the LPAR, which is moved in the destination server that matches the current configuration of that LPAR. Then, LPM migrates all the profiles that are associated with the LPAR to the destination server. Then, the migration profile is used to map the original resources (VSCSI, Fibre Channel adapters, and network adapters) to the corresponding targets on the destination VIOS LPAR.

The time that is required for LPM to complete depends on several factors, including the memory size of the partition being moved and the network connectivity between the two servers.

The LPM main steps are as follows:

1. Configuration checking between both servers for required resources, such as main memory, processor resources, and storage visibility.
2. Creating a shell LPAR on the destination server with the necessary memory and processor pool, with the devices defined on the source profile, such as NPIV, VSCSI, and network devices. On the target VIOS, the NPIV and VSCSI devices also are created and mapped to the destination LPAR to enable access to the network and storage subsystem.
3. The migration of the main memory from the source LPAR to the destination LPAR starts while the users are still using the source LPAR.

4. The source LPAR is suspended when only a few memory pages are pending to migrate to the destination LPAR. Users cannot work with the source LPAR at this point, and can experience a delay in the response of the system in this point.
5. When the memory migration completes, the original LPAR definition and the I/O definitions of the VIOS are deleted on the source system. This action occurs while the target LPAR resumes activity so that the user transactions start without a need to log in to the system again.

Important:

- The LPM migration process fails if any LPAR with the same name exists on the destination server. However, this name can be found during migration validation.
- The best moment to perform an LPM is during a period of low activity on the system.

2.2.2 Remote restart and Simplified Remote Restart

Remote restart is a HA option for LPARs. When an error causes a server outage, a partition that is configured with the remote restart capability can be restarted on a different physical server. This feature is useful when it might take longer to restart the failed server and then restart the partition. So, the remote restart feature provides faster re-provisioning of the partition.

Remote restart is supported on IBM POWER7 and IBM POWER8® processor-based systems. Starting with both firmware and HMC 8.2.0 or later, SRR is supported. As a best practice, use the simplified version of the remote restart feature when your firmware and HMC support the SRR capability (includes POWER8, IBM POWER9™, and IBM Power10 processors). Remote restart requires that a reserved storage device is assigned to the partition on both the source and the target servers. SRR removes this requirement, which makes it a better choice for any servers that can support SRR.

If the source physical host has an error that causes it to halt, you can configure the LPARs to start on another (target) host. This feature might sound like inactive partition mobility, but the key difference is that the source physical host itself is no longer available or accessible.

SRR with HMC 8.2.0 or later running on POWER8 firmware 8.2.0 or later removes the need to assign reserved storage to each LPAR and is recommended.

The characteristics of SRR are as follows:

- SRR is *not* a suspend and resume or migration operation of the partition that preserves the active running state of the partition. During the remote restart operation, the halted or failed LPAR is started on a different system.
- SRR preserves the resource configuration of the partition. If processors, memory, or I/O are added or removed while the partition is running, the remote restart operation activates the partition with the most recent configuration.

When an LPAR is restarted by using SRR, a new profile is automatically created on the target host that matches the profile on the source host. Then, that new profile is mapped to the storage logical unit numbers (LUNs) that were being used by the original partition (that partition being inactive). Then, the new profile on the target host is activated and the partition is again active. When the source host becomes active, you must remove the old profile to ensure that the partition is not accidentally restarted on that host, especially automatically. The automatic cleanup runs without the force option, which means that if a failure occurs during the cleanup (for example, RMC communications with the VIOS fails), the LPAR is left on the original source host and its status is marked as `Source Side Cleanup Failed`.

The prerequisites for SRR are similar to LPM. In short, if LPM does not work for an LPAR, then SRR does not work either.

Other than the minimum required firmware, HMC versions, and VIOS versions, the high-level SRR prerequisites include the following items:

- Remote restart must be enabled on the VM. You can set this option while deploying or resizing the VM.
- Remote restart must be enabled on the host.
- The hosts and VMs must be SRR-capable.
- The source system must be in a state of *Initializing, Power Off, Powering Off, No connection, Error,* or *Error - Dump in progress*.
- The source systems VIOSs that provide the I/O for the LPAR must be *inactive*.
- The target system must be in an *active* state.
- The target systems VIOSs that provide the I/O for the LPAR must be *active*.
- The LPAR that will be restarted must be in an *inactive* state.
- The LMB size is the *same* on the source and the target system.
- The target system must have enough available resources (processors and memory) to host the partition.
- The target system VIOSs must provide the networks that are required for the LPAR.

Figure 2-3 Shows an environment that is configured for SSR, with two frames and shared storage. If one frame fails, the LPARs with remote restart enabled are started on the second frame.

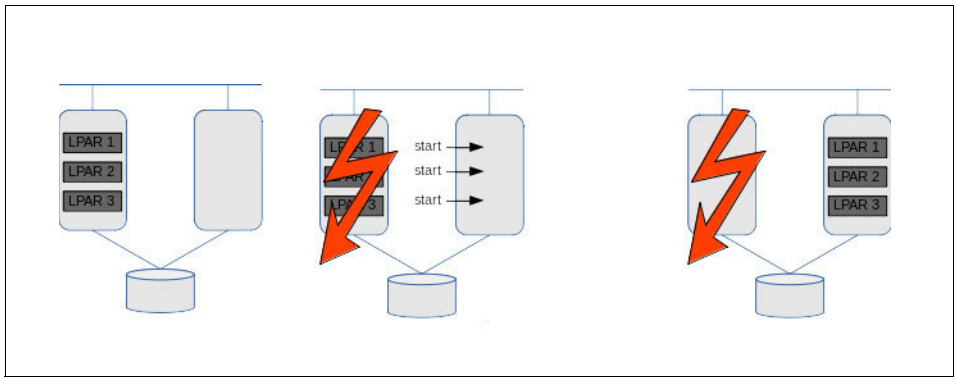

Figure 2-3 Simplified Remote Restart

2.2.3 Managing LPM and SRR

LPM and SRR are driven through the interaction of the HMC and the IBM Power servers (source and target). You can use the HMC GUI or the HMC CLI to manage the process. Figure 2-4 on page 33 shows an example of the HMC GUI window that is used to start a live migration.

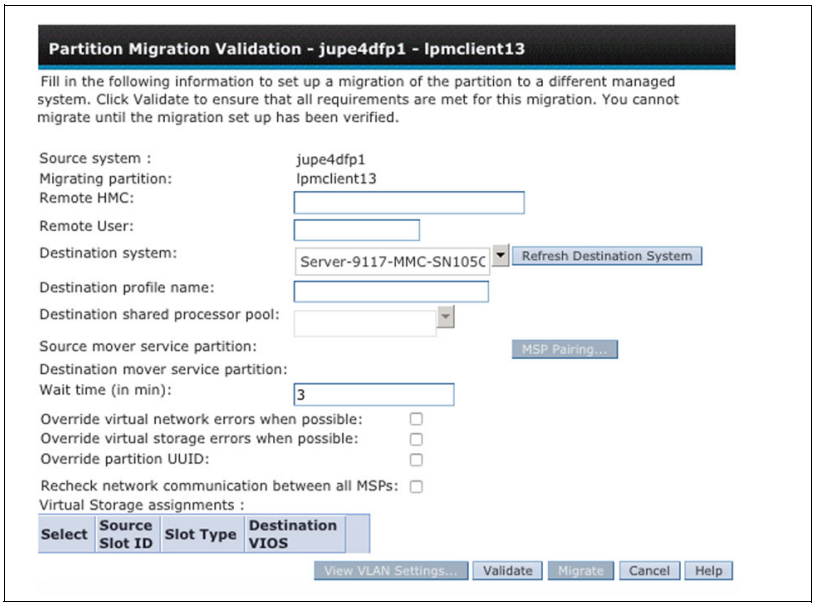

Figure 2-4 Hardware Management Console GUI for LPM

PowerVC

SRR is available through both the HMC and PowerVC. However, PowerVC adds another level of HA by automating the operation for SRR and adding features such as placement options.

Automated remote restart monitors hosts for failure by using the Platform Resource Scheduler (PRS) HA service. If a host fails, PowerVC automatically remote restarts the VMs from the failed host to another host within a host group.

Without automated remote restart enabled for an LPAR, if the host goes into an *Error* or *Down* state, then a manual remote restart can be triggered to restart the LPAR on another server. Manual remote restarts can be used at any time regardless of the LPARs remote restart setting.

For more information about automated remote restart with PowerVC, see Automated remote restart.

Demo: A demonstration of the automated remote restart capability is available on YouTube.

IBM PowerVM LPM and the SRR Automation tool

IBM Lab Services has a solution for managing LPM and SRR. With the PowerVM LPM and SRR Automation tool, you can leverage many of the features of LPM and SRR more quickly than by using the base management functions that are provided by the HMC GUI or HMC CLI through its GUI or spreadsheet support.

Also, built in to the tool are back-end capabilities that have no equivalent on the HMC:

- Return all the partitions back after LPM or SRR operations.
- Daily health checks of LPM and SRR readiness.
- Scripting capabilities.
- Automatic movement of Mobile Capacity (also known as Power Enterprise Pools (PEP)).

Demonstration: A demonstration of the PowerVM LPM and SRR Automation tool is available on YouTube.

2.2.4 IBM VM Recovery Manager High Availability

IBM VMRM HA for IBM Power is an HA solution that is easy to deploy and provides an automated solution to recover the LPARs on a failed host. It supports all three of the OS types that are supported on IBM Power: AIX, IBM i, and Linux.

The VMRM HA solution implements recovery of the VMs based on VM restart technology. The VM restart technology relies on an out-of-band monitoring and management component that is used to restart the VMs on another host when the current host fails. The VM restart technology is different from the conventional cluster-based technology that deploys redundant hardware and software components for a near real-time failover operation when a component fails.

The VMRM HA solution is ideal to ensure HA for many VMs. Also, the VMRM HA solution is easier to manage than cluster-based tools because there is no requirement to manage consistency and communication across a cluster.

The VMRM HA solution provides the following capabilities:

Host health monitoring The VMRM HA solution monitors hosts for any failures. If a host fails, the VMs in the failed host are automatically restarted on other hosts. The VMRM HA solution uses the host monitor module of the VIOS partition in a host to monitor the health of hosts.

VM and app monitoring The VMRM HA solution monitors the VMs, registered applications, and hosts for any failures. If a VM or a critical application fails, the corresponding VMs are started automatically on other hosts. The VMRM HA solution uses the VM monitor agent that must be installed in each VM to monitor the health of VMs and registered applications.

Unplanned HA events During an unplanned outage, when the VMRM HA solution detects a failure in the environment, the VMs are restarted automatically on other hosts. You also can change the auto-restart policy to advisory mode. In advisory mode, failed VMs are not relocated automatically; instead, email or text messages are sent to the administrator. The administrator can use the interfaces to manually restart the VMs.

Planned HA events During a planned outage, for example, a firmware update, VMRM HA uses LPM to vacate the host and move all the VMs to other hosts in the group. After the maintenance operation, VMRM HA can be used to restore the VMs to their original host in a single operation.

Advanced HA policies The VMRM HA solution provides advanced policies to define relationships between VMs, such as colocation and anti-collocation of VMs, the priority in which the VMs are restarted, and the capacity of VMs during failover operations.

GUI and CLI management Either the GUI or the CLI can be used to manage the resources in the VMRM HA solution after the managing server is installed. The GUI is accessed by a web browser and from the CLI, where the **`ksysmgr`** command is used.

Figure 2-5 shows an environment that is managed by VMRM HA, with three hosts and shared storage. In this example, host 1 fails and the three VMs running there are restarted on the remaining two hosts.

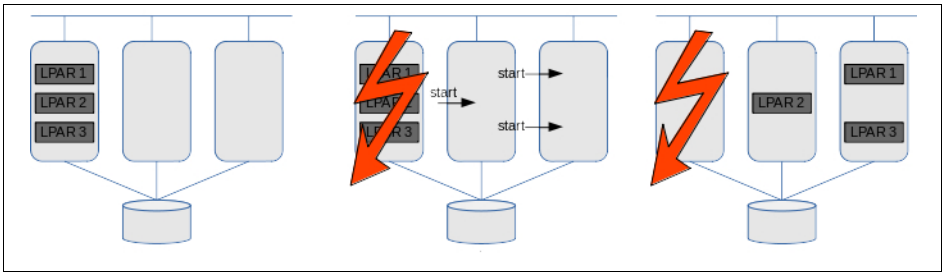

Figure 2-5 VMRM HA in operation

2.2.5 IBM VM Recovery Manager Disaster Recovery

IBM VMRM DR for IBM Power, formerly known as IBM Geographically Dispersed Resiliency (IBM GDR), consists of both HADR offerings in the same package. This solution is a DR solution that is easy to deploy and provides automated operations to recover the production site. The VMRM DR solution is based on the IBM Geographically Dispersed Parallel Sysplex (IBM GDPS®) offering that optimizes the usage of resources. This solution does not require you to deploy the backup VMs for DR. Thus, the VMRM DR solution reduces the software license and administrative costs.

Clustered HADR solutions typically deploy redundant hardware and software components to provide near real-time failover when one or more components fail. The VM restart-based HADR solution relies on an out-of-band monitoring and management component that restarts the VMs on other hardware when the host infrastructure fails. The VMRM DR solution is based on the VM restart technology.

The VMRM DR solution automates the operations to recover your production site. This solution provides an easy deployment model that uses a controller system (KSYS) to monitor the entire VM environment. This solution also provides flexible failover policies and storage replication management.

Table 2-1 identifies the differences between the conventional cluster-based DR model and the VMRM DR solution.

Table 2-1 Clustered DR versus VM Recovery Manager DR

Parameters	Cluster-based DR model	VM restart DR model that is used by the VM Recovery Manager DR solution
Deployment method	Redundant hardware and software components are deployed at the beginning of the implementation to provide near real-time failovers when some of the components fail.	With virtualization technology, many images of the OS are deployed in a system. These VMs are deployed on physical hardware by the hypervisor that allocates and manages the CPU, memory, and I/O physical resources that are shared among the VMs.
Dependency	This solution relies on the monitoring and heartbeat capabilities within the cluster to monitor the health of the cluster and take recovery action if a failure condition is detected.	This solution relies on out-of-band monitoring software that works closely with the hypervisor to monitor the VM environment and to provide a DR mechanism for the VM environment.
Workload startup time	The workload startup time is faster because the VMs and the software stack are already available.	The VMs require more time to restart in the backup environment.
Cluster administration required	Yes.	No.
Error coverage	Comprehensive. This solution monitors the entire cluster for any errors.	Limited. This solution monitors the servers and the VMs for errors.
Deployment simplicity	This solution must be set up in each VM.	Aggregated deployment at the site level.
Protected workload type	Critical workloads can be protected by using this solution.	Critical workloads can be protected by using this solution.
Software license and administrative costs	This solution costs more because redundant software and hardware are required to deploy this solution.	This solution costs less because of optimized usage of resources.

Demonstration: A demonstration of VMRM DR under its original name of IBM GDR is available at YouTube.

Figure 2-6 on page 37 shows an example of three VMs running on a host at one site, with the storage being replicated to the remote site. If the KSYS manager recognizes that the host at the first site has failed, the storage replication can be reversed and the VMs started in the second site.

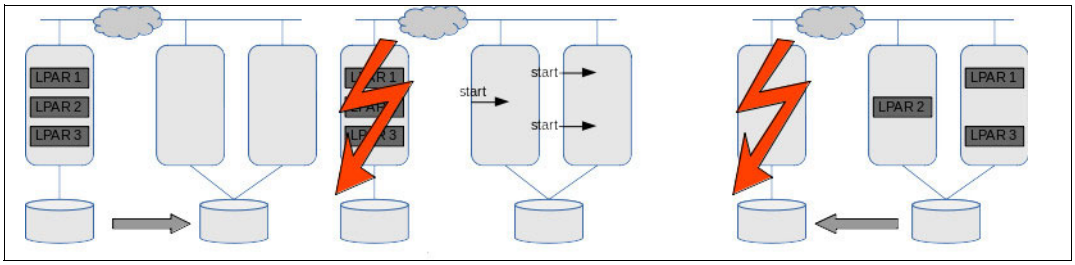

Figure 2-6 Operation of VMRM DR

VMRM DR has the added feature of supporting DR testing. A copy of the storage at the remote data center is copied and zoned to the DR test VMs, which then can be started in a sandbox environment for testing.

2.2.6 Summary of LPAR availability management options

Table 2-2 compares the features of the different LPAR management options.

Table 2-2 Comparing features of the LPAR management solutions in the IBM portfolio

Feature	Live Partition Mobility	Simplified Remote Restart	VM Restart HA	VM Restart DR
Support	≥ IBM POWER6	≥ IBM POWER7	≥ IBM POWER7+	≥ POWER7
Frame failure	N	Y	Y	Y
VM Monitor	N	N	Agent (AIX)	Agent (AIX)
Auto failover	N	N	Y	Y
Storage	Shared	Shared	Shared	Replicated
Clustering	N	N	N	N
Active-passive	Y	Y	Y	Y
DR	N	N	N	Y
Automated Failover	N	N	Y	N
Source Server Status	Active	Inactive	Active or Inactive	Active or Inactive
Source VIOS Status	Active	Inactive	Active or Inactive	Active or Inactive
VM/Application Outage	No (if LPAR active)	Y	Only if there is a frame or LPAR outage	Yes
RTO	N/A	Operator + IPL + App start	IPL + App start	VMRM HA time if local, DR+
RPO	N/A	0	0	sync 0; async cache
Tier	N/A	5[a<Default ¬¹ Font>]	6[a<Default ¬¹ Font>]	6(async); 7(sync)

Feature	Live Partition Mobility	Simplified Remote Restart	VM Restart HA	VM Restart DR
License usage	N/A	N + 0	N + 0	N + 0
Cost	N/A[a]	$	$$	$$

a. Within one data center.

2.3 Clustering options

The following solutions consist of a cluster of IBM Power servers that monitor themselves, the networks, the storage, and the applications that are moving the applications from LPAR to LPAR to work around failures or maintenance in the environment.

2.3.1 IBM PowerHA SystemMirror for AIX

IBM PowerHA SystemMirror (PowerHA) has versions for AIX and IBM i. There was a Linux version, but it was withdrawn. PowerHA for AIX, formerly known as IBM High Availability Cluster Multi-Processing (HACMP), has been popular in its 30+ year history. Originally designed as a stand-alone product (known as HACMP Classic) after the IBM HA infrastructure known as Reliable Scalable Cluster Technology (RSCT) became available, HACMP adopted this technology and became HACMP Enhanced Scalability (HACMP/ES) because it provides performance and functional advantages over the Classic version. Starting with HACMP 5.1, there are no more Classic versions. Later, HACMP terminology was replaced with PowerHA in Version 5.5 and then PowerHA SystemMirror 6.1.

PowerHA 7.1 was the first version to use the Cluster Aware AIX (CAA) component of AIX. This major change improves the reliability of PowerHA because the cluster service functions now run in kernel space rather than user space. CAA was introduced in AIX 6.1 TL6 and AIX 7.1 TL0. At the time of writing, the current release of PowerHA is Version 7.2.7.

A PowerHA cluster must contain a minimum of two LPARs that communicate with each other by using heartbeats and keepalive packets over several networks. The cluster contains many resources, such as IP addresses, shared storage, and application scripts, which are grouped as resource groups (RGs). An RG contains everything that the application needs to run, and it is the object that PowerHA moves around the cluster to ensure availability.

If PowerHA detects an event within the cluster, it automatically responds to ensure that the RG is placed on the most appropriate node in the cluster to ensure continued availability. A correctly configured PowerHA cluster after setup requires no manual intervention to ensure application availability in the case of a single failure, such as failures of physical servers, nodes, applications, adapters, cables, ports, network switches, and storage area network (SAN) switches. The cluster can also be controlled manually if the RGs must be balanced across the clusters or moved for planned outages.

PowerHA for AIX comes in two editions: *Standard* and *Enterprise*.

Standard edition is more synonymous with local HA, and in some configurations even near-distance DR. It depends on both shared local area network (LAN) and SAN connectivity between servers and storage.

A basic local cluster is shown in Figure 2-7 on page 39.

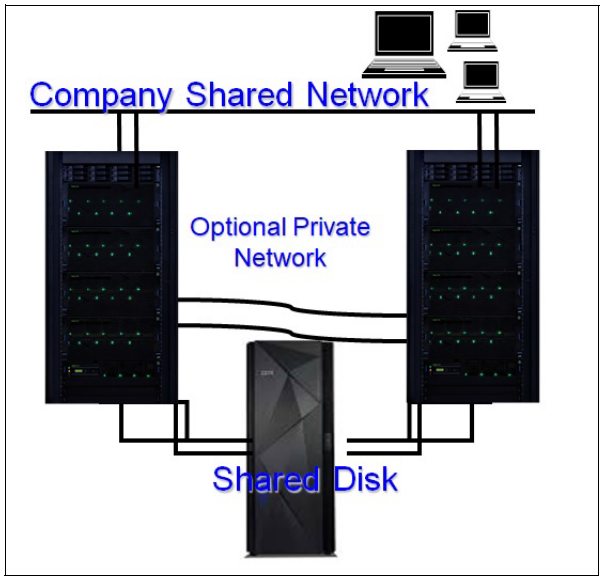

Figure 2-7 PowerHA SystemMirror Standard Edition cluster

Figure 2-8 shows an example of a 2-node cluster with shared storage and two applications. In this example, Application 2 experiences an error operating on the second host, which is detected by PowerHA, which responds by activating the resources in Application 2's RG on the first host and running the application start script.

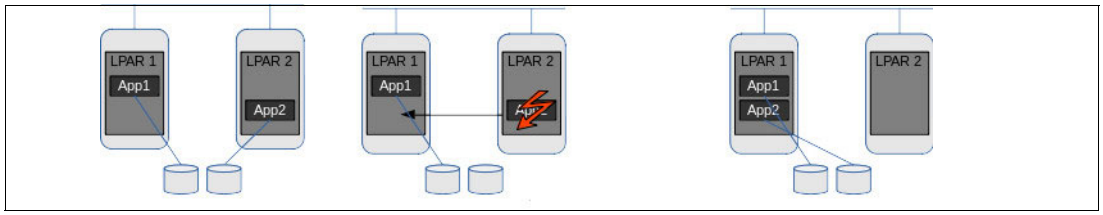

Figure 2-8 Operation of PowerHA SystemMirror

PowerHA clusters can be configured in many ways:

- Active-passive: One node in the cluster runs the RG, and its partners are in standby mode waiting to take on the resources when required. The passive nodes in the cluster must be running for them to participate in the cluster.
- Active-active: All nodes in the cluster are running an RG, but also are the standby nodes for another RG in the cluster. Many resources groups can be configured within a cluster, so how they are spread out across the nodes and in which order they move is highly configurable.
- Concurrent: All nodes in the cluster run the same RG. Historically, this configuration was most common with Oracle Real Application Clusters (RAC) environments, but some application servers also can be used in this configuration.

Clustered multiprocessing

In addition to HA, PowerHA also provides the *multiprocessing* component. The multiprocessing capability comes from the fact that in a cluster there are multiple hardware and software resources that are managed by PowerHA to provide complex application functions and better resource utilization.

A short definition for *cluster multiprocessing* might be multiple applications running over several nodes with shared or concurrent access to the data.

The cluster multiprocessing component depends on the application capabilities and system implementation to efficiently use all resources that are available in a multi-node (cluster) environment. This solution must be implemented by starting with the cluster planning and design phase.

PowerHA is only one of the HA technologies, and it builds on increasingly reliable OSs, hot-swappable hardware, and increasingly resilient applications by offering monitoring and automated response.

A HA solution that is based on PowerHA provides automated failure detection, diagnosis, application recovery, and node reintegration. PowerHA also can provide excellent horizontal and vertical scalability by combining other advanced functions, such as dynamic logical partitioning (DLPAR) and Capacity on Demand (CoD). In more recent versions of PowerHA this approach is known as Resource Optimized High Availability (ROHA).

Summary

Although most clusters are simple two-node active-passive clusters, PowerHA SystemMirror for AIX supports 32 nodes in a cluster for various failover options. Some of these options are shown in Figure 2-9.

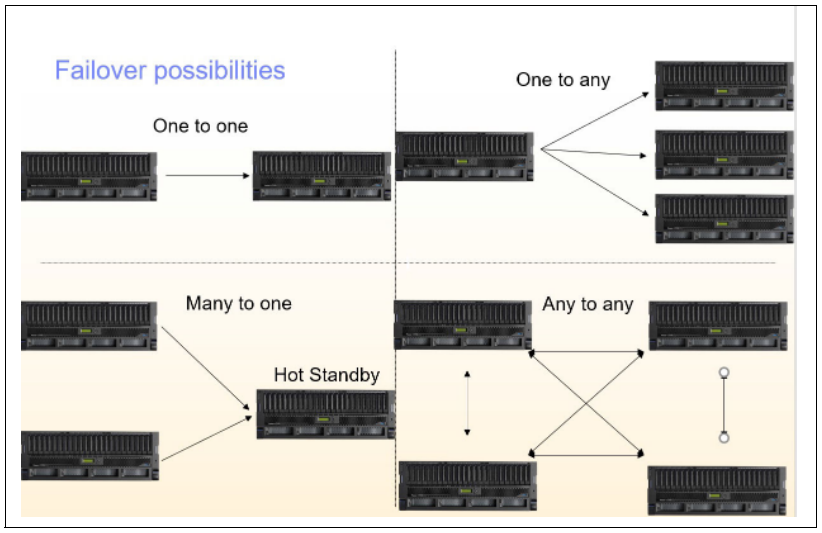

Figure 2-9 PowerHA failover options

PowerHA has many options and features, and many of them are tightly integrated into both AIX and PowerVM specific features. Some of these Standard Edition features are shown as follows and also linked to online demos where available:

- Dynamic Node Priority (DNP)

  The target and fallover node are chosen by available resources, such as the following ones:

  - Available CPU
  - Paging space
  - Disk I/O

- DLPAR with extra CPU or memory during startup or fallover

  Includes resource-optimized failovers by using enterprise pools (ROHA).
- LPM awareness
- Live Kernel Update (LKU) awareness
- RG dependencies, which are designed for multitier environments:
  - Parent and child
  - Same node or same site
  - Different node or site
- RG priorities:
  - Low
  - Intermediate
  - High
- AIX Logical Volume Manager (LVM) or Enhanced Journaled File System (JFS2) specialized option utilization:
  - File system Concurrent Mount Protection, also known as Mount Guard
  - Active-passive mode of concurrent volume groups
- Non-disruptive cluster updates and upgrades by using **`cl_ezupdate`**
- NovaLink managed LPAR support
- Rootvg and critical volume group loss detection
- User-defined events
- Customizable processing order
- Automatic Repository Replacement (ARR)
- Cluster testing, both automated and customizable
- Delayed Fallback Timer

2.3.2 IBM PowerHA SystemMirror Enterprise Edition for AIX

PowerHA Enterprise Edition includes everything Standard Edition does, but also provides cross-site clustering where shared storage is not an option but SAN-based replication is available. In this environment, PowerHA uses the remote copy facilities, either IP address- or storage-based to ensure that the nodes at each site have access to the same data, but on different storage devices. It is possible to combine both local and remote nodes within a PowerHA cluster to provide local HA and cross-site DR.

Figure 2-10 shows a simple example with two sites and a host at each one. PowerHA monitors the health of the environment and manages the storage replication for each application. If a site, host, or application fail, PowerHA reverses the storage replication and starts the application running on the surviving site.

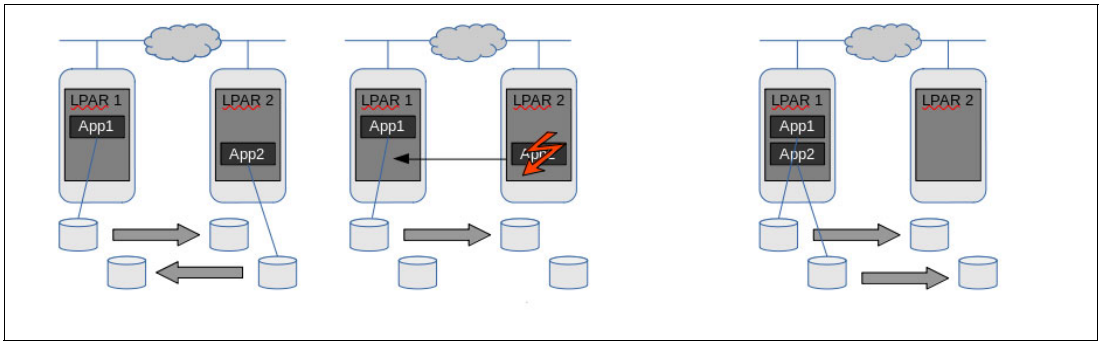

Figure 2-10 Operation of PowerHA Enterprise Edition

Enterprise Edition provides extra integrated support that primarily focuses on DR. Here are some of these features:

- IP-based replication and Geographic Logical Volume Manager (GLVM)
- IBM Spectrum Virtualize Storage Replication:
  - Metro Mirror
  - Global Mirror
  - IBM HyperSwap®
- Dell EMC: Symmetrix Remote Data Facility (SRDF) (Synchronous or Asynchronous)
- Hitachi:
  - TrueCopy for synchronous
  - Hitachi Universal Replicator for asynchronous
- IBM XIV: Remote Mirror
- User confirmation on split-site failure
- Site-specific service addresses

For more information about planning, installing, and configuring PowerHA SystemMirror for AIX, see the following resources:

- *IBM PowerHA SystemMirror for AIX Cookbook*, SG24-7739
- *IBM System Storage Solutions Handbook*, SG24-5250
- PowerHA SystemMirror Version 7.2 for AIX PDFs

Demonstration: A series of demos of PowerHA SystemMirror for AIX Enterprise Edition for AIX can be found as follows:

- Utilizing GLVM IP Replication
- Utilizing IBM SAN Volume Controller (SVC) Replicaton
- Utilizing XIV Replication
- Utilizing Hitachi TrueCopy

2.3.3 IBM PowerHA SystemMirror for IBM i

PowerHA SystemMirror for IBM i has been around since 2008 and shares many similarities with the AIX version. It is deeply integrated into IBM i and hardware-dependent System Licensed Internal Code (SLIC). Until V7R4, it offered three editions: Express, Standard, and Enterprise:

- Express Edition enables single-node, full-system HyperSwap with the IBM DS8700 server, which provides continuously available storage through either planned or unplanned storage outage events.
- Standard Edition is generally for local data center HA, and for using GUI commands, synchronous geographic mirroring, switched logical units (LUs), or FlashCopy.
- Enterprise Edition for multi-site DR solutions, which uses asynchronous geographic mirroring, Metro Mirroring, Global Mirroring, or DS8000 HyperSwap with independent auxiliary storage pools (IASPs).

When V7R5 was announced, a new version of PowerHA SystemMirror for IBM i was presented, only with one simplified edition. Now, PowerHA SystemMirror for IBM i is shipped only with *BASE and Option 1, which is called *PowerHA enablement*).

PowerHA for IBM i cluster configurations are flexible. It is more common for IBM i customers to deploy multi-site PowerHA clusters where the data is replicated either by IBM storage or by IBM i Geographic Mirroring. PowerHA integrates the IBM i OS with storage replication technologies that provide solutions that meet the HA needs of clients regardless of size.

Configurations range from a simple 2-system, 2-site cluster that uses IBM i Geographic Mirroring with internal storage to an IBM FlashSystem cluster or a 3-site HyperSwap cluster with DS8000 storage. Using IBM storage adds the extra benefit of FlashCopy, which is used to eliminate the backup window, conduct query operations, and create point-in-time copies for data protection purposes.

The production data, including the local journals, is contained within an IASP, and planned switchovers between nodes in the cluster consists of a single command. Unplanned failovers can be configured to be automatic and require minimal operator intervention. The administration domain takes care of synchronizing security and configuration objects, such as user profiles. All these tasks are done with integration between PowerHA and the IBM i, and there is no dependency on third-party replication tools. Because there is at least one active OS on each node in the cluster, you can conduct software maintenance and OS upgrades on an alternative node without disrupting production.

Implementing IASPs is a simple task consisting of moving your application libraries and Integrated File System (IFS) data into the IASP, thus separating business data from the OS. The application binary files do not change, and most users are unaware of the migration in their daily workflow because their jobs automatically have access to libraries both in the system ASP and the IASP simultaneously.

Demonstration: A demonstration of PowerHA for IBM i that uses IBM i Geographic Mirroring can be found at this YouTube link.

2.4 Comparing the clustering options (RTO, RPO, and cost and complexity)

Table 2-3 compares the features of the different clustering options that were described in this chapter.

Table 2-3 Comparing features of the HADR clustering options

Feature	PowerHA	PowerHA SystemMirror Enterprise Edition
Support	≥ POWER6	≥ POWER6
Frame failure	Y	Y
VM Monitor	Y	Y
Auto failover	Y	Y
Storage	Shared	Replicated
Clustering	Y	Y
DR	N (except cross-site LVM)	Y
Automated Failover	Y	Y
VM/Application Outage	Yes	Yes
RTO	App start	App start
RPO	0	sync 0; async +
Tier	7[a]	7
Node license usage	N + 1	N + 1
Cost	$$	$$$

a. Within one data center

3

IBM VM Recovery Manager High Availability and Disaster Recovery

This chapter focuses on IBM VM Recovery Manager HA (VMRM HA) and IBM VM Recovery Manager DR (VMRM DR) and provides guidance about what is required to install and run VM Recovery Manager (VMRM) in your environment.

This chapter describes the following topics:

- Terminology and concepts
- History and evolution
- New features in V1.7
- Installation planning and prerequisites
- IBM VM Recovery Manager installation
- Migrating and upgrading from previous releases

3.1 Terminology and concepts

The VMRM solution uses the following concepts.

IBM VM Recovery Manager HA and IBM VM Recovery Manager DR

VMRM HA is a single-site option that provides the ability to move LPARs to other servers in that location (no DR). VMRM DR is the DR version where LPARs that fail are moved to the backup site.

There are two more high availability (HA) options that are available for VMRM DR:

- HADR adds the ability to handle some failures by moving LPARs to another server in the primary site instead of failing everything over to the backup site.
- HADRHA adds the HA function to the backup site and the primary site.

Controller system

The controlling system (KSYS) is the central component that provides a single point of control for the entire environment that is managed by the VMRM. The KSYS interacts with the Hardware Management Console (HMC) to gather the configuration and health information of managed systems. The KSYS is responsible for recovery actions if a disaster or a potential disaster occurs, and should be protected for failure by using PowerHA SystemMirror for AIX.

Hardware Management Console

The HMC (virtual or physical appliance) is used to manage IBM Power servers, Virtual I/O Server (VIOS), and logical partitions (LPARs) that can be managed by the controller system.

Sites

A site is a logical name that represents a site that includes mapped HMCs, hosts, VIOS, and storage devices.

Hosts

A host is a managed, physical system in the HMC that runs the virtual machines (VMs) that provide the workload.

Host pair

A host pair is a set of hosts that are logically paired across the sites for HADR.

Host group

A host group is a group of hosts that are logically chosen and named by the administrator. When failures in any of the hosts are detected, VMs in the failed host are relocated and restarted on other healthy hosts within the group of hosts.

Virtual machines

A VM is an LPAR that is associated with a specific VIOS partition to provide virtual resources to run a workload.

Host monitor

The host monitor is a daemon that is deployed during the VIOS installation. The controller system subsystem communicates with the host monitor daemon through the HMC to monitor the hosts for HA.

VM agents

The VM agent subsystem provides HA features at the VM and application level and monitors for failures. The VM agent is supported on AIX and Linux (Red Hat Enterprise Linux (RHEL) and SUSE Linux Enterprise Server) guest VMs only. At the time of writing, the VM agent subsystem is not supported for the IBM i and Ubuntu VMs. Therefore, IBM i and Ubuntu VMs are relocated from one host to another host within the host group only after a host failure.

These optionally installed agents monitor the following issues in the production environment:

VM failures If the operating system (OS) of a VM is not working correctly, or if the VM stopped working because of an error, the VM is restarted on another host within the host group. The KSYS subsystem uses the VM monitor module to monitor the heartbeat from the VM to the host monitor subsystem in a VIOS.

Application failures Optionally, you can register the applications in the VM agent to enable application monitoring. The VM agent uses the Application HA monitoring framework to monitor the health of the application periodically by running the application-specific monitor scripts; by identifying whether the application failed; and by identifying whether the VM must be restarted in the same host or another host. This framework can also manage the sequence in which applications are started and stopped within a VM.

Storage agents

The storage agent is software that is used to perform replication operations from the active site to the backup site across or between specific support storage units. The supported unit types are described in 3.4.2, “VM Recovery Manager DR” on page 52.

3.2 History and evolution

Announced in October 2016, IBM Geographically Dispersed Resiliency (IBM GDR) for IBM Power is an automated disaster recovery (DR) solution that enables IBM Power users to achieve low recovery times for both planned and unplanned outages. IBM GDR provides systems and data replication monitoring along with server and network management, which simplify the tasks that are associated with recovery and maintaining HA.

The first version of IBM GDR was delivered as an IBM service offering for either the client's own DR site or an IBM hosted site through this DR as a service offering.

IBM GDR 1.2 was renamed to VMRM DR 1.3 to implement a common naming strategy for VM restart technology for both high availability and disaster recovery (HADR). Therefore, VMRM HA is the shared storage data center solution, and VMRM DR is based on storage replication of VMs from production to a auxiliary storage system that is at a secondary location.

Starting in April 2020, VMRM HA is included with AIX Enterprise Edition 1.3. VMRM DR is included in the Cloud Solution bundle for no additional charge.

Figure 3-1 shows the history of the VMRM product.

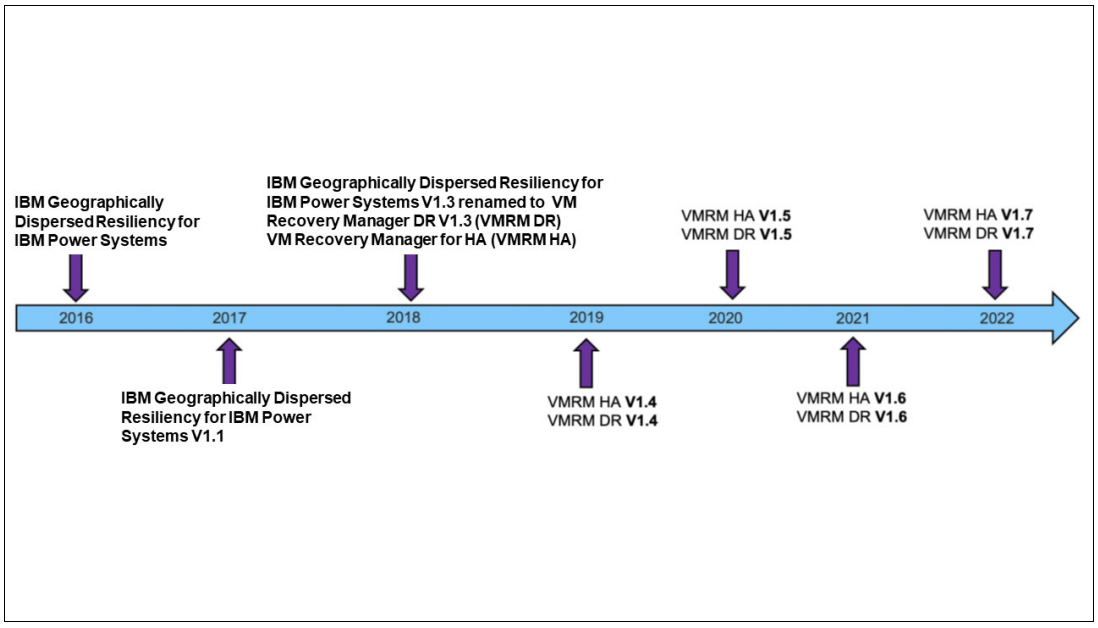

Figure 3-1 VM Recovery Manager roadmap

Table 3-1 provides highlights of the VMRM features for the recent releases.

Table 3-1 IBM GDR and VMRM HADR main enhancements

VMRM or IBM GDR version	Main enhancements
IBM GDR 1.1	► Supports IBM i as a guest OS, which adds to the current support for IBM AIX and Linux. ► Supports IBM DS8000 Global Mirror. ► Supports IBM SAN Volume Controller (SVC) and IBM Storwize Metro and Global Mirror. ► Supports EMC Symmetrix Remote Data Facility (SRDF) synchronous replication. ► Supports boot device selection for IBM POWER8 processor-based systems.
IBM GDR 1.2 / VMRM 1.3	► Supports IBM FlashSystem A9000 (XIV) storage mirroring-based DR management (1.3.0.2). ► HA agent for a Postgres database. ► Graphical management enhancements for VMRM DR. ► VMRM for DR includes VMDR HA, which provides the ability to deploy either in DR mode or HA mode. ► HA of the KSYS subsystem through PowerHA SystemMirror can be used to handle a scenario in which the KSYS node itself went down.
VMRM 1.4	► Supports parent-child dependency and primary-secondary dependency. ► Workgroup support. ► Supports SAP HANA scale-up configuration with host-based replication. ► Supports monitoring of processes at the VM level. ► Supports VIOS health monitoring and application monitoring in the local mode. ► Provides Proactive HA management support. ► Support for HADR deployments provides the capability for one KSYS to manage both HADR operations, and manage from the UI. ► Asymmetric host groups enable M to N server pairing across sites. ► Graphical management updates for HADR. ► Supports HADR policies. ► Supports the creation of workgroups or VM groups for all DR operations from the GUI.

VMRM or IBM GDR version	Main enhancements
VMRM 1.5	► Supporst HA management on both DR sites. ► Expands role-based management. ► Single root i/O virtualization (SR-IOV) override support for DR failovers. ► Health monitoring improvements for HA management. ► Enhances the unmanaged disk feature to support non-replicated disk. ► Supports port-level validation through a tunable. If the port-level validation is enabled, the KSYS subsystem performs port-level validation instead of logical unit number (LUN) level validation. ► Improves the error messages that are displayed by using the **ksysmgr** CLI. ► A DR-only verification option is available at the host group level for HADR and HADRHA configurations. The DR-only verification operation verifies only the DR features and configuration; it does not verify the HA features. ► In the HADRHA cluster type, HA features are enabled at the backup site. The HA feature is enabled after the VMs are migrated to the backup site when a failure occurs at the primary site. ► Supports configuration changes such as add or remove host in a user-defined Shared Storage Pool (SSP) for HADR and HADRHA cluster types. ► The host monitor log collection feature includes enhancements for spooling files. ► The size of KSYS trace log files increased from 4 MB to 20 MB. ► Performance improvement for recovery and cleanup operations. ► You can use custom scripts to configure a site-specific IP address for a Linux VM.
VMRM 1.6	► Disables Live Partition Mobility (LPM) to address potential independent software vendor (ISV) licensing issues. ► Deployment and usability enhancements: Prepopulates network configuration details and configuration error validation. ► VMRM DR support expands up to 1,000 VMs. ► EMC Unity sync replication support. ► Asymmetric SAN paths and ports between source and target site when all paths are redundant. ► HA restart with port-level validation. ► Automatic LUN masking validation is available for EMC and SVC Storage Systems. ► DR support is available for VMs through SR-IOV or vNIC. ► Multiboot disk is supported for DR rehearsal operation. ► During the DR test operation, if the volume protection attribute is enabled on the SVC storage system, the **ksysmgr** command displays a warning message. ► The KSYS subsystem automatically unmanages a VM after the VM is removed from a workgroup. ► A utility is available to restore the KSYS snapshot in multi-node KSYS cluster. ► The **query disk_group status=yes** command displays the disk group status of the SVC storage agent, EMC storage agent, Hitachi storage agent, and EMC Unity storage agent. ► The failover rehearsal discovery operation on the SVC storage subsystem initiates the point-in-time SVC snap operation for DR test activity. Each failover rehearsal discovery (`dr_test`) operation initiates the latest point-in-time copy task through the SVC snap operation. ► The **ksysmgr** command displays a warning message if the specified virtual local area network (VLAN) or vSwitch does not exist in the target site. ► Supports IBM Power10 processor–based systems.
VMRM 1.7	For more information, see Section 3.3, “New features in V1.7” on page 50.

3.3 New features in V1.7

VMRM 1.7 delivers the following enhancements:

- Support for multi-tenant hosting.
- Support for maintaining ISV compliance with LPAR shared processor pools management between the source and target to preserve processor pool configuration on restart at the DR host.
- Enhanced security with multifactor authentication (MFA) for the VMRM GUI.
- A VLAN ID and vSwitch mapping policy can be created at the workgroup level.
- VMRM redundant orchestrator capability (KSYS): The redundant orchestrator enables multiple instances of KSYS to be incorporated into an HADR cluster configuration. If the primary KSYS goes offline, a secondary KSYS node automatically becomes the master.
- VMRM DR workgroup mapping support: VLAN or virtual switch mapping between the source and target at the workgroup level is enabled. VMs in a workgroup no longer require custom scripts to map connections between sites, which make deployment and management more efficient.
- Reduced VMRM DR infrastructure: DR host and target configurations can now have reduced paths to the target so that there are fewer Fibre Channel (FC) adapters on the target side.
- VMRM DR provides more detailed messages: More detailed messaging and enhanced documentation helps enable client installation with minimal outside assistance.

3.4 Installation planning and prerequisites

Before installing VMRM in your environment, you must understand the prerequisite requirements that are described in the following sections.

3.4.1 VM Recovery Manager HA

The following requirements must be met before you can install the VMRM HA solution:

- The KSYS LPAR must be running IBM AIX 7.2 with Technology Level 2 or later.
- Install the latest version of OpenSSL software for the AIX OS (Version 1.0.2.800 or later).
- Use HMC Version 9 Release 9.1.0 or later.
- The LPAR in the host must have one of the following OSs:
  - AIX 7.1 or later
  - PowerLinux (Red Hat Enterprise Linux, SUSE Linux Enterprise Server, and Ubuntu Linux distributions)
  - IBM i 7.1 or later

Table 3-2 lists the host monitor daemon versions and their corresponding VIOS versions.

Table 3-2 Host monitor daemon version and corresponding VIOS versions

Host monitor daemon version	VIOS version	VM Recovery Manager version
1.4.0.0	3.1.1.0	1.4
1.4.0.1	3.1.1.20	1.4 SP1
1.5.0.0	3.1.2.0	1.5
1.5.0.1	3.1.2.20	1.5 SP1
1.6.0.0	3.1.3.14	1.6
1.7.0.0	3.1.4.10	1.7

- Use a VM agent to monitor the VM and applications for AIX, Red Hat Enterprise Linux, and SUSE Linux Enterprise Server.
- Here are the minimum required firmware (FW) levels of IBM Power servers:
  - POWER8 servers that have one of the following firmware levels:
    - FW840.60 or later
    - FW860.30 or later
  - IBM POWER9 servers with FW910 or later
  - Power10 servers with FW1010 or later
- Install the mandatory APAR fixes and VIOS fixes.
- Use Supported Storage agents, as described in VM Recovery Manager HA requirements.
- You must have root authority to perform any installation tasks.
- The KSYS LPAR must have at least a 1-core CPU and 8 GB of memory. These requirements can be higher if you have a large environment of more than 100 LPARs in the data center.
- Deploy AIX rules in the VIOS. The VIOS must have enough available space in the / (root), `/var`, and /usr file systems. Extra CPU and memory resources are needed in each VIOS for VMRM HA management. Add at least a 1-core CPU and 4 GB of memory to the /, `/var`, and `/usr` file systems, apart from the VIOS sizing that you plan to deploy on your production environment, and have at least a 1-core CPU and 10 GB of memory in a scalable environment.
- Ensure that you have enough space in the LPAR so that KSYS file sets can be installed successfully. You must have 30 MB of disk space in the `/opt` directory and 200 MB of disk space in the `/var` directory.
- Check whether a KSYS installation is in progress by using the **`ksysmgr q cl`** command. If the KSYS software was installed previously, you must uninstall the KSYS software.
- For the installation of the VM agent, ensure that each VM meets the following disk space requirements:
  - At least 100 MB disk space in the `/usr` directory to install the VM agent file sets
  - At least 1 GB disk space in the `/var` directory for log files

- Your production environment must have two VIOSs per host. You can have a maximum of 24 VIOSs in a single host group. If more than two VIOSs are in a host, they can be excluded from the KSYS configuration settings.
- For a multi-node KSYS cluster, the same version of OS must be installed on all nodes. If the OS that is running on a node is on an earlier version than the other node, you must specify the node that is running the earlier version of the OS before the node that is running the later version of the OS when you run a command.

3.4.2 VM Recovery Manager DR

The following requirements must be met before you can install the VMRM DR solution:

- The KSYS LPAR must be running IBM AIX 7.2 with Technology Level 1 Service Pack 1 (7200-01-01) or later.
- Install the latest version of the OpenSSL software for the AIX OS. You can download the latest version of OpenSSL software from AIX Web Download Pack Programs. The latest version of the OpenSSL software is also included on the AIX base media.
- Each LPAR in the host must have one of the following OSs:
  - AIX Version 6.1 or later
  - PowerLinux:
    - Red Hat Enterprise Linux (Little Endian, Big Endian) 7.2 or later
    - SUSE Linux Enterprise Server 12.1 or later
    - Ubuntu Linux distribution 16.04 or later
  - IBM i 7.1 or later
- The VMs that are recovered during a disaster situation must be running on POWER7 processor-based servers or later that are PowerVM based systems that are managed by HMCs.

  For HADR and HADRHA support, a POWER7+ processor-based server is required.
- For a multi-node KSYS cluster, the same version of OS must be installed on all nodes. If the OS that is running on a node is on an earlier version than the other node, you must specify the node that is running the earlier version of the OS before the node that is running the later version of the OS when you run a command.
- All VMs that are managed by the VMRM DR solution must use virtual i/O resources through VIOS. The VMs must not be connected to a physical network adapter or any dedicated devices.
- The VIOS must have a Shared Ethernet Adapter (SEA) configuration to bridge to the same Ethernet network between the hosts at the same site.
- The same VLAN must be configured across the site. If a different VLAN is required on the target site, you must update the KSYS configuration for the different VLAN ID on the backup site.
- Ensure a redundant connection from the KSYS to the HMC and from the HMC to VIOS LPARs. Any connectivity issues between the KSYS, the HMC, and VIOS LPARs can disrupt the regular data collection activity and DR operations. To support VLAN and vSwitch at the workgroup level, create the required shared Ethernet adapter on all VIOSs in the VM group.

- The VMRM DR solution supports the following storage devices:
  - EMC storage system

    The VMRM DR solution supports storage devices for the EMC VMAX family (VMAX1, VMAX2, and VMAX3). The EMC storage devices must be SRDF-capable. The EMC storage must have Solutions Enabler SRDF family 8.1.0.0 or later installed. Both SRDF/S (Synchronous) and SRDF/A (Asynchronous) replication modes are supported. The SYMCLI interface on the KSYS node must be the same version or later as the version of the SYMCLI interface on the storage agent.
  - IBM SVC and Storwize Storage Systems

    The VMRM DR solution supports IBM SVC 6.1.0 or later and IBM Storwize V7000 7.1.0 or later. Both Metro Mirror (synchronous) and Global Mirror (asynchronous) modes of data replication are supported across sites.
  - IBM System Storage DS8000 series

    The VMRM DR solution supports IBM DS8700 or later and DS8000 Storage Systems with DSCLI 7.7.51.48 or later. Only the Global Mirror (asynchronous) mode of data replication is supported across sites.
  - Hitachi storage systems

    The VMRM DR solution supports the Hitachi Virtual Storage Platform (VSP) G1000 and Hitachi VSP G400 with CCI 01-39-03/04 and model RAID-Manager/AIX. Both synchronous and asynchronous modes of data replication are supported across sites.
  - IBM XIV Storage System and IBM FlashSystem A9000

    The VMRM DR solution supports IBM XIV Storage System and IBM FlashSystem A9000. Both Metro Mirror (synchronous) and Global Mirror (asynchronous) modes of data replication are supported across sites.
  - EMC Unity Storage System

    The VMRM DR supports EMC Unity storage system 5.0.6.0.6.252 or later. Both synchronous and asynchronous modes of data replication are supported across sites.
  - Use Supported Storage agents, as described in Software requirements.

For more information, see the following documents:

- *IBM VM Recovery Manager HA for Power Systems Version 1.7 Deployment Guide*
- *IBM VM Recovery Manager DR for Power Systems Version 1.7 Deployment Guide*

3.5 IBM VM Recovery Manager installation

This section details the installation process for both the HA and DR versions of VMRM. It also describes how to install the GUI to control your VMRM installation.

3.5.1 Common tasks

This section describes common tasks for installing both VMRM HA and VMRM DR.

Installing the VIOS interim fix

Install the VIOS efixes that are shown in Table 3-3 for the VIOS version that you are using.

Table 3-3 VIOS fixes that are required

VIOS version	APARs	Download URL
3.1.3.10	IJ35732	https://aix.software.ibm.com/aix/efixes/IJ35732/IJ35732m3b.211201.epkg.Z
3.1.1.21	IJ33050	https://aix.software.ibm.com/aix/ifixes/ij33050/ (Unity storage)
	IJ25175, IJ25173, IJ25171, IJ25170, IJ25169, IJ25168, IJ27427, IJ25166, IJ25174, and IJ25165	http://aix.software.ibm.com/aix/efixes/ij25165/IJ25165m2c.200727.epkg.Z
3.1.2.20	IJ33042, IJ33043, IJ33041, IJ33044, IJ33034, and IJ33045	https://aix.software.ibm.com/aix/ifixes/ij33034/z
	IJ33050	https://aix.software.ibm.com/aix/ifixes/ij33050/ (Unity storage)
3.1.1.0	IJ21043 and IJ22767	http://aix.software.ibm.com/aix/ifixes/IJ21043/IJ21043m1b.200218.epkg.Z
3.1.1.25	IJ25175, IJ25173, IJ25171, IJ25170, IJ25169, IJ25168, IJ27427, IJ25166, IJ25174, and IJ25165	http://aix.software.ibm.com/aix/efixes/ij25165/IJ25165m2c.200727.epkg.Z
3.1.1.30	IJ25171	https://aix.software.ibm.com/aix/ifixes/IJ25171/IJ25171s3a.210709.epkg.Z

VIOS version	APARs	Download URL
3.1.2.10	IJ28933, IJ28934, IJ28935, and IJ28937	https://aix.software.ibm.com/aix/efixes/IJ28933/IJ28933m1a.201106.epkg.Z
3.1.1.10	IJ21043 and IJ22767	http://aix.software.ibm.com/aix/ifixes/IJ21043/IJ21043m1b.200218.epkg.Z

For other versions of VMRM HA, install the following VIOS interim fix on all VIOS instances that are included in the KSYS subsystem, as shown in Table 3-4.

Table 3-4 Required interim fix

VMRM version	Required interim fix
1.3	IJ10896m2a.181101.epkg.Z
1.4	IJ21043m1a.191118.epkg.Z
1.5	IJ28933m1a.201106.epkg.Z
1.6	IJ28933m1a.201106.epkg.Z
1.7	None that is known at the time of writing

To install and check for interim fixes on a VIOS, see the following resources:

- Managing an interim fix on a VIOS
- Index of AIX efixes

Installing the KSYS software

To install the KSYS software, complete the following steps:

1. Go to the directory that contains the installation images and list the installation files, as shown in Example 3-1.

Example 3-1 VMRM installation files

```
# ls -alrt
total 1012432
drwxr-xr-x   25 root     system         4096 Nov 14 16:35 ..
-rw-r--r--    1 root     system       188416 Nov 14 16:42 ksys.drutils.rte
-rw-r--r--    1 root     system      1040384 Nov 14 16:42 ksys.ha.license
-rw-r--r--    1 root     system        29696 Nov 14 16:42 ksys.hautils.rte
-rw-r--r--    1 root     system      5746688 Nov 14 16:42 ksys.hsmon.rte
-rw-r--r--    1 root     system      1040384 Nov 14 16:42 ksys.license
-rw-r--r--    1 root     system     12876800 Nov 14 16:43 ksys.main.cmds
-rw-r--r--    1 root     system       237568 Nov 14 16:43 ksys.main.msg.DE_DE.cmds
-rw-r--r--    1 root     system       231424 Nov 14 16:43 ksys.main.msg.ES_ES.cmds
-rw-r--r--    1 root     system       237568 Nov 14 16:43 ksys.main.msg.FR_FR.cmds
-rw-r--r--    1 root     system       222208 Nov 14 16:43 ksys.main.msg.IT_IT.cmds
-rw-r--r--    1 root     system       279552 Nov 14 16:43 ksys.main.msg.JA_JP.cmds
-rw-r--r--    1 root     system       221184 Nov 14 16:43 ksys.main.msg.PT_BR.cmds
-rw-r--r--    1 root     system       224256 Nov 14 16:43 ksys.main.msg.ZH_CN.cmds
```

```
-rw-r--r--    1 root     system       228352 Nov 14 16:43 ksys.main.msg.ZH_TW.cmds
-rw-r--r--    1 root     system       202752 Nov 14 16:43 ksys.main.msg.en_US.cmds
-rw-r--r--    1 root     system     68172800 Nov 14 16:44 ksys.main.rte
-rw-r--r--    1 root     system       171008 Nov 14 16:44 ksys.mirror.ds8k.rte
-rw-r--r--    1 root     system       269312 Nov 14 16:44 ksys.mirror.emc.rte
-rw-r--r--    1 root     system       194560 Nov 14 16:44 ksys.mirror.hitachi.rte
-rw-r--r--    1 root     system       238592 Nov 14 16:44 ksys.mirror.svc.rte
-rw-r--r--    1 root     system       107520 Nov 14 16:44 ksys.mirror.unity.rte
-rw-r--r--    1 root     system       167936 Nov 14 16:44 ksys.mirror.xiv.rte
-rw-r--r--    1 root     system     15281152 Nov 14 16:44 ksys.ui.agent
-rw-r--r--    1 root     system    142644224 Nov 14 16:45 ksys.ui.common
-rw-r--r--    1 root     system    256133120 Nov 14 16:47 ksys.ui.server
-rw-r--r--    1 root     system      9052160 Nov 14 16:47 ksys.vmmon.rte
-rw-r--r--    1 root     system       872324 Nov 14 16:47
vmagent-1.7.0-1.0.el7.ppc64le.rpm
-rw-r--r--    1 root     system      1738680 Nov 14 16:47
vmagent-1.7.0-1.0.suse123.ppc64le.rpm
drwxr-xr-x    2 root     system         4096 Nov 14 18:53 .
-rw-r--r--    1 root     system       266765 Nov 14 18:53 .toc
```

2. Run the installation in verbose mode, as shown in Example 3-2.

Example 3-2 installing the VMRM file sets

```
$ installp -acFXYd /vmrm -V2 ksys.hautils.rte ksys.ha.license ksys.main.cmds
ksys.main.msg.en_US.cmds ksys.main.rte ksys.ui.agent ksys.ui.common
+-----------------------------------------------------------------------------+
                    Pre-installation Verification...
+-----------------------------------------------------------------------------+
Verifying selections...done
Verifying requisites...done
Results...

SUCCESSES
---------
  Filesets listed in this section passed pre-installation verification
  and will be installed.
    -- Filesets are listed in the order in which they will be installed.
    -- The reason for installing each fileset is indicated with a keyword
       in parentheses and explained by a "Success Key" following this list.

  ksys.ui.common 1.7.0.0 (Selected)
    VMRestart User Interface - common part
  ksys.ui.agent 1.7.0.0 (Selected)
    VMRestart User Interface - agent part
  ksys.main.rte 1.7.0.0 (Selected)
    Base Server Runtime
  ksys.main.msg.en_US.cmds 1.7.0.0 (Selected)
    Base Server Runtime - US English
  ksys.main.cmds 1.7.0.0 (Selected)
    Base Server Runtime
  ksys.hautils.rte 1.7.0.0 (Selected)
    Base Server Runtime
  ksys.ha.license 1.7.0.0 (Selected)
    Base Server Runtime
```

```
(....)

Finished processing all file sets.  (Total time:  5 mins 31 secs).

Please wait...

        /opt/rsct/install/bin/ctposti
0513-059 The ctrmc Subsystem has been started. Subsystem PID is 5898442.
0513-059 The IBM.ConfigRM Subsystem has been started. Subsystem PID is 7078008.
done
+-----------------------------------------------------------------------------+
                                Summaries:
+-----------------------------------------------------------------------------+

Installation Summary
--------------------
Name                          Level           Part        Event       Result
-------------------------------------------------------------------------------
ksys.ui.common                1.7.0.0         USR         APPLY       SUCCESS
ksys.ui.agent                 1.7.0.0         USR         APPLY       SUCCESS
ksys.ui.agent                 1.7.0.0         ROOT        APPLY       SUCCESS
ksys.main.rte                 1.7.0.0         USR         APPLY       SUCCESS
ksys.main.rte                 1.7.0.0         ROOT        APPLY       SUCCESS
ksys.main.msg.en_US.cmds      1.7.0.0         USR         APPLY       SUCCESS
ksys.main.cmds                1.7.0.0         USR         APPLY       SUCCESS
ksys.main.cmds                1.7.0.0         ROOT        APPLY       SUCCESS
ksys.hautils.rte              1.7.0.0         USR         APPLY       SUCCESS
ksys.hautils.rte              1.7.0.0         ROOT        APPLY       SUCCESS
ksys.ha.license               1.7.0.0         USR         APPLY       SUCCESS
```

3. Verify that the installation of the file sets was successful by running the command that is shown in Example 3-3. The expected results are also shown.

Example 3-3 Verifying the file sets installation

```
$ lslpp -l ksys.ha.license ksys.hautils.rte ksys.main.cmds
ksys.main.msg.en_US.cmds ksys.main.rte
  File set                      Level  State      Description
  ----------------------------------------------------------------------------
Path: /usr/lib/objrepos
  ksys.ha.license            1.7.0.0  COMMITTED  Base Server Runtime
  ksys.hautils.rte           1.7.0.0  COMMITTED  Base Server Runtime
  ksys.main.cmds             1.7.0.0  COMMITTED  Base Server Runtime
  ksys.main.msg.en_US.cmds   1.7.0.0  COMMITTED  Base Server Runtime - US
                                                 English
  ksys.main.rte              1.7.0.0  COMMITTED  Base Server Runtime

Path: /etc/objrepos
  ksys.hautils.rte           1.7.0.0  COMMITTED  Base Server Runtime
  ksys.main.cmds             1.7.0.0  COMMITTED  Base Server Runtime
  ksys.main.rte              1.7.0.0  COMMITTED  Base Server Runtime
```

4. Check the version of the KSYS software, as shown in Example 3-4.

Example 3-4 Checking the KSYS version

```
/opt/IBM/ksys/ksysmgr query version
Ksysmgr version: 1.7.0.0
Ksys version: 1.7.0.0
```

5. After the successful installation of the KSYS file sets, run a command to check whether the class IDs are reserved. The command and its expected results are shown in Example 3-5.

Example 3-5 Checking the class IDs for IBM VMRM

```
$ cat /usr/sbin/rsct/cfg/ct_class_ids | grep IBM.VMR

IBM.VMR_HMC                                     510
IBM.VMR_CEC                                     511
IBM.VMR_LPAR                                    512
IBM.VMR_VIOS                                    513
IBM.VMR_SSP                                     514
IBM.VMR_SITE                                    515
IBM.VMR_SA                                      516
IBM.VMR_DP                                      517
IBM.VMR_DG                                      518
IBM.VMR_KNODE                                   519
IBM.VMR_KCLUSTER                                520
IBM.VMR_HG                                      521
IBM.VMR_APP                                     522
IBM.VMR_CLOUD                                   523
IBM.VMR_DP_CLD                                  524
IBM.VMR_SA_CLD                                  525
IBM.VMR_LPAR_CLD                                526
IBM.VMR_SITE_CLD                                527
IBM.VMR_VMG_CLD                                 528
IBM.VMR_APP_CLD                                 529
```

Now that the KSYS LPAR is installed, you can continue with installing the GUI.

Installing the GUI server file sets

The GUI server can be run either on a KSYS node or on another node in the cluster. To install the GUI on a KSYS node, continue with “Installing GUI server file sets on a KSYS node” on page 59. If you are installing the GUI on a non-KSYS server, go to “Installing open-source software packages” on page 60.

Note: The LPAR in which you want to install the GUI file sets must be running IBM AIX 7.2 with Technology Level 2 Service Pack 1 (7200-02-01) or later.

Installing GUI server file sets on a KSYS node

To install GUI server file sets on one of the KSYS nodes, use the command that is shown in Example 3-6, which also shows the output from the command.

Example 3-6 Installing the VMRM GUI server file sets

```
$ installp -acFXYd /vmrm -V2  ksys.ui.server
+-----------------------------------------------------------------------------+
                    Pre-installation Verification...
+-----------------------------------------------------------------------------+
Verifying selections...done
Verifying requisites...done
Results...
SUCCESSES
---------
  Filesets that are listed in this section passed pre-installation verification
  and will be installed.
    -- Filesets are listed in the order in which they will be installed.
    -- The reason for installing each file set is indicated with a keyword
       in parentheses and explained by a "Success Key" following this list.
  ksys.ui.server 1.7.0.0 (Selected)
    VMRestart User Interface - server part
  Success Key:
  Selected    -- Explicitly selected by user for installation.
  Maintenance -- Maintenance Level file set update; being installed
                 automatically to enable the level of the system to be
                 tracked.
  Mandatory   -- Considered to be important to the system; will always
                 be installed when detected on the installation media.
  Requisite   -- Requisite of other file sets being installed.
  P_Requisite -- Previously installed file set's requisite; being installed
                 automatically now to ensure system's consistency.  (Only
                 installed automatically when "auto-install" (-g flag)
                 is specified.)
  Supersedes  -- Superseding file set update; not selected, chosen instead
                 of an older, selected update.  (Only chosen in this fashion
                 when "auto-install" is specified (-g flag)).
  << End of Success Section >>
+-----------------------------------------------------------------------------+
                   BUILDDATE Verification ...
+-----------------------------------------------------------------------------+
Verifying build dates...done
+-----------------------------------------------------------------------------+
                   BUILDDATE Verification ...
+-----------------------------------------------------------------------------+
Verifying build dates...done
FILESET STATISTICS
------------------
    1  Selected to be installed, of which:
        1  Passed pre-installation verification
  ----
    1  Total to be installed
+-----------------------------------------------------------------------------+
                         Installing Software...
+-----------------------------------------------------------------------------+
installp:  APPLYING software for:
```

```
        ksys.ui.server 1.7.0.0

(....)
Generating RSA private key, 4096 bit long modulus
.....................................................................................
.....................................................................+++
.........+++
e is 65537 (0x10001)
writing RSA key
Creating the "vmruiserver" subsystem
0513-071 The vmruiserver Subsystem has been added.
Creating an inittab entry for vmruiserver
Finished processing all file sets.  (Total time:  13 mins 16 secs).
+-----------------------------------------------------------------------------+
                                Summaries:
+-----------------------------------------------------------------------------+
Installation Summary
--------------------
Name                        Level           Part        Event       Result
-------------------------------------------------------------------------------
ksys.ui.server              1.7.0.0         USR         APPLY       SUCCESS
ksys.ui.server              1.7.0.0         ROOT        APPLY       SUCCESS
```

After this task is complete, go to “Installing open-source software packages”.

Installing GUI server file sets on a non-KSYS node

To install GUI server file sets on a separate system that manages all the KSYS nodes, run the command that is shown in Example 3-7.

Example 3-7 Installing GUI server file sets on a non-KSYS node

```
$ installp -acFXYd /vmrm -V2   ksys.ha.license ksys.ui.server ksys.ui.common
```

After this step is complete, go to “Installing open-source software packages”.

Installing open-source software packages

Install the open-source software packages. The process is different based on whether your system has access to the internet or not.

- If the GUI server LPAR is connected to the internet.

  Use the command that is shown in Example 3-8. The command and the expected results are displayed.

Example 3-8 Using vmruiinst.ksh

```
$ /opt/IBM/ksys/ui/server/dist/server/bin/vmruiinst.ksh
Warning: "/tmp/vmruiinst.downloads" does not exist. Creating...
"/tmp/vmruiinst.downloads" has been created.

Checking if the requisites have already been downloaded...
        ** "info-6.6-2.aix6.1.ppc.rpm" needs to be retrieved.
        ** "cpio-2.13-1.aix6.1.ppc.rpm" needs to be retrieved.
        ** "readline-8.0-2.aix6.1.ppc.rpm" needs to be retrieved.
        ** "libiconv-1.16-1.aix6.1.ppc.rpm" needs to be retrieved.
        ** "bash-5.0.18-1.aix6.1.ppc.rpm" needs to be retrieved.
        ** "gettext-0.20.2-1.aix6.1.ppc.rpm" needs to be retrieved.
```

```
        ** "libgcc-8.3.0-2.aix7.3.ppc.rpm" needs to be retrieved.
        ** "libstdcplusplus-8.3.0-2.aix7.3.ppc.rpm" needs to be retrieved.

        Total number of requisite downloads needed: 8

/tmp/vmruiinst.downloads/ ...
Checking for sufficient download space in "/tmp"...
The disk space needed for the required download(s) is 28 MB in /tmp.
The available space in /tmp is 510 MB.
The available space is sufficient.

Checking for sufficient download space in "/opt"...
The disk space needed for the required download(s) is 39 MB in /opt.
The available space in /opt is 510 MB.
The available space is sufficient.

Checking for sufficient installation space in "/opt"...
The disk space needed for the required install is 36 MB in /opt.
The available space in /opt is 132 MB.
The available space is sufficient.

To use the VM Recovery Manager HADR for AIX GUI, you
must install third-party files. These third-party files were not included in
the server file sets because they are licensed under the General Public License
(GPL). This script downloads and installs the files that are required by
SQLite and Node.js.

    SQLite dependencies:                      | Node.js dependencies:
    ========================================= | ======================
        bash                info              | libgcc
        libiconv            readline          | libstdc++
        gettext             cpio              |

The files that are required for Node.js are used on the server, and are also
installed on each cluster node automatically during the cluster discovery
operation.

IBM does not offer support for these files if the files are used outside the
context of the VM Recovery Manager HADR GUI.

Do you want to download and install these files? (y/n)  y

***
    The downloads have started and can take several minutes.
    Please wait for the downloads to complete.
***

(NOTE: the progress bar is a best-guess approximation,
       and might not be perfectly accurate)

    Downloading "/tmp/vmruiinst.downloads/info-6.6-2.aix6.1.ppc.rpm"...
    Downloading "/tmp/vmruiinst.downloads/cpio-2.13-1.aix6.1.ppc.rpm"...
    Downloading "/tmp/vmruiinst.downloads/readline-8.0-2.aix6.1.ppc.rpm"...
    Downloading "/tmp/vmruiinst.downloads/libiconv-1.16-1.aix6.1.ppc.rpm"...
    Downloading "/tmp/vmruiinst.downloads/bash-5.0.18-1.aix6.1.ppc.rpm"...
```

```
    Downloading "/tmp/vmruiinst.downloads/gettext-0.20.2-1.aix6.1.ppc.rpm"...
    Downloading "/tmp/vmruiinst.downloads/libgcc-8.3.0-2.aix7.3.ppc.rpm"...
    Downloading
"/tmp/vmruiinst.downloads/libstdcplusplus-8.3.0-2.aix7.3.ppc.rpm"...

Adding execution permission to the PAM module
(/opt/IBM/ksys/ui/server/lib/auth/smuiauth)
Configuring the PAM system for authentication support for our module.
Attempting to start the server...
The server was successfully started.
The installation completed successfully. To use the VM Recovery Manager HADR GUI,
open a web browser and enter the following URL:
    https://192.168.100.23:3000/login
After you log in, you can add existing clusters in your environment to the
VM Recovery Manager HADR GUI.
Log collected Successfully at
/tmp/GUILOGS/vmruiinst.log
```

- If the GUI server LPAR is configured to use an HTTP proxy to access the internet.

  Use the command that is shown in Example 3-9.

Example 3-9 HTTP proxy installation

```
/opt/IBM/ksys/ui/server/dist/server/bin/vmruiinst.ksh -p
```

- If the GUI server LPAR is not connected to the internet.

  You can copy and run the **vmruiinst.ksh** scripts to download the package from any host that runs AIX and that has internet access. Download the following packages:

  - info-4.13-3.aix5.3.ppc.rpm
  - cpio-2.11-2.aix6.1.ppc.rpm
  - readline-6.2-2.aix5.3.ppc.rpm
  - libiconv-1.13.1-2.aix5.3.ppc.rpm
  - bash-4.2-5.aix5.3.ppc.rpm
  - gettext-0.17-6.aix5.3.ppc.rpm
  - libgcc-4.9.2-1.aix6.1.ppc.rpm
  - libgcc-4.9.2-1.aix7.1.ppc.rpm
  - libstdc++-4.9.2-1.aix6.1.ppc.rpm
  - libstdc++-4.9.2-1.aix7.1.ppc.rpm

  In the GUI server LPAR, run the command that is shown in Example 3-10.

Example 3-10 Non-internet connected installation

```
vmruiinst.ksh -i /vmrm
```

  /vmrm is the location where you copied the downloaded files.

To start the server and validate that the port is open, run the commands that are shown in Example 3-11.

Example 3-11 Starting the VMRM GUI

```
# startsrc -s vmruiserver
0513-059 The vmruiserver Subsystem has been started. Subsystem PID is 15597690.
#
# netstat -an | grep 3000
tcp        0      0  *.3000                 *.*                    LISTEN
```

Note: If you install VMRM HA GUI on AIX 7.3, manually download and install the `libgcc` and `libstdc++` libraries that are compatible with the AIX 7.3 OS.

3.6 Migrating and upgrading from previous releases

Before starting the migration process, download the required VMRM HA and VMRM DR packages by completing the following steps:

1. Download the VMRM HA packages from IBM Entitled Systems Support by using the following link:

   https://www.ibm.com/servers/eserver/ess/landing/index.html

2. Decompress the file sets. The KSYS package consists of the following file sets:

   - `ksys.main.rte`
   - `ksys.main.msg.en_US.cmds`
   - `ksys.license`
   - `ksys.ha.license`
   - `ksys.hautils.rte`
   - `ksys.ui.agent`
   - `ksys.ui.common`
   - `ksys.ui.server`
   - `ksys.vmmon.rte`
   - `ksys.drutils.rte`
   - `ksys.mirror.unity.rte`
   - `./RPMS/linux/vmagent-1.7.0-1.0.el7.ppc64le.rpm`
   - `./RPMS/linux/vmagent-1.7.0-1.0.suse123.ppc64le.rpm`

3.6.1 Upgrading VM Recovery Manager HA to Version 1.7.0

To upgrade the VMRM HA solution to the latest version, you must upgrade the KSYS software and the VM agents in the VMs. To upgrade the VMRM HA solution to Version 1.7.0, follow the steps in these sections:

- Upgrading the KSYS software
- Upgrading VM agents

Upgrading the KSYS software

Complete the following steps:

1. Log in as root.
2. Check the status of the KSYS-related operations to ensure that the KSYS software is not running any active operations, as shown in Example 3-12.

Example 3-12 Checking the status of KSYS

```
$ ksysmgr query system status monitor=yes
```

3. Stop the `IBM.VMR` daemon, as shown in Example 3-13.

Example 3-13 Stopping the IBM.VMR daemon

```
$ stopsrc -s IBM.VMR
```

4. Use SMIT to update the installed software to the latest level, as shown in Example 3-14.

Example 3-14 Using SMIT to install and update software

```
$ smit install

Install and Update Software

Move cursor to wanted item and press Enter.

  Install Software
  Update Installed Software to Latest Level (Update All)
  Update Installed Software to Latest Level (Live Update)
  Install Software Bundle
  Update Software by Fix (APAR)
  Install and Update from ALL Available Software


F1=Help             F2=Refresh          F3=Cancel           F8=Image
F9=Shell            F10=Exit            Enter=Do
```

5. In the Update Installed Software to Latest Level (Update All) panel, accept the new license agreements, as shown in Example 3-15. Press Enter after you make all the changes.

Example 3-15 Updating the installed software to the latest level by using SMIT

```
Type or select values in entry fields.
Press Enter AFTER making all wanted changes.

[TOP]                                                 [Entry Fields]
* INPUT device / directory for software              .
* SOFTWARE to update                                 _update_all
  PREVIEW only? (update operation will NOT occur)    no                  +
  COMMIT software updates?                           yes                 +
  SAVE replaced files?                               no                  +
  AUTOMATICALLY install requisite software?          yes                 +
  EXTEND file systems if space needed?               yes                 +
  VERIFY install and check file sizes?               no                  +
  DETAILED output?                                   no                  +
  Process multiple volumes?                          yes                 +
  ACCEPT new license agreements?                     yes                 +
  PREVIEW new LICENSE agreements?                    no                  +

[MORE...6]

F1=Help             F2=Refresh          F3=Cancel           F4=List
F5=Reset            F6=Command          F7=Edit             F8=Image
F9=Shell            F10=Exit            Enter=Do
```

6. Start the `IBM.VMR` daemon, as shown in Example 3-16.

Example 3-16 Starting the IBM.VMR daemon

```
$ startsrc -s IBM.VMR
```

7. Run the script that is shown in Example 3-17 on page 65 on the GUI server.

Example 3-17 Running the script in the GUI server

```
$ opt/IBM/ksys/ui/server/dist/server/bin/vmruiinst.ksh
```

8. Stop the GUI server, as shown in Example 3-18.

Example 3-18 Stopping the GUI server

```
$ stopsrc -s vmruiserver
```

9. Start the GUI server, as shown in Example 3-19.

Example 3-19 Starting the GUI server

```
startsrc -s vmruiserver
```

Upgrading VM agents

To upgrade the VM agent RPM packages, complete one of the following tasks:

- For Red Hat Enterprise Linux (Little Endian) VMs, run the rpm upgrade that is shown in Example 3-20.

Example 3-20 Upgrading the VM agent in Red Hat Enterprise Linux

```
$ rpm -Uvh vmagent-1.7.0-1.0.el7.ppc64le.rpm
```

- For SUSE Linux Enterprise Server (Little Endian) VMs, run the rpm upgrade that is shown in Example 3-21.

Example 3-21 Upgrading the VM agent for SUSE Linux Enterprise Server

```
$ rpm -Uvh vmagent-1.7.0-1.0.suse123.ppc64le.rpm
```

3.6.2 Upgrading VM Recovery Manager DR to Version 1.7.0

To upgrade the VMRM DR solution to the latest version, you must upgrade the KSYS software. To upgrade the VMRM DR solution to Version 1.7.0, complete the following steps:

1. Connect as root.
2. Check the status of KSYS-related operations to ensure that the KSYS software is not running any active operations, as shown in Example 3-22.

Example 3-22 Checking the status of KSYS

```
$ ksysmgr query system status monitor=yes
```

3. Stop the `IBM.VMR` daemon, as shown in Example 3-23.

Example 3-23 Stopping the IBM.VMR daemon

```
stopsrc -s IBM.VMR
```

4. Use SMIT to update the installed software to the latest level, as shown in Example 3-24.

Example 3-24 Using SMIT to install and update the software

```
$ smit install

Install and Update Software

Move cursor to wanted item and press Enter.

  Install Software
  Update Installed Software to Latest Level (Update All)
  Update Installed Software to Latest Level (Live Update)
  Install Software Bundle
  Update Software by Fix (APAR)
  Install and Update from ALL Available Software


F1=Help             F2=Refresh          F3=Cancel           F8=Image
F9=Shell            F10=Exit            Enter=Do
```

5. In the Update Installed Software to Latest Level (Update All) panel, accept the new license agreements. Press Enter after you make all your changes, as shown in Example 3-25.

Example 3-25 Updating the installed software to the latest level by using SMIT

```
Type or select values in entry fields.
Press Enter AFTER making all wanted changes.

[TOP]                                                 [Entry Fields]
* INPUT device / directory for software              .
* SOFTWARE to update                                 _update_all
  PREVIEW only? (update operation will NOT occur)    no                   +
  COMMIT software updates?                           yes                  +
  SAVE replaced files?                               no                   +
  AUTOMATICALLY install requisite software?          yes                  +
  EXTEND file systems if space needed?               yes                  +
  VERIFY install and check file sizes?               no                   +
  DETAILED output?                                   no                   +
  Process multiple volumes?                          yes                  +
  ACCEPT new license agreements?                     yes                  +
  PREVIEW new LICENSE agreements?                    no                   +

[MORE...6]

F1=Help             F2=Refresh          F3=Cancel           F4=List
F5=Reset            F6=Command          F7=Edit             F8=Image
F9=Shell            F10=Exit            Enter=Do
```

6. Start the `IBM.VMR` daemon, as shown in Example 3-26.

Example 3-26 Starting the IBM.VMR daemon

```
startsrc -s IBM.VMR
```

7. Run the script on the GUI server, as shown in Example 3-27 on page 67.

Example 3-27 Running the script in the GUI server

```
$ opt/IBM/ksys/ui/server/dist/server/bin/vmruiinst.ksh
```

8. Stop the GUI server, as shown in Example 3-28.

Example 3-28 Stopping the GUI server

```
$ stopsrc -s vmruiserver
```

9. Start the GUI server, as shown in Example 3-29.

Example 3-29 Starting the GUI server

```
$ startsrc -s vmruiserver
```

After the upgrade operation is complete, you must run a discovery operation before you add, delete, or modify any configuration settings in the KSYS subsystem.

Notes:

- To upgrade the KSYS software in a KSYS LPAR with PowerHA configured, stop the PowerHA cluster services on all KSYS nodes, and upgrade the KSYS file set on all KSYS nodes before starting the cluster services.
- Migrating from Version 1.4 to Version 1.7 is not supported. First upgrade to Version 1.5 and then upgrade to Version 1.6.
- Migrating from Version 1.5 to Version 1.7 is not supported. First upgrade to Version 1.6.

3.6.3 Upgrading KSYS file sets for KSYS high availability

Before VMRM 1.7, providing HA for a KSYS LPAR required a clustering solution, namely PowerHA SystemMirror. Version 1.7 provides HA features on your KSYS natively within VMRM without needing PowerHA.

If you want to upgrade your KSYS to take advantage of the new HA option in VMRM 1.7, complete the following steps:

1. Check the group leader (GL) node through the KSYS subsystem by running the following command:

   ```
   lssrc -ls IBM.VMR | grep Group
   ```

2. Check the GL node through the ConfigRM Reliable Scalable Cluster Technology (RSCT) tool by running the following command:

   ```
   lssrc -ls IBM.ConfigRM | grep Group
   ```

3. On the GL, run the PowerHA SystemMirror operations by using the **`smit clstop`** or **`clmgr`** command with the resource group (RG) **`offline`** option. This command stops all PowerHA SystemMirror services on both nodes.

4. The PowerHA SystemMirror cluster state should be `INIT` on both nodes. Sometimes, a conflict occurs in the cluster state of the GL node that is displayed in the KSYS subsystem and the GL node that is displayed in RSCT. Ignore this conflict.

   a. To view the GL node, run the following command in the KSYS subsystem:

      ```
      lssrc -ls IBM.VMR | grep Group
      ```

   b. To view the GL node, run the following command in RSCT:

      ```
      lssrc -ls IBM.ConfigRM | grep Group
      ```

   The output of these commands might show different GLs, which is expected.

5. To start the IBM.VMR daemon on the first node (for example, Node1), run the following command:

   ```
   startsrc -s IBM.VMR
   ```

6. Upgrade the VMRM HA solution on the first node (for example, Node1) from Version 1.6 to Version 1.7.

7. To verify the version of VMRM HA on the first node, run the following command:

   ```
   lslpp -l | grep ksys
   ```

   You cannot use the run **`ksysmgr query version`** command on the GL node.

8. Upgrade the VMRM HA solution on the other node (for example, Node2) from Version 1.6 to Version 1.7. The `IBM.VMR` daemon becomes stable on both nodes after the upgrade operation completes. The GL node in the KSYS subsystem changes and the Node1 becomes the GL. Because the daemon was stopped and then restated during upgrade operations, the restart operation of the daemon triggers the failover operation.

9. Check the GL node both from RSCT and the KSYS subsystem. The KSYS subsystem and RSCT both point to the same GL node. The HA feature is active and monitoring the resources.

10. To use the HA features of VMRM 1.7, do not use PowerHA SystemMirror services. You can view resources of both nodes (for example, Node1 and Node2) of the cluster by using the **`lsrsrc IBM.VMR_SITE_CLD`** command. If the resources of only one node are displayed, restart the `IBM.VMR` daemon on both nodes.

    To stop the `IBM.VMR` daemon, run the following command:

    ```
    stopsrc -s IBM.VMR -c ; sleep 10
    ```

    To start the `IBM.VMR` daemon, run the following command:

    ```
    startsrc -s IBM.VMR
    ```

4

IBM PowerHA SystemMirror Standard Edition and PowerHA SystemMirror Enterprise Edition for AIX

This chapter provides an introduction to IBM PowerHA SystemMirror (both Standard and Enterprise Editions) for newcomers to this solution and a refresher for those users that implemented PowerHA SystemMirror and used it for many years. There also is a focus on new features that are available in the most recent version of PowerHA SystemMirror. For more information about these features and their setup and operation, see *IBM PowerHA SystemMirror for AIX Cookbook*, SG24-7739.

This chapter describes the following topics:

- Terminology and concepts
- History and evolution
- New features in PowerHA SystemMirror 7.2.7
- New features in PowerHA GUI 7.2.7
- Installation prerequisites
- Additional PowerHA resources

4.1 Terminology and concepts

PowerHA SystemMirror for AIX (also referred to as PowerHA) is the IBM Power data center solution that helps protect critical business applications from outages, both planned and unplanned. One of the major objectives of PowerHA is to offer automatically continued business services by providing redundancy despite different component failures. PowerHA depends on Reliable Scalable Cluster Technology (RSCT) and Cluster Aware AIX (CAA).

RSCT is a set of low-level operating system (OS) components that allow the implementation of clustering technologies. RSCT is distributed with AIX. On AIX 7.2, RSCT is Version 3.2.1.0, and AIX 7.3 includes RSCT Version 3.3.0.0. After installing the PowerHA and CAA file sets, the RSCT topology services subsystem is deactivated, and all its functions are performed by CAA.

PowerHA 7.2 or later relies heavily on the CAA infrastructure that was introduced in AIX 6.1 TL6 (which is not supported anymore) and AIX 7.1 (supported from Technology Level 3 with Service Pack 9 or later). CAA provides communication interfaces and monitoring provisions for PowerHA and execution by using CAA commands with **`clcmd`**.

PowerHA SystemMirror Enterprise Edition also provides disaster recovery (DR) functions, such as cross-site mirroring, IBM HyperSwap, Geographic Logical Volume Manager (GLVM), and many storage-based replication methods. These cross-site clustering methods support PowerHA functions between two geographic sites. For more information, see *IBM PowerHA SystemMirror for AIX Cookbook*, SG24-7739.

For more information about features that were added, see 4.2, “History and evolution” on page 73.

While many general concepts were described in Chapter 1, “Introducing high availability and disaster recovery” on page 1, the terminology that is used to describe PowerHA configuration and operation continues to evolve. The following terms are used when describing PowerHA solutions:

Node
An IBM Power server (or logical partition (LPAR)) running AIX and PowerHA that are defined as part of a cluster. Each node has a collection of resources (disks, file systems, IP addresses, and applications) that can be transferred to another node in the cluster if the node or a component fails.

Cluster
A loosely coupled collection of independent systems (nodes) or LPARs that are organized into a network for sharing resources and communicating with each other.

PowerHA defines relationships among OSs where peer cluster nodes provide the services that are offered by a cluster node if that node cannot do so. These individual nodes are responsible for maintaining the functions of one or more applications in a failure of any cluster component.

Client
A client is a system that can access the application running on the cluster nodes over a local area network (LAN). Clients run a client application that connects to the server (node) where the application runs.

Topology
Contains basic cluster components nodes, networks, communication interfaces, and communication adapters.

Resources	Logical components or entities that are being made highly available (HA) (for example, file systems, raw devices, service IP labels, and applications) by being moved from one node to another node. All resources that together form a high availability (HA) application or service are grouped in resource groups (RGs). PowerHA keeps the RG HA as a single entity that can be moved from node to node if a component or node fail. RGs can be available from a single node, or in concurrent applications, available simultaneously from multiple nodes. A cluster can host more than one RG, thus allowing for efficient usage of the cluster nodes.
Dependencies	With PowerHA, you can define dependencies and relationships between RGs that can be used to control their location, order of processing, and whether to bring online or take offline depending on the state of other resources.
Service IP label	A label that matches to a service IP address and is used for communications between clients and the node. A service IP label is part of an RG, which means that PowerHA can monitor it and keep it HA.
IP address takeover (IPAT)	The process where an IP address is moved from one adapter to another adapter on the same logical network. This adapter can be on the same node or another node in the cluster. If aliasing is used as the method of assigning addresses to adapters, then more than one address can be on a single adapter.
Resource takeover	The operation of transferring resources between nodes inside the cluster. If one component or node fails because of a hardware or OS problem, its RGs are moved to another node.
Fallover	Represents the movement of an RG from one active node to another node (backup node) in response to a failure on the active node or in the environment affecting the active node.
Fallback	Represents the movement of an RG back from the backup node to the previous node when it becomes available. This movement is typically in response to the reintegration of the previously failed node.
Heartbeat packet	A packet that is sent between communication interfaces in the cluster and used by the various cluster daemons to monitor the state of the cluster components (nodes, networks, and adapters).
RSCT daemons	These daemons consist of two types of processes: topology and Group Services. PowerHA uses Group Services, but depends on CAA for topology services. The cluster manager receives event information that is generated by these daemons and takes corresponding (response) actions in any failure.

Smart assists

A set of HA agents that are bundled with the PowerHA SystemMirror Standard Edition to help discover and define HA policies for the most common middleware products.

Split-brain or split-cluster

A cluster split-brain can occur when a subset of nodes in a cluster cannot communicate with the remaining nodes. Although it is possible for this situation to occur within the data center, it is more likely to happen to a cluster across data centers due to the greater exposure of the interconnecting networks to potential risk.

In a split-brain situation, the two partitions have no knowledge of each other's status, and each of them believe that the nodes in the other partition are offline. Therefore, each partition tries to bring online the other partition's applications and access the shared resources, which is an action that is likely to result in lost or corrupted data on the shared storage.

Tie breaker or third site

In high availability and disaster recovery (HADR) clusters, it is a best practice to use a tie breaker or a third site to prevent a split-brain situation. Although it is still important to avoid this situation for clusters within a single data center, it is less likely because multiple communication paths connect all nodes in the cluster, which is a less common situation between sites.

The tie-breaker feature uses a tie-breaker resource to choose which partition survives and continues to operate when a cluster split-brain situation occurs. This feature prevents data corruption on the shared or replicated disks.

PowerHA SystemMirror uses tie-breaker disks or a Network File System (NFS) share file to act as the tie-breaker and split-merge policies to control the behavior of the cluster.

Split policy

When a split-brain situation occurs, each partition attempts to acquire the tie breaker by placing a lock on the tie-breaker disk or on the NFS file. The partition that holds the lock on the SCSI disk or reserves the NFS file wins, and the other loses.

All nodes in the winning partition continue to process cluster events, and all nodes in the losing partition attempt to recover according to the defined split and merge action plan. This plan most often implies either the restart of the cluster nodes or the restart of cluster services on those nodes.

Merge policy

There are situations where, depending on the cluster split-brain policy, the cluster can have two partitions that run independent of each other. However, most often, it is a best practice to configure a merge policy that allows the partitions to operate together again after communications are restored between them.

In this second approach, when partitions that were part of the cluster are brought back online after the communication failure, they must be able to communicate with the partition that owns the tie-breaker disk or NFS file. If a partition that is brought back online cannot communicate with the tie-breaker disk or the NFS file, it does not join the cluster. The tie-breaker device is released when all nodes in the configuration have rejoined the cluster.

The merge policy configuration must be the same type as the one for the split policy, for example, it uses an NFS-based tie breaker.

Synchronous replication Writes are committed at the remote storage before an acknowledgment can be returned to the application. This delay degrades the application performance and limits the distance between the application and the remote storage to ~ 80 - 120 km.

Asynchronous replication Writes are cached locally in some form of non-volatile storage, and an acknowledgment is returned to the application. Later, the write is committed to the remote storage, and then the record is removed from the local cache. Asynchronous mode allows for greater distances between sites, smoother peaks in I/O, and a lower bandwidth network. However, in a disaster, data is lost, with the cache size representing the maximum amount of data that can potentially be lost.

4.2 History and evolution

IBM High Availability Cluster Multi-Processing (HACMP) development started in 1990 to provide HA solutions for applications running on IBM RS/6000 servers. We do not provide information about the early releases, which are no longer supported or were not in use at the time of writing. Instead, we provide highlights about the most recent versions.

HACMP was designed as a stand-alone product (known as HACMP Classic). After the IBM HA infrastructure (known as RSCT) became available, HACMP adopted this technology and became HACMP Enhanced Scalability (HACMP/ES) because it provided performance and functional advantages over the Classic version. Starting with HACMP 5.1, there are no more Classic versions. Later, HACMP terminology was replaced with PowerHA starting with Version 5.5 and then PowerHA SystemMirror in Version 6.1.

Starting with PowerHA 7.1, the CAA feature of the OS is used to configure, verify, and monitor the cluster services. This major change improves the reliability of PowerHA because the cluster service functions now run in the kernel space rather than in user space. CAA was introduced in AIX 6.1 TL6. At the time of writing, the current release is PowerHA 7.2.7, which was announced on 11 October 2022 and was available starting on 9 December 2022.

4.2.1 PowerHA SystemMirror 7.2.0

Released in December 2015, PowerHA 7.2 continued the development of PowerHA SystemMirror by adding further improvements in management, configuration simplification, automation, and performance areas. The previous version (Version 7.1) was withdrawn from marketing in July 2017, and service was continued in April 2018. The following list summarizes the improvements in PowerHA 7.2:

- Resiliency enhancements:
  - Integrated support for AIX Live Kernel Update (LKU)
  - Automatic Repository Replacement (ARR)
  - Verification enhancements
  - Exploitation of Logical Volume Manager (LVM) rootvg failure monitoring
  - Live Partition Mobility (LPM) automation
- CAA enhancements:
  - Network Failure Detection Tunable per interface
  - Built-in netmon logic
  - Traffic simulation for better interface failure detection
- Enhanced split-brain handling:
  - Quarantine protection against “sick but not dead” nodes
  - NFS TieBreaker support for split and merge policies
- Resource Optimized failovers by using Enterprise Pools (Resource Optimized High Availability (ROHA))
- Non-disruptive upgrades

The Systems Director plug-in was discontinued in PowerHA 7.2.0.

4.2.2 PowerHA SystemMirror 7.2.1

Released in December 2016, PowerHA 7.2.1 added the following improvements:

- Verification enhancements, some that are carried over from Version 7.2.0:
  - The reserve policy value must not be single path.
  - Checks for the consistency of `/etc/filesystem`, that is, do mount points exist and so on.
  - LVM physical volume identifier (PVID) checks across LVM and Object Data Manager (ODM) on various nodes.
  - Uses AIX Runtime Expert checks for LVM and NFS.
  - Checks for network errors. If they cross a threshold (5% of packet count receive and transmit), warn the administrator about the network issue.
  - GLVM buffer size checks.
  - Security configuration (password rules).
  - Kernel parameters: Tunables that are related to AIX network, virtual memory manager (VMM), and security.
- Expanded support of resource-optimized failovers by using Enterprise Pools (ROHA).

- Browser-based GUI, which is called PowerHA SystemMirror User Interface (SMUI). The initial release is for monitoring and troubleshooting, not configuring clusters.
- All split and merge policies are now available to both standard and stretched clusters when using AIX 7.2.1.

4.2.3 PowerHA SystemMirror 7.2.2

Released in December 2017, PowerHA 7.2.2 added the following improvements:

- Log Analyzer provides capabilities for scanning and extracting detailed information about different types of errors from the PowerHA SystemMirror, AIX, and other system components log files.
- NovaLink supports the logical partitioning (LPAR) that is managed by PowerVM NovaLink.
- Easy Update, which is the **`cl_ezupdate`** tool.
- Shared listener support and added support for individual monitors of each of the Oracle listener threads.
- Oracle DB Shared Memory Clean Up. The shared memory that is associated with the Oracle database instance is cleaned up before starting the database.
- CAA autostart on the DR site. The CAA function stores the primary and backup repository disks' PVIDs and uses them to identify the repository disks during DR when UUID-based identification fails.
- Monitor Restart Count adds support for a Monitor Restart Count function for long-running Custom Application Monitors.
- Capturing CAA tunables in PowerHA Snapshot. Captures all the CAA tunables and customer security preferences as part of the snapshot database feature.
- The clRGinfo updates added the **`-i`** flag to show the status of applications with administrative control operations.
- Failover rehearsals (Enterprise version only).
- GLVM (Enterprise version only).

For more information about these improvements, see *IBM PowerHA SystemMirror V7.2 for IBM AIX Updates*, SG24-8278.

4.2.4 PowerHA SystemMirror 7.2.3

Released in December 2018, PowerHA 7.2.3 added the following improvements:

- SMUI new features
- Availability metrics
- Cloud Backup Management (CBM)
- Oracle database shutdown option
- Reliable Syslog facility (rsyslog) support
- LVM preferred read option
- Log analyzer improvements
- Support for stand-alone enqueue server 2

For more information about these improvements, see IBM PowerHA SystemMirror V7.2.3 for AIX offers new enhancements.

4.2.5 PowerHA SystemMirror 7.2.4

Released in December 2019, PowerHA 7.2.4 added the following improvements:

- The Cross-Cluster Verification (CCV) utility: With Version 7.2.4 and later, CCV allows cross-cluster verification and comparison.
- Cluster snapshot and **clmgr** command enhancements: The **clmgr** command supports all snapshot functions, and **clsnapshot** was deprecated.
- Availability metrics enhancements: These metrics were enhanced with extra reporting capabilities (**cl_availability**).
- Support for up to 256 RGs: Support was increased from 64 to 256 RGs.
- Db2 multiple database support: Configure and monitor multiple Db2 databases for the same Db2 server with SmartAssist.
- WebSphere MQ listener support: Multiple listeners for the same WebSphere MQ application are supported.
- Support for CAA 4 K disk block size for repository disks.
- Support for iSCSI disk as a cluster repository disk.
- Enhanced PowerHA SystemMirror RBAC to include a new condition.

The PowerHA SystemMirror GUI has the following updates:

- Enhanced visual representation of availability metrics and the newly available resource-centric views.
- Create and manage logical volumes, mirror pools, and RG dependencies.
- Generate the snapshot report and view the contents of a snapshot before you restore a snapshot.
- Enhanced cluster reports to display details about repository disks and methods.
- Non-root support for cloning cluster from snapshots.
- Configure the PowerHA SystemMirror GUI server to be HA by using the High Availability wizard option in the PowerHA SystemMirror GUI.
- Import multiple clusters by using the Add Multiple Clusters wizard.
- Enhanced activity log to display the activity ID, start time and end time of the activity, and the duration of the activity.
- Enhanced security features with options to disable anonymous login and global access.
- Automatic download and installation of the remaining files that are required to complete the PowerHA SystemMirror GUI installation process from the IBM website by using the **smuiinst.ksh** command.

4.2.6 PowerHA SystemMirror 7.2.5

Released in December 2020, PowerHA 7.2.5 added the following improvements:

- The EMC Symmetrix Remote Data Facility (SRDF)/Metro SmartDR configuration.
- PowerHA SystemMirror 7.2.5 SP1 added the EMC SRDF/Metro SmartDR configuration, which is a 2-region HADR framework that integrates SRDF/Metro and SRDF/Async replicated resources.
- GLVM Configuration Assistant enhancements: GLVM Configuration Assistant is enhanced with new features that convert an existing volume group to a GLVM-based volume group and updates an existing RG to include GLVM resources. Also, the delete or rollback feature that is used for removing resources and configurations that are performed by using the GLVM Configuration Assistant was improved.

The following updates are new in the PowerHA SystemMirror GUI:

- GLVM asynchronous configuration is now supported in the PowerHA SystemMirror GUI.
- The Operation Center Support (OCS) is used to configure the PowerHA SystemMirror GUI for long-term use, with visual and audio alerts.
- The CCV feature in PowerHA SystemMirror compares two cluster configurations to show differences in the PowerHA SystemMirror configuration, file sets, interim fixes, and more.
- All the supported application monitor configuration parameters are now available in the PowerHA SystemMirror GUI.
- The PowerHA SystemMirror GUI server's hostname is logged in the **clmgr** log file for cluster changes that are initiated from the PowerHA SystemMirror GUI.
- PowerHA SystemMirror can create a backup communication method for the PowerHA SystemMirror GUI server by configuring a Secure Shell (SSH) key while adding a cluster to the PowerHA SystemMirror GUI.
- The PowerHA SystemMirror GUI removed the usage of the **hostname** command to determine how to communicate with nodes. The PowerHA SystemMirror GUI server now collects either a public boot IP address or persistent IP address from each cluster node and uses that IP address to communicate with that node.
- Importing multiple clusters was enhanced to provide a progress indicator.
- Improved the clarity of events that are displayed in the PowerHA SystemMirror GUI by adding a start and complete indicator for two-phase events.

4.2.7 PowerHA SystemMirror 7.2.6

Released in December 2021, PowerHA 7.2.6 added the following improvements:

- Support for logical volume encryption.
- Starting with PowerHA SystemMirror 7.2.6 and IBM AIX 7.3, the LVM enables data encryption for data volume groups that are configured in the PowerHA SystemMirror environment.
- PowerHA SystemMirror 7.2.6 or later supports platform keystore (PKS) and key server authentication methods to enable the logical volume encryption. For more information about encrypting logical volumes, see Encrypting logical volumes.
- Support for EMC SRDF/Metro SmartDR configuration, which is a 2-region HADR framework that integrates SRDF/Metro and SRDF/Async replicated resources.

- Additional GLVM Configuration wizard enhancements:
  - Allows the dynamic update of the `aio_cache` logical volume.
  - The GLVM Configuration wizard collects the Remote Physical Volume (RPV) mirroring statistics and stores the information in JavaScript Object Notation (JSON) format. You can use the RPV statistics by using any tool that can display the JSON format. This data also is sent automatically to the PowerHA SystemMirror GUI, which displays it in a graphical format.
  - Additional GLVM policies, such as compression, io_grp_latency and no_parallel_ls.
- Cloud reliability, availability, and serviceability (RAS) enhancements:
  - For CBM, RAS was enhanced with an improved logging process.
  - Standard to linked cluster conversion.
  - In PowerHA SystemMirror 7.2.6, you can convert an existing standard cluster to a linked cluster by using the **`clmgr`** command. This feature is useful for converting a standard cluster to an IBM Power Virtual Server (Power VS) cloud cluster. For more information, see Converting a standard cluster to a linked cluster.

The following updates are new in the PowerHA SystemMirror GUI:

- GLVM historical charts: GLVM historical charts provide information about cache and network utilization.
- Asynchronous cache size in GLVM: Ability to view and modify the asynchronous cache size.
- GLVM policies: Ability to configure the physical volumes at the remote site and set the following tunables:
  - Compression
  - I/O group latency
  - Number of parallel logical volumes
- Multi-factor authentication: The GUI now uses multifactor authentication (MFA) through mobile or email authentication.

  Note: MFA is configured by using the IBM Security® Verify application.

- CBM: Ability to create, view, and delete backup profiles of RGs in IBM or Amazon cloud services.
- Multiple CCVs: Can compare primary cluster with multiple clusters with ability to select attributes to display.

4.3 New features in PowerHA SystemMirror 7.2.7

PowerHA SystemMirror Enterprise Edition became generally available on 6 December 2022. At the time of writing, it is supported on the following AIX levels:

- AIX 7.1 TL05 SP10
- AIX 7.2 TL01 SP6
- AIX 7.2 TL02 SP6
- AIX 7.2 TL03 SP7
- AIX 7.2 TL04 SP6

- AIX 7.2 TL05 SP5
- AIX 7.3 TL00 SP2
- AIX 7.3 TL01 SP1

The new features in the 7.2.7 release areas follow:

- ROHA in IBM PowerVS Cloud
- Cloud Backup and Restore (CBR)
- GLVM DR Sizing Tool
- CAA PVM WatchDog Timer support
- French Catalog message support
- Hardware Management Console (HMC) 10 support in ROHA
- Enhancement to **`cl_extendvg`**

4.3.1 Resource Optimized High Availability in IBM PowerVS Cloud

ROHA is a function in PowerHA SystemMirror that automatically and dynamically manages dynamic logical partitioning (DLPAR), Enterprise Pool Capacity on Demand (EPCoD), and On/Off CoD resources. You can configure ROHA with the HMC, hardware resource provisioning, and cluster tunable configurations.

ROHA is designed to assist customers manage the following costs:

- Hardware costs
- Software license costs

Before you use ROHA, you must determine which HMC manages a specific LPAR and any LPARs that you plan to use in the future. Plan for the necessary resources for your applications, identify your workloads, and your requirements for physical resources (CPU cores, virtual CPUs, and memory). After you identify all these requirements, you can configure ROHA.

Complete the following steps:

1. Create a definition for each HMC that manages any of the cluster nodes.
2. Determine the resource requirement for each application server. This resource is the extra resource that is added to the LPAR before the application server starts, and the same resources that are removed when the application server stops.

Because a PowerHA cluster can be created on IBM Cloud PowerVS, customers request the ROHA feature in IBM Cloud environment too. At the time of writing, PowerHA achieves this goal by interacting with the HMC or NovaLink either through SSH or the REST API.

In PowerVS, PowerHA uses a ServiceBroker API to perform any action on the PowerVS instance by using the REST API endpoint URL. To achieve this goal, all PowerVS instances must be defined and PowerHA invokes the REST API to the specific instance based on the application server requirements. PowerHA must be aware of all the PowerVS instances to invoke the REST API to the specific instance based on the application availability.

ROHA can be configured both through SMIT and the **`clmgr`** utility.

Dependencies

The following items are required to configure ROHA in PowerVS:

- Python Interpreter 2.7 or later
- REST API module `libcurl.a(libcurl.so.4)`
- Cloud authentication and connectivity from PowerVS cloud LPARs

Troubleshooting

In case of any issue with ROHA, the following logs might contain information that is relevant to debugging the problem:

- `/smit.log`
- `/var/hacmp/log/cloudroha.log`
- `/var/hacmp/log/hacmp.out` (Hint: Use the ROHALOG pattern search.)
- `/var/hacmp/log/clutils.log`

Note: With Version, 7.2.6.1, ROHA in the cloud is supported only with a Standard Cluster in PowerVS. Hybrid and multicloud clusters are not supported.

For more information about the Service Broker, see Get a Cloud Instance's current state/information.

The following videos on PowerVS are available Power Virtual Servers on IBM Cloud Series: Part 1:

- IBM Power Virtual Servers on IBM Cloud Series: Part 1
- IBM Power Virtual Servers on IBM Cloud Series: Part 2

4.3.2 Cloud Backup and Restore

IBM PowerHA SystemMirror 7.2.3 introduced a feature that is called CBM for IBM AIX, which enables users to create a backup of their critical data by using either of the following two methods:

- Cloud backup

  PowerHA uses the IBM SAN Volume Controller (SVC) FlashCopy function to create a backup of your data. Then, the created backup file is uploaded to the cloud by using cloud APIs. At the time of writing, PowerHA supports the IBM and AWS cloud providers.

- Remote storage

  Back up the data to remote storage by using the storage replication method.

To create a backup of application data, a backup profile must be created in PowerHA for the application's RGs configured, as shown in Figure 4-1 on page 81.

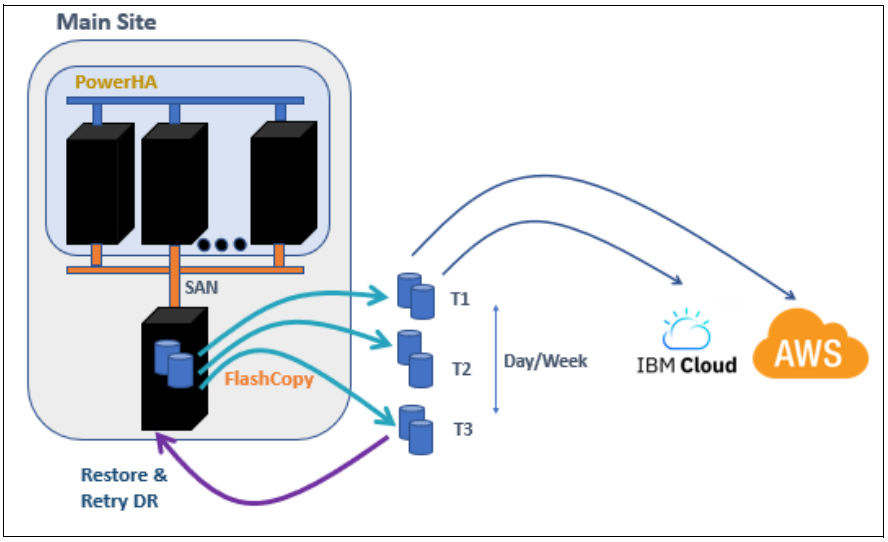

Figure 4-1 Backup management diagram

Backing up data by using backup management

If the backup method is `Cloud` and the associated RG is online, the backup process starts at the scheduled time, and an image file of the backup data is created at the target location that is set in the backup profile. Then, the image file is uploaded to the bucket in the cloud storage.

If the backup method is `remote_storage`, the data replication from the local SVC storage disk to the remote SVC storage disk starts and is synchronized with the primary storage.

Recovering data by using backup management

Backup management helps you to back up the application data at a set frequency and upload the backup to cloud storage. This feature ensures that the most recently updated application data is stored and can be recovered after a disaster.

Restoration is done by downloading the selected backup file from cloud and converting the file to a disk by running the **`dd`** command. Then, this disk is accessible by running **`importvg [disk#]`**.

Software and hardware requirements

Here are the requirements to create backup profiles in PowerHA:

- PowerHA SystemMirror 7.2.3 or later.
- Python 2.7 or later in all nodes of the cluster.
- The disks from SVC storage that is assigned to the nodes of the cluster.
- The PowerHA nodes must have cloud (IBM or AWS) connectivity before you configure the backup profile.
- The size of the source and target storage must be the same.

Changes in Version 7.2.7

DS8000 storage is now supported for addition, modification, and deletion with CBM. At the time of writing, the support is offered only through **clmgr** (support for SMIT will be implemented soon). The configuration options are similar to the SVC Storage configuration, except for the **PASSWORD** option, which is mandatory (Example 4-1).

Example 4-1 Adding DS8000 CBM

```
clmgr add storage_system <storage_system_name> \
             TYPE=ds8k \
             ADDRESSES=<IP>[<IP#2>,...] \
             USER=<username> \
             PASSWORD=<password> \
             BACKUP_PROFILE=true
```

Restrictions

The following restrictions apply to using DS8000 in CBM:

- At the time of writing, SMIT is not supported.
- The remote storage backup method is not supported.
- The only prerequisite is that **dscli** must be installed.

For more information, see *Cloud Backup Management with PowerHA SystemMirror*, REDP-5651.

4.3.3 GLVM DR sizing tool

The GLVM DR sizing tool analyzes the GLVM configuration and the I/O profile to estimate the network bandwidth and asynchronous cache requirements. The tool collects I/O data over a period to form the basis for the recommendations.

Operation

Complete the following steps

1. To run the tool, run the following command:

   ```
   /usr/es/sbin/cluster/cl_survey
   ```

2. Specify the data collection period by running the following command:

   ```
   -t <Collection time in seconds> -i<interval in seconds>.
   ```

Data is collected into files at both the local and remote site. Then, the data is analyzed to produce recommendations for network bandwidth and the AIO cache size. A report is produced and placed into `/var/hacmp/log/clsurvey/Clsurvey_Run_Analysis_information`.

Prerequisites

The following prerequisites must be met:

- Python Interpreter 2.7 or later.
- PowerHA 7.2.7 or later.
- A GLVM Asynchronous Volume Group is configured.
- Cluster Services are online.

Also use the following best practices:

- Do not make any cluster changes during the run.
- There can be only one run at a time and the report saved. Any additional runs overwrite the previous run and results.

For more information, see *Asynchronous Geographic Logical Volume Mirroring Best Practices for Cloud Deployment*, REDP-5665.

4.3.4 CAA PowerVM WatchDog Timer support

With this feature, a PowerHA administrator can set the CAA tunable **`pvm_wdog`**. CAA supports a new tunable PowerVM Watchdog Timer in IBM AIX 7.3 Technology Level 1 or later. The PowerVM Watchdog Timer tunable is supported in PowerHA SystemMirror 7.2.7 or later. When you enable this tunable, CAA resets the Power hypervisor (PHYP) timer of every cluster node. If the reset of the PHYP timer fails because, for example, the LPAR fails, PHYP manages that LPAR based on the option that is set by the user.

The following values can be configured:

- **`DISABLE`** (the default)
- **`DUMP_RESTART`**
- **`HARD_RESET`**
- **`HARD_POWER_OFF`**

This variable can be set and viewed through SMIT by running **`smitty sysmirror`**, and then selecting **Custom Cluster Configuration** → **Cluster Nodes and Networks** → **Manage the Cluster** → **Cluster heartbeat settings**.

4.3.5 French Catalog message support

Before Version 7.2.7, PowerHA supported English and Japanese messages. With Version 7.2.7, French message catalogs were added. The following file sets are included with PowerHA:

- `cluster.msg.fr_FR.*`
- `cluster.msg.Fr_FR.*`

Prerequisites

Install the AIX French file sets and set the `LANG` variable to "`fr_FR`". Then, installing PowerHA installs the French file sets.

4.3.6 Hardware Management Console 10 support in ROHA

HMC 10 introduced several changes when compared to earlier versions. These changes include the format for version and release information and several return codes. PowerHA ROHA was updated to support HMC 10, with both SSH and REST API connectivity supported. REST API now handles return codes.

These changes also will be available in the following upcoming service pack releases for PowerHA:

- 7.2.4 SP6
- 7.2.5 SP4
- 7.2.6 SP2

4.3.7 Enhancement to cl_extendvg

An enhancement was requested for a particular issue that relates to adding a disk to a volume group through C-SPOC when the disk is part of another volume group.

Previously, if a disk was shared between cluster nodes and one or more non-cluster nodes and was added to an existing cluster volume group, no error was returned, which led to potential corruption or loss of data.

As a solution to this issue, a check is added in PowerHA to make sure that the disk that you use to extend the volume group does not have any VGDA data remaining. If any data is found, an error is returned that states that VGDA data was found, which must be cleaned before trying again.

4.4 New features in PowerHA GUI 7.2.7

Here are the new features in the PowerHA 7.2.7 GUI:

- GVLM sizing tool
- Email notification
- Cloud Backup Management with DS8000 storage
- Cloud Backup Management restoration
- GLVM Tunables in Multi-CCV
- Automatic Expiration of Events
- Fix Central

4.4.1 GVLM sizing tool

The GLVM sizing tool provides suitable recommendations that are based on the network bandwidth and cache requirements for the asynchronous GLVM network that supports GLVM traffic. The GLVM sizing tool monitors disks and network usage for the specified duration and provides suggestions.

4.4.2 Email notification

The user can now select to receive an email notification from the PowerHA GUI based on selected events for particular clusters.

An email notification about a cluster-based event is sent to the email ID that is set by the user. In PowerHA SystemMirror GUI, a user can modify and delete the email notification configuration that is set. A user also can delete the email notification configuration. To use the Send Email Notification feature, you must use the Simple Mail Transport Protocol (SMTP) configuration on the server node.

4.4.3 Cloud Backup Management with DS8000 storage

With the CBM feature, users can create, view, edit, and delete backup volume group profiles of IBM DS8000 storage disks in IBM Cloud. You can back up the volume group data and store it on IBM or Amazon cloud services.

4.4.4 Cloud Backup Management restoration

With CBM Restoration, users can restore backup profiles of an RG from IBM Cloud or Amazon Web Services (AWS) cloud. You can back up the volume group data and store it on IBM or Amazon cloud services. You can restore the backed up data by using the DR operation to prevent data loss or data corruption.

4.4.5 GLVM Tunables in Multi-CCV

The GLVM Tunables in Multi-CCV feature provides GLVM policy information for cross-cluster comparison.

4.4.6 Automatic Expiration of Events

The Automatic Expiration of Events feature closes all notifications in the PowerHA SystemMirror GUI dashboard page after 1 minute (the default time).

4.4.7 Fix Central

This feature integrates the Fix Central support with the PowerHA GUI so that a user can check and upgrade their PowerHA SystemMirror cluster with the required service pack.

With the GUI, users see what PowerHA service pack versions are available (Figure 4-2), and then download the selected Fix Pack to any or all of nodes in the cluster. You are prompted to provide your IBM credentials before the download will start. A warning is issued if the selected service pack already is downloaded.

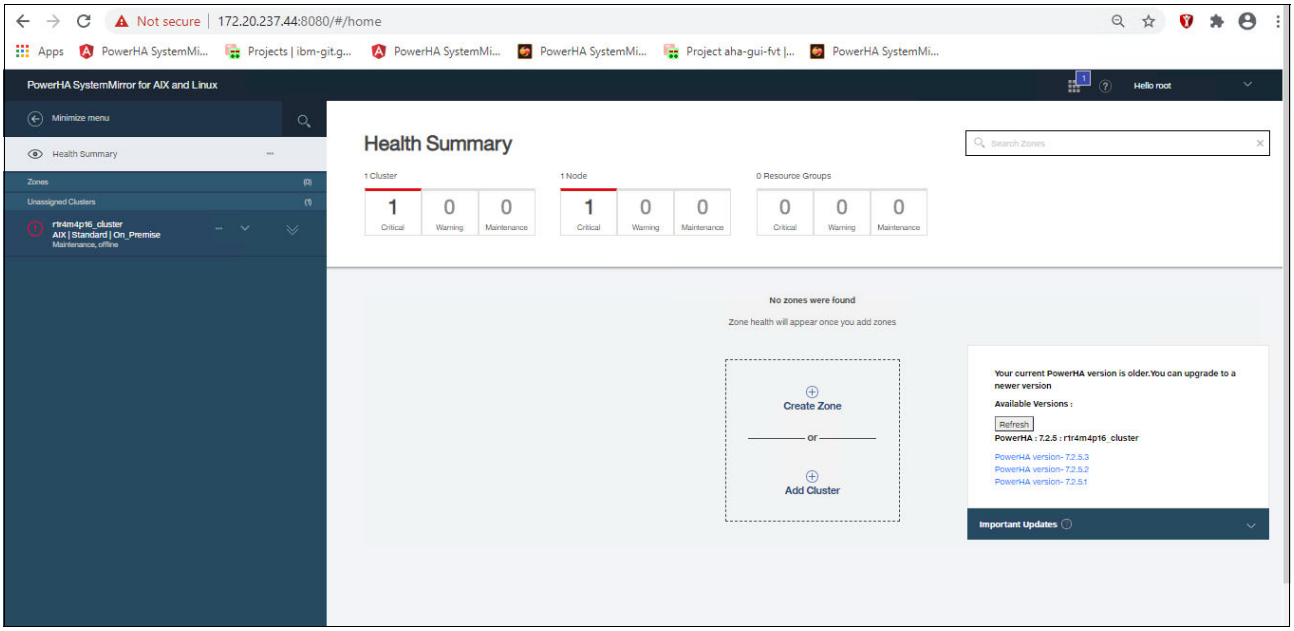

Figure 4-2 PowerHA GUI showing fixes available through Fix Central

4.5 Installation prerequisites

A primary requirement for providing a HA environment is to identify and remove all single points of failure (SPOFs) from the environment. This task requires redundancy options for servers, networks, storage, data centers, and the surrounding infrastructure (people, printers, backups, and so on).

In this section, we examine the following items:

- Basic system requirements
- Network configuration
- Storage configurations
- Site requirements
- CAA for PowerHA SystemMirror
- Other prerequisites

Generally, applications can be broken down into two types:

- Scale-out or concurrent
- Clustered

Scale-out and concurrent solutions provide redundancy by using multiple instances, so the focus is on ensuring that the surrounding infrastructure provides client access to several application instances while ensuring that sufficient instances of the application are available to meet workload requirements.

Clustered solutions rely heavily on knowing the status of the infrastructure to keep the individual applications available. The focus is on ensuring that the applications are online only when required while ensuring that they have consistent access to the data. If the cluster splits, then nodes on either side should not start to operate independently.

4.5.1 Basic system requirements

There are many different ways to build a HA environment. This section describes a subset of options.

Mirrored architecture

In a mirrored architecture, you have identical or nearly identical physical components in each part of the data center. You can have this type of setup in a single room (although it is not recommended), in different rooms in the same building, or in different buildings.

Figure 4-3 on page 87 shows a high-level diagram of a typical cluster. In this example, there are two networks, two managed systems, two Virtual I/O Servers (VIOSs) per managed system, and two storage subsystems. This example also uses LVM mirroring for maintaining a complete copy of data within each storage subsystem. There also is a disk for the CAA repository disk on each storage subsystem. For more information about how to set up the CAA repository disk, see Cluster Repository Disks.

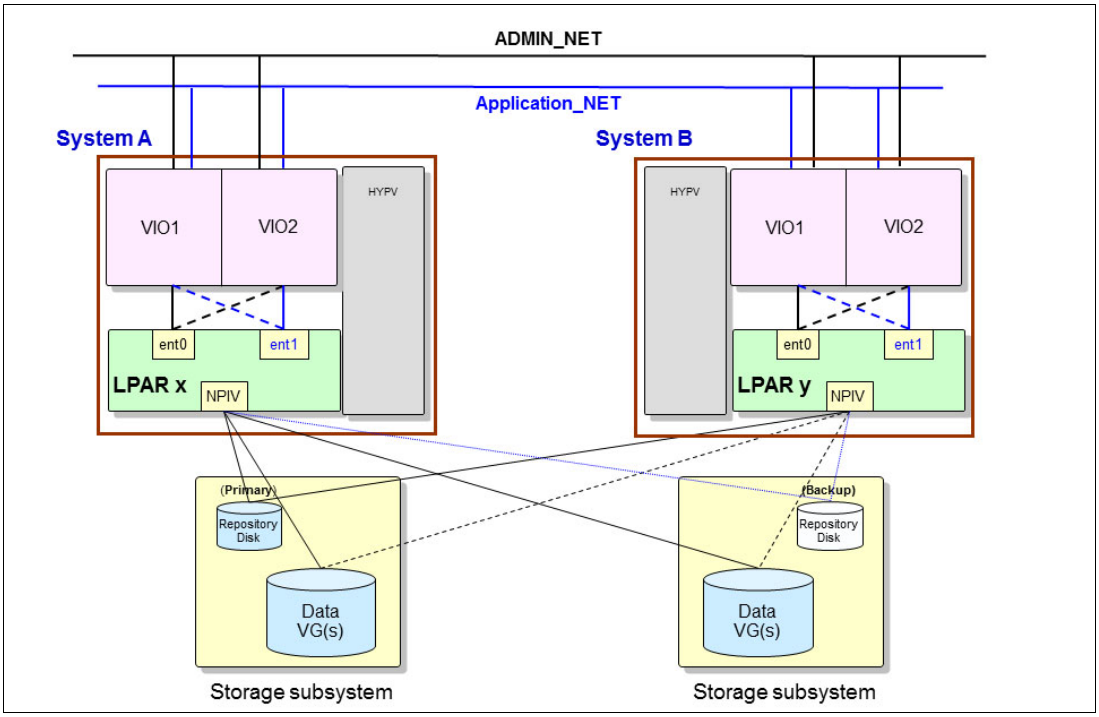

Figure 4-3 Cluster with multiple storage subsystems

4.5.2 Network configuration

This section focuses on the network considerations from an availability point of view and examines the following items:

- Types of network (physical or virtual)
- Network adapters
- Redundancy in networks
- Intersite considerations

Physical or virtual network

Using technologies such as LPM, Simplified Remote Restart (SRR), and VM Recovery Manager (VMRM) require that the environment be fully virtualized. In clustered solutions such as PowerHA SystemMirror, although there are some configuration differences, they operate equally well in both physical or virtual environments.

Network adapters

Network redundancy traditionally is provided by using dual-adapter networks. More recently, single logical adapters are used with their redundancy that is provided by multiple physical backing devices. This approach uses bonding (Etherchannel), failover (Network Interface Backup), virtualization (dual VIOSs and Shared Ethernet Adapters (SEAs)), or a combination.

Redundancy in networks

In the past, redundant networks were common. Now, these networks are not as prevalent because improvements in the design and operations of the network hardware with the ability to have multiple paths introduced greater redundancy.

Intersite considerations

Ensure that the intersite connection does not become a SPOF. To accomplish this task, avoid the following items:

- A single provider
- A common entry point for client access to the applications at both sites
- A common entry or exit point for the intersite links
- Common intermediate points

Often, different data centers use different subnets. and although they can be handled by PowerHA SystemMirror and VMRM DR, manual intervention might be required if other HA solutions are operated across sites.

This publication describes the importance of planning the network bandwidth and latency to meet the application response time requirements. What is equally important is to plan a bandwidth that is sufficient for both normal operations and the extra throughput that is required to recover and resynchronize a site after a disaster.

4.6 Storage configurations

This section describes different storage configurations.

4.6.1 Single storage architecture

In a *single storage architecture*, the common storage subsystem is shared by all the nodes. This solution can be used when there are lower availability requirements for the data, and it is not an uncommon architecture when all nodes are in the same location.

If you use storage-based mirroring and replication, such as IBM SVC, the physical layout is similar to the mirrored architecture that is described in “Mirrored architecture” on page 86. However, from the OS perspective, it is a single storage architecture because it is aware of only a single set of logical unit numbers (LUNs). However, from the cluster management perspective, this architecture requires some extra administration to manage the underlying replication. For more information about the layout in an IBM SVC stretched cluster, see “Stretched cluster” on page 89.

Figure 4-4 shows such a layout from a logical point of view.

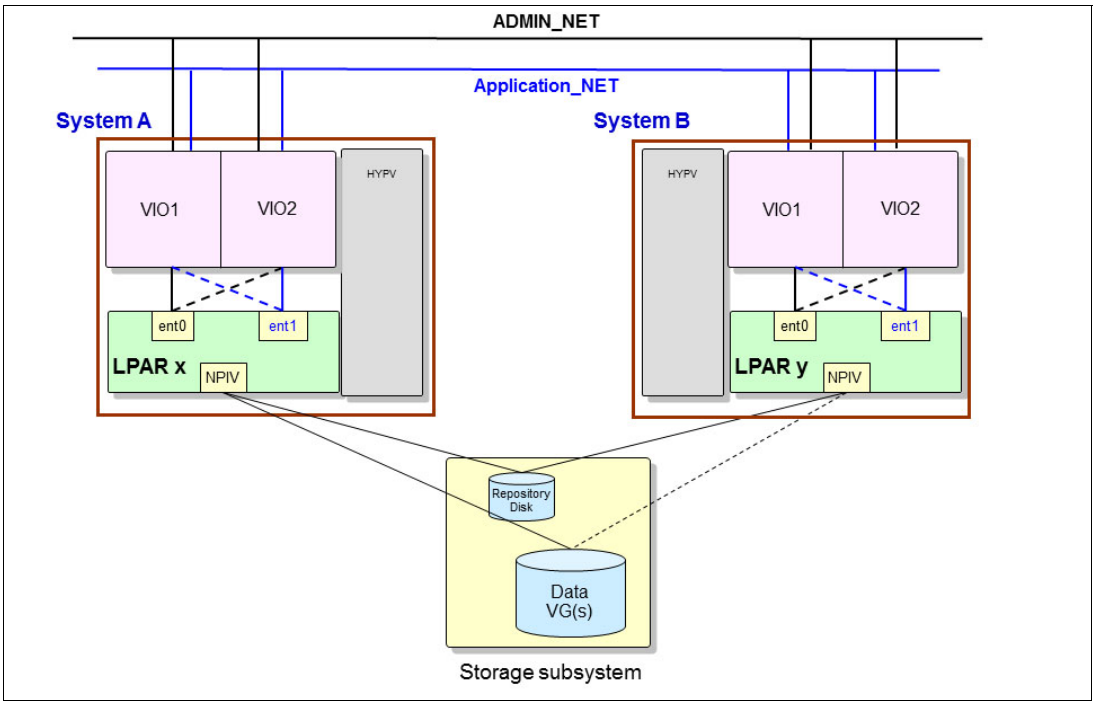

Figure 4-4 Cluster with a single storage subsystem

Stretched cluster

A *stretched cluster* involves separating the cluster nodes into *sites*. A site can be in a different building within a campus or separated by typically less than 120 kilometers. In this configuration, the storage area network (SAN) spans the sites, and storage can be presented across sites.

Having both SAN and TCP/IP connectivity between sites removes the site network as a SPOF. Steps must still be taken to ensure that both different providers and different routes are used so that there is not a common point that can be broken, preferably for both SAN and IP networks.

Another main concern is having redundant storage and verifying that the data within the storage devices is synchronized across sites. The following section presents a method for synchronizing the shared data.

Storage subsystem that uses a stretched configuration

The SAN storage subsystems can be configured in a *stretched* configuration. In the stretched configuration, the storage controller presents the two storage devices as one unit even though they are separated by distance. The storage subsystem keeps the data between the sites consistent.

With a storage subsystem in a stretched configuration, the cluster software can provide continuous availability of the storage LUNs, even through the failure of a single component. With this combination, the behavior of the cluster is similar in terms of function and failure scenarios in a local cluster (Figure 4-5).

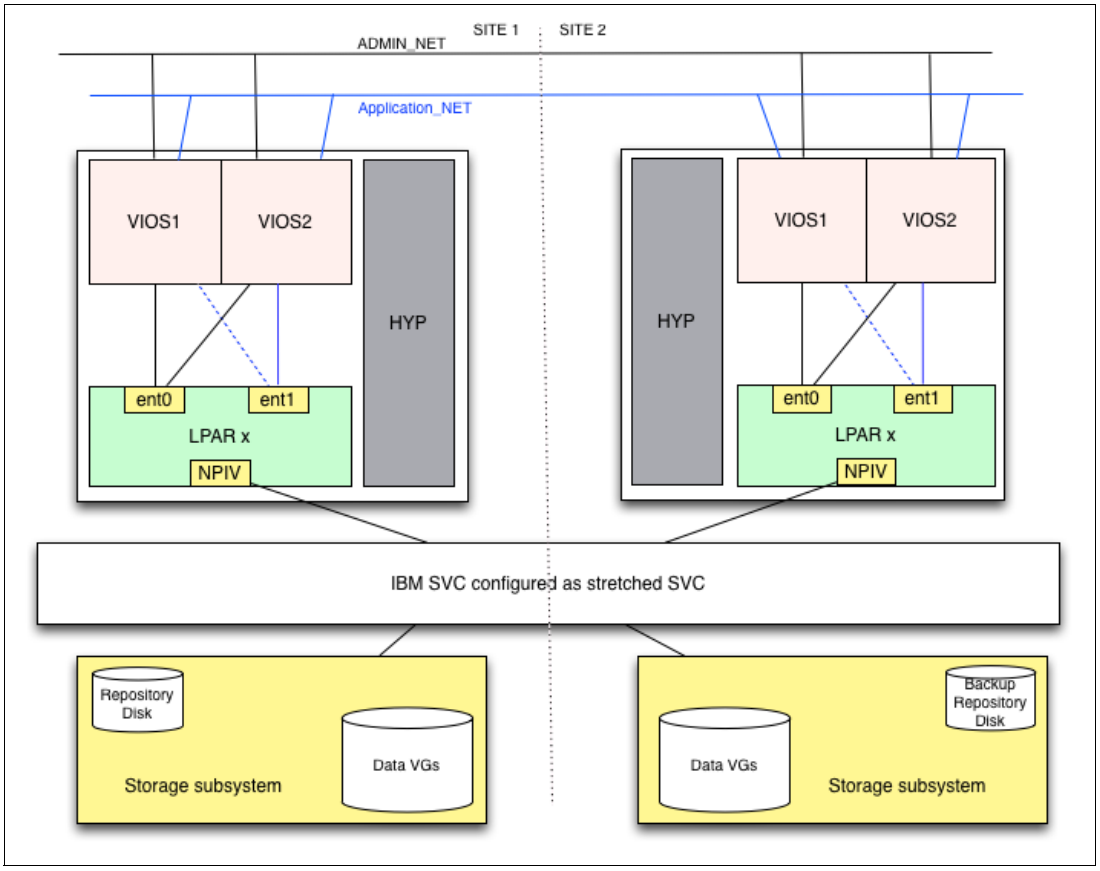

Figure 4-5 IBM SAN Volume Controller stretched configuration

Linked cluster

A *linked cluster* is another type of cluster that involves multiple sites. In this case, there is no SAN link between sites due to a combination of cost and distance.

In this configuration, each site has its own copy of the repository disk, and PowerHA SystemMirror keeps those disks synchronized.

Because there is only one type of intersite network, the IP network is a SPOF, so we must reduce the possibility of it failing. Ensure that there are multiple providers and routes so that there is no loss of IP communication between the sites.

For more information about linked clusters, see *IBM PowerHA SystemMirror 7.1.2 Enterprise Edition for AIX*, SG24-8106.

IBM supported storage that uses copy services

Although there are several IBM supported storage devices with copy services capabilities, we use IBM SVC for the following example. IBM SVC can replicate data across long distances by using the IBM SVC copy services functions. The data can be replicated in either synchronous or asynchronous modes.

If there is a failure that requires moving the workload to the remaining site, the cluster software interacts directly with the storage to switch the direction of the replication. The LUNs are presented to nodes at the surviving site and the clustering software activates the applications to grant access to users by using the addresses for that site.

An example of this concept is shown in Figure 4-6.

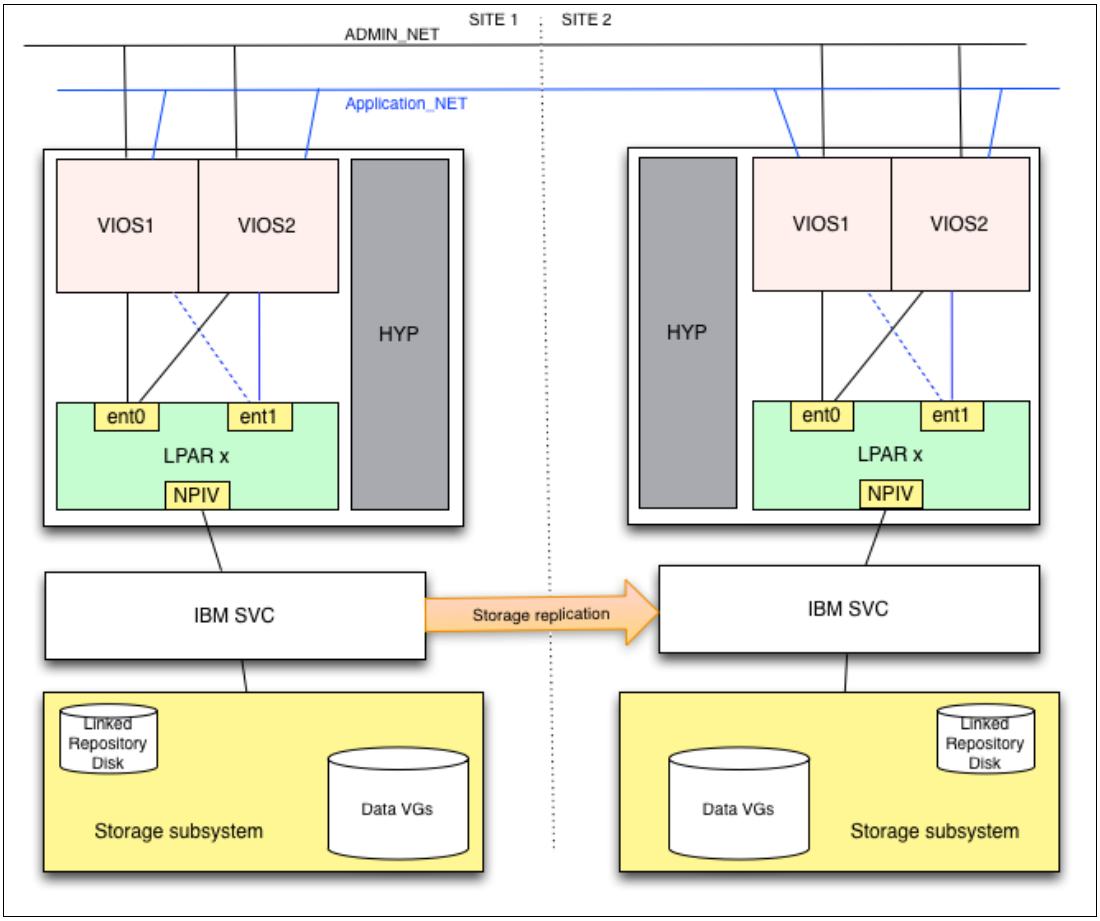

Figure 4-6 PowerHA SystemMirror and IBM SAN Volume Controller storage replication

4.7 Additional PowerHA resources

Here is a list of additional PowerHA resources:

- Entitled Software Support (download images)
- PowerHA fixes
- PowerHA, CAA, and RSCT migration interim fixes
- PowerHA wiki

  This comprehensive resource contains links to all the following references and much more.
- PowerHA LinkedIn group

- Base publications:

  All the following PowerHA 7 publications are available at IBM Documentation:

  - *Administering PowerHA SystemMirror*
  - *Developing Smart Assist applications for PowerHA SystemMirror*
  - *Geographic Logical Volume Manager for PowerHA SystemMirror Enterprise Edition*
  - *Installing PowerHA SystemMirror*
  - *Planning PowerHA SystemMirror*
  - *PowerHA SystemMirror concepts*
  - *PowerHA SystemMirror for IBM Systems Director*
  - *Programming client applications for PowerHA SystemMirror*
  - *Quick reference: clmgr command*
  - *Smart Assists for PowerHA SystemMirror*
  - *Storage-based high availability and disaster recovery for PowerHA SystemMirror Enterprise Edition*
  - *Troubleshooting PowerHA SystemMirror*
- PowerHA and Capacity Backup
- Videos

  Shawn Bodily has several PowerHA related videos on his YouTube channel.
- Developer Discussion Forum
- IBM Redbooks publications

  The main focus of each IBM PowerHA Redbooks publication differs a bit, but usually their main focus is covering what is new in a release. They generally have more details and advanced tips than the base publications.

  Each new publication is rarely a complete replacement for the previous one. The closest exception is *IBM PowerHA SystemMirror for AIX Cookbook*, SG24-7739-02. At the time of writing, it is being updated to Version 7.2.7 in parallel with this paper, and it replaces three previous cookbooks. It is probably the most comprehensive of all the IBM Redbooks publications regarding PowerHA Standard Edition specifically. Although there is some overlap across them with multiple versions supported, reference the version of the book that is relevant to the version that you are using.

  - *Deploying PowerHA Solution with AIX HyperSwap*, REDP-49544
  - *Guide to IBM PowerHA SystemMirror for AIX Version 7.1.3*, SG24-81677
  - *High Availability for Oracle Database with IBM PowerHA SystemMirror and IBM Spectrum Virtualize HyperSwap*, REDP-5459
  - *IBM PowerHA SystemMirror 7.1.2 Enterprise Edition for AIX, SG24-8106*
  - *IBM PowerHA SystemMirror for AIX Cookbook*, SG24-7739
  - *IBM PowerHA SystemMirror for AIX 7.1.3 Best Practices and Migration Guide*, SG24-82344
  - *IBM PowerHA SystemMirror Standard Edition 7.1.1 for AIX Update*, SG24-8030
  - *IBM PowerHA SystemMirror V7.2 for IBM AIX Updates*, SG24-8278
  - *IBM PowerHA SystemMirror V7.2.1 for IBM AIX Updates*, SG24-8372

- *IBM PowerHA SystemMirror V7.2.3 for IBM AIX and V7.22 for Linux*, SG24-8434
- *IBM SAN Volume Controller Stretched Cluster with PowerVM and PowerHA*, SG24-81422

5

IBM PowerHA SystemMirror for IBM i

This chapter provides an overview of IBM PowerHA SystemMirror for IBM i. IBM i integrated support for clustering and high availability (HA) to help you build a resilient and highly available (HA) platform for your business applications.

This chapter describes the following topics:

- Terminology and concepts
- History and evolution
- New features
- Installation, prerequisites, and options
- Data resilience
- High availability management
- Backup solutions in an HA environment
- Third-party options for IBM i HA: logical replication
- Migrating and upgrading from previous releases

5.1 Terminology and concepts

This section provides an overview of the terminology and concepts that you must understand when considering the high availability and disaster recovery (HADR) options that are available on the IBM Power platform for IBM i.

5.1.1 Cluster

A *cluster* in the IBM i environment refers to a group of IBM i systems or logical partitions (LPARs) that work together like a single system entity to provide resilient resources for business continuity, which provides HA to applications, data, and services.

A cluster must be flexible and easily scalable and provide the fastest recovery for as wide range of outages as possible with minimal cost and effort.

5.1.2 Cluster node

A *cluster node* refers to an IBM i system or LPAR that is a member of a cluster. A cluster is 2 - 128 nodes. Each node is associated with one or two IP addresses that represent the system. Communications between the nodes of the cluster run over TCP/IP by using those IP addresses.

When the node is added to the cluster, a name is given to the node that identifies that node as one of the members of the cluster. When configuring a cluster, you can use any name that you want for a node in the cluster. However, as a best practice, the node name should be the same as the hostname or the system name. The set of cluster nodes that are configured as part of the cluster is referred to as the *cluster membership list*.

There are three main roles that a cluster node can adopt, depending on the cluster configuration:

Primary Node / Active Node Contains the principal copy of the replicated resources (for example, the production independent auxiliary storage pool (IASP)), and it is the owner of any device resource. It can fail over to the backup node when necessary to maintain the availability of the system.

Backup node / Passive Node Can take over the role of primary access when a failure occurs on the primary node. Contains a valid replicated-copy of the cluster resource.

Data Node / Replicate Node Contains copies of cluster resources, but cannot assume the role of the primary or backup. It is used for data warehousing or backup processes of the primary node through FlashCopy, for example.

5.1.3 Cluster resource group

A *cluster resource group* (CRG) is an IBM i system object that is a set or grouping of cluster resources that is used to manage events that occur in a HA environment. A CRG is the foundation for all types of resilience.

Cluster resources can be moved or replicated to one or more nodes within the cluster, and they consist of the resources that are required to be HA for the business.

The CRG manages and monitors the collection of cluster resources and can define the relationship between nodes that are associated to them. For example, you can specify which nodes can contain the resources, which node is assigned those resources, and which node should get the resources in a failure.

The IBM i cluster defines four types of CRG:

Device CRG Supports device resiliency in IBM i HA environments. A device CRG contains one or a list of IASPs, which are replicated or switched between systems. IASPs can be switched or replicated only between nodes in the recovery domain that is defined for that device CRG. Access to all the IASPs that are listed in the device CRG is switched to the backup node when an outage, planned or unplanned, occurs.

Application CRG Can start an application and monitor for application failure. An application is defined as a program or set of programs that can be called to deliver some business solution. The application CRG does not manage any data that is associated with the application. The data is managed by a data or device CRG.

Data CRG An IBM i system object that helps data replication between the primary and backup nodes in the recovery domain. A data CRG does not do the replication, but 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.

Data CRGs are primarily used with logical replication applications, which are provided by several HA IBM Business Partners.

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

Unlike the other CRG types, in which only the primary CRG node is the node doing the work, in a peer CRG all the nodes in the recovery domain work together. The purpose of peer CRG is to provide a general distributed-computing framework for which programmers can write applications. The peer CRG provides peer resiliency for groups of objects or services. The user, user applications, and IBM Business Partner applications, not the system, choose the groups of objects.

Within these types of CRG, you can find two common elements: a recovery domain and an exit point.

The *recovery domain* is the subset of nodes that can be used in recovering the resources in th CRG. The *exit point* defines the actions that are performed when certain events are detected by the CRG, such as a primary node failing.

For example, a device CRG enables a hardware resource to be switched between nodes. This CRG is represented by a device configuration object that is called an IASP.

5.1.4 Administrative domain

The cluster *administrative domain* provides a mechanism for maintaining operational environment consistency across all the cluster nodes within an HA environment. This approach ensures the expected behavior of the different CRGs when they are switched to or fail over to the backup nodes.

When a change is made in one of the nodes, the same change must be made in all the other nodes. With the administrative domain, you can identify and monitor for changes and maintain consistency across the nodes, by synchronizing any changes.

For example, it is important that the user profiles on the primary node and backup node have the same parameters, such as job description (JOBD), library lists, UID, and GID.

5.1.5 Device domain

When a switchable IASP is created, a *device domain* is the first of the cluster constructs to be defined, which ensures that there are no configuration conflicts that can prevent a switchover or failover from occurring.

The set of configuration resources that is 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 a time, that device can be switched to another node and brought online.

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

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

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

Virtual address assignments The cluster configuration determines the virtual address space that is required for the IASP. Virtual address assignments (the cluster configuration) cannot conflict across any of the nodes in the device domain.

5.1.6 ASP Copy Descriptions

ASP Copy Descriptions provide the information that is used in an ASP Session. Each copy description contains the information that is needed by PowerHA SystemMirror for IBM i to manage and control a specific copy of an IASP in a HA solution. Two types of copy Descriptions are used by PowerHA:

ASP Copy Descriptions Establishes the connectivity between an IBM i system with the ASP Copy Description.

IBM SAN Volume Controller Copy Descriptions Establishes the connectivity between an IBM i system and the IBM SAN Volume Controller (SVC).

5.1.7 ASP sessions

An *ASP session* is used to link ASP Copy Descriptions to create a data replication session. The status of the ASP session describes the status of the replication.

ASP sessions define the type of data replication to be used by PowerHA SystemMirror for IBM i. The choices are Geographic Mirror, Metro Mirror, Global Mirror, or FlashCopy. An ASP session is not required when you are using logical unit number (LUN) level switching. The ASP session is used to manage and monitor the status of data replication.

5.1.8 Data replication

There are multiple PowerHA SystemMirror for IBM i technologies that are available to switch or replicate the data between IASPs.

Switchable logical units

The *switchable logical unit* (LU) technology can switch the access of a copy of the IASP from the primary to the backup server.

Replication technologies

For *replication technologies* such as geographic mirroring, Metro Mirror, or Global Mirror, data is mirrored from the IASP at the production site to another IASP at the backup site. Replication technologies can be divided into two categories:

- Host-based replication

  Also called Geographic Mirroring, host mirroring uses data copy over the Internet Protocol network, which is used to keep a consistent copy of an IASP in one of the cluster nodes (production copy) in another IASP that is assigned on another node (target copy).

- Storage-based replication

  In storage-based replication, the data replication is done by the storage subsystem by using one of the following technologies:

  - Metro Mirror
  - Global Mirror
  - FlashCopy

Usually, these sites are geographically separate from each other, which provides disaster recovery (DR) protection. In these environments, the device CRG controls switching access between mirrored copies of the IASP. When the production site experiences an outage event, the device CRG switches access to the IASP to the backup site.

5.1.9 Cluster jobs

When managing an IBM i cluster, you must know the job structures that are created on the system and how they are organized. These jobs are multithreaded and run under a QSYS profile. Critical cluster resource jobs are system jobs. When the cluster is active on a system, you can find the following system jobs:

QCSTCTL	The cluster control job.
QCSTCRGM	The CRG job manager.

CRG jobs	You can find a job per CRG object with the same name as the CRG name.
Administrative domain jobs	There is a single system job with the same name as the administrative domain name of the cluster that is running on each node of the cluster.

QCSTCTL and QCSTCRGM are critical jobs for the cluster, so they must be running for the node to be active in the cluster. If any of these jobs end, clustering ends, and a failover occurs.

5.1.10 Cluster events

This section details the different cluster events that can occur within the PowerHA SystemMirror for IBM i environment.

Switchover

Switchover is the manual switch of a cluster resource, for example, PowerHA SystemMirror for IBM i, from one node to another one. The most common usage for the switchover is to do system maintenance on a node, such as PTF installations, system upgrades, or to install a new OS release.

When the switchover occurs, the access to the resource is switched from active node to the first node that is assigned as backup node in the recovery domain of the CRG.

To perform an ordered switchover of a device CRG, complete the following steps:

1. Quiesce the application on the primary node to quiesce changes to the data.
2. Switch the device CRG to the new primary node.
3. Start the applications or jobs in the new primary node.

Failover

Failover is the automatic switch of a cluster resource from one node to another node in a system failure.

A failover event works in the same way as a switchover, but the difference is how the event is triggered. The access to the resource is switched from the active node to the first node that is assigned as the backup node in the recovery domain of the CRG.

If multiple CRGs are involved in the failover event, device CRGs are processed first, then data CRGs, and then the application CRGs.

The failover processing of device CRGs varies off the devices that are associated with the CRG, which means that the devices are varied off even if the failover event is canceled.

Cluster partition

Within IBM i HA environments, a *cluster partition* is a subset of the active cluster nodes that results from a communications failure. Members of a partition maintain connectivity with each other.

A cluster partition occurs in a cluster when the communication between one or more nodes is lost and the failure of the lost nodes cannot be confirmed. When this condition is detected, the cluster services limit the types of actions that can be performed by the nodes in the cluster partition. When the problem is solved, cluster resource services can merge the nodes in partition state.

5.1.11 PowerHA policies

PowerHA policies are designed to modify the behavior of the HA environment. These policies establish the actions and capabilities that PowerHA uses when processing instructions and provide a single point of control for the administration of changes to operations.

There are different types of PowerHA policies based on the modified behavior:

- Administrative domain policies

  These policies permit adding, deleting, and restoring resources to nodes in the cluster administrative domain:

  - QCST_AD_CREATE: With this policy, you can specify which resources automatically are added to the administrative domain when that resource is created.
  - QCST_AD_DELETE: Automates and simplifies management of the removal of monitored resources and monitored resource entries (MREs) from a cluster.
  - QCST_AD_RESTORE: When restoring a resource that is monitored by the administrative domain, the restoration of attributes is kept. Normally, resources with an MRE in the cluster administrative domain that are restored are immediately updated to match the attribute values in the cluster administrative domain, which ignores the restore operation.
  - QHA_AD_ANZCADPRF: This policy is used with the **ANZCADPRF** command so that you can act on detected inactive user profiles by using distributed information across the administrative domain.

- Cluster resource group

  QCST_CRG_CANCEL_FAILOVER: Used to prevent some failovers from taking place.

- Communication

  QHA_COMM_STRICT_CERT_CHECK: Controls security behavior in the HA network and the configuration settings that govern communication between PowerHA and the storage device by using IBM Copy Services Manager.

5.2 History and evolution

In IBM i V6R1, 5761-HAS shipped with the name IBM System i High Availability Solutions Manager. Shortly after it shipped, the name was changed in all marketing brochures to IBM PowerHA for i. The product ID continued as 5761-HAS, and it was PowerHA 1.0.

In IBM i V7R1, 5770-HAS shipped with the name IBM PowerHA for i. At announcement time, the name was changed in all marketing brochures to PowerHA SystemMirror for IBM i. The product ID continued as 5770-HAS, and it was PowerHA Versions 2.0 - 2.2.

In IBM i versions V7R2, V7R3 kept the same product ID and name. These OS versions both supported PowerHA Versions 3.0 - 3.9 starting in 2014 through June of 2022. PowerHA was offered as three versions: Express, Enterprise, and Standard Editions.

In IBM i V7R4, 5770-HAS again was used, and still with the same three versions. These versions were PowerHA 4.0 - 4.7.2, June 2019 - June 2022.

In IBM i V7R5, PowerHA SystemMirror for IBM i was simplified to a single edition that contains all the functions. It uses product ID 5770-HAS.

Section 5.3, “New features” on page 102 describes the new features and enhancements that were introduced since V7R3.

At the time of writing, the versions for PowerHA are the following ones:

- IBM i V7R3: PowerHA 3.9
- IBM i V7R4: PowerHA 4.7.2
- IBM i V7R5: PowerHA 5.1.2

Tip: For more information about PowerHA SystemMirror for IBM i, see Welcome to PowerHA SystemMirror for IBM i.

5.3 New features

The new features and enhancements of PowerHA SystemMirror for IBM i have been numerous during different Technology Release (TR) launches since V7R3. You find the most important ones in the following list:

- V7R3 OS Base Enhancements: IASP HyperSwap Cluster
- V7R3 TR1:
  - Support for IBM SVC and IBM Storwize HyperSwap.
  - New Hardware Management Console (HMC) interfaces for Advanced Cluster Node Failure Detection.
- V7R3TR6 - V7R4 Base Enhancements:
  - Support for DS8000 HyperSwap with a Global Mirror link.
  - Automation of administrative domain operations.
- V7R3 TR8 - V7R4 TR2:
  - Introduction of a new policy (QHA_AD_ANZCADPRF) and the command **ANZCADPRF** to analyze and cache usage information across the PowerHA for IBM i environment to disable inactive user profiles.
  - Real-time recovery point objective (RPO) information that shows how much data will be lost in a disaster for SVC based replication and for DS8000 based replication (CSM).
  - PowerHA for IBM i SQL Services to retrieve PowerHA information about cluster and cluster node information and status, administrative domain and monitored resource information and status, and CRG information and status.
  - PowerHA for IBM i with this TR automatically updates the PowerHA version when the upgrade or PTF is applied to all nodes within the environment. The **CHGCLUVER** command is not required for enabling the new enhancements.
  - There was a 132-column display redesign for a cleaner vision of the replication status.
  - Simplified configuration and management of PowerHA for IBM i to provide contextual F4 prompt options to some commands, which now are optimized to the specific configuration and environment.

- V7R3 TR9 - V7R4 TR3:
  - Clustering enhancement: Retries cluster node activation after IPL. Allows an automatic retry to start the cluster node after waiting for a period for the IP addresses to become fully active so that the start is successful (specific in Group SI99876 Level 13).
  - Simplified management and deployment of clusters with automation and default settings that can be customized. The manual steps that are required to configure a PowerHA environment were substantially reduced.
  - Updated PowerHA dashboards.
  - PowerHA for IBM i SQL Services: Introduced a query to extract information about the state of data replication.
  - RPO real-time information is now displayed on the geographic mirroring session window, along with an estimate of the impact of the network on the performance of geographic mirroring.
  - Easier to configure firewall security and QoS in a geographic mirroring environment. With this TR, a port number for geographic mirroring to use can be specified.
- V7R3 TR10 - V7R4 TR4:
  - Streamlined switchover: Provides simplified node selection with automated recovery domain management.
  - FlashCopy automation to simplify FlashCopy operations.
- V7R3 TR6:
  - Geographic Mirroring compression: Geographic mirroring performs disk-level replication. When PowerHA performs a synchronization, whether for the initial synchronization, after an unplanned failure, or when resuming from a suspend operation, PowerHA replicates data over the Internet Protocol network.

    With this enhancement, PowerHA can now compress geographic mirroring data during resynchronization, which reduces the data transfer that is required. In addition, on Power10 servers running in Power10 compatibility mode, PowerHA uses the on-chip NX GZIP accelerator, which provides hardware-accelerated compression.
  - Reduced the time that is required to perform a CRG switchover for some configurations.
  - The **CHGCRGPRI** command automatically displays the progress during a switchover.
- V7R5 base enhancements:
  - PowerHA SystemMirror for IBM i is now offered in a single edition that contains all the functions.
  - Geographic Mirroring compression: Geographic mirroring performs disk-level replication. When PowerHA performs a synchronization, whether for the initial synchronization, after an unplanned failure, or when resuming from a suspend operation, PowerHA replicates data over the Internet Protocol network.

    With this enhancement, PowerHA can now compress geographic mirroring data during resynchronization, which reduces the data transfer that is required. In addition, on Power10 servers running in Power10 compatibility mode, PowerHA uses the on-chip NX GZIP accelerator, which provides hardware-accelerated compression.
  - Reduced the time that is required to perform a CRG switchover for some configurations.
  - The **CHGCRGPRI** command automatically displays the progress during a switchover.

- V7R5 TR1:
  - A new, modern web-based interface for monitoring PowerHA environments.
  - Enhanced administrative domain integration with support for synchronizing user auditing (**CHGUSRAUD**) and maximum sign-on attempts (**MAXSIGN**) profile attributes for user profiles across the environment.
  - New IBM FlashSystem and IBM Spectrum Virtualize FlashCopy enhancements, including more automation and simplification of IASP-based FlashCopy.

  For more information about V7R5 TR1, see this announcement letter.

5.4 Installation, prerequisites, and options

To implement a HA solution, plan and configure a cluster. A cluster groups systems and resources in an HA environment.

Prerequisites

Here are the minimum hardware requirements for clusters:

- You need at least two systems or partitions running IBM i. The cluster supports up to 128 systems, which are defined as a cluster:
  - Any IBM i model that can run IBM i V4R4M0 or later can be clustered.
  - An external uninterruptible power supply or its equivalent is recommended to protect from a sudden power loss, which could result in a cluster partition.
- If you plan to use data resiliency technologies that require independent disk pools, plan for hardware that is specific to your chosen data resiliency technology. You can use different methods of disk protection to prevent failover from occurring if a protected disk fails.
- Install PowerHA SystemMirror for IBM i on each system that participates in the HA environment, and introduce the corresponding license codes on each system.

Installing PowerHA SystemMirror for IBM i

Starting with PowerHA SystemMirror IBM i V7R5M0, PowerHA is a single product version, which makes the installation process slightly different. The different processes are shown in this section.

V7R4M0

In an IBM i system under the V7R4M0 OS version, the first step to install the PowerHA SystemMirror for IBM i licensed program (5770-HAS) is to install IBM i 5770SS1 option 41 (HA Switchable Resources), which is a prerequisite for installing PowerHA SystemMirror for IBM i. Complete the following steps:

1. Enter **Go licpgm** from a CLI to access and work with the licensed programs display and select option 11 (Install licensed programs).
2. Select 5770SS1 option 41 (HA Switchable resources).
3. When the installation options panel opens, type the name of your installation device as requested (for example, OPTVRT01) and press Enter to start the installation.
4. After 5770SS1 option 41 is installed, repeat steps 1 and 3, but select the product 5770-HAS option *Base to install PowerHA SystemMirror for IBM i and press Enter.

5. Depending on your PowerHA scenario, select additional options for the 5770-HAS product:
   - IBM PowerHA for i Enterprise Edition (Option1) is required if you plan to use asynchronous geographic mirroring, Metro Mirror, Global Mirror, or DS8000 HyperSwap with IASPs.
   - IBM PowerHA for i Standard Edition (Option2) is required if you plan to use a GUI, commands, synchronous geographic mirroring, switched LUs, or FlashCopy.
   - IBM PowerHA for i Express Edition (Option3) is required if you plan to use DS8000 FullSystem HyperSwap.

After the installation completes successfully, the InetD server must be restarted.

V7R5M0

In an IBM i system under the V7R5M0 OS version, the first step to install PowerHA SystemMirror for IBM i (5770-HAS) is to install IBM i 5770SS1 option 41 (HA Switchable resources).

In this OS version, licensed program PowerHA SystemMirror for IBM i is now a single edition, so 5770-HAS *BASE and option1 are necessary to configure a HA environment.

The installation process is the same that is used for V7R4 (see “V7R4M0” on page 104), but the only choice is *Base and Option1 (now called PowerHA for i enablement).

5.5 Data resilience

Data resilience is the ability of data to be available to users or applications. You can reach data resilience by using IBM i cluster technology with PowerHA technologies or logical replication technologies.

You have several choices of technologies for IBM i supported implementations of data resilience, and combined with IBM i cluster services, you can build a solid and complete HA solution.

5.5.1 IBM i independent disk pool technologies

These technologies are based on an IBM i IASP. The data must be migrated to this IASP to be resilient. The supported IBM i technologies that are based on independent disk pools are shown in the following list:

- Switched LUs
- Geographic mirroring
- Metro Mirror
- Global Mirror
- HyperSwap with IASPs

Switched logical units

You can create a cost-effective HA solution by combining the IBM i cluster technology with switched LUs. This model supports planned and some unplanned outages. The CRG controls these LUs, which can be switched over manually or automatically in a failover. These switchable LUs must be in an IBM System Storage server that is connected through a storage area network (SAN). When a switchover or failover occurs, the LUs are reassigned from one system to another one.

Geographic mirroring

Geographic mirroring uses IBM i cluster technology to provide a HA solution that replicates the data between two copies of an IASP. This technology supports synchronous and asynchronous replication, and can be used with any storage, internal or external. Geographic mirroring protects against server and storage outages.

Geographic mirroring allows geographic separation between production and mirrored copies, but the distance between them can affect the response time of the applications. With asynchronous replication, the response times are not impacted as with synchronous replication.

Metro Mirror

Metro Mirror maintains a consistent copy of data between IBM System Storage external storage units. Metro Mirror with cluster technology provides a HADR solution. This technology mirrors data that is stored in independent disk pools from the source storage (production) unit to the target storage unit (backup), and provides availability for planned and unplanned outages. Source and target volumes can be on the same storage unit or on a separate one. If the storage units are separated, the target storage unit can be at another site up to 300 km away, but the performance can be affected when using synchronous data replication.

Global Mirror

Global Mirror provides a remote, long-distance remote copy across two sites by using asynchronous technology. It is designed to maintain a complete and consistent remote mirror of data asynchronously at unlimited distances with almost no impact to application response time.

HyperSwap with IASPs

HyperSwap is a storage HA solution that allows for LUs that are mirrored between two IBM System Storage DS8000 units to be switched with a near-zero outage time. When HyperSwap is implemented at the IASP level, it can be combined with other PowerHA technologies to provide a minimal downtime solution for planned and unplanned storage outages, and a minimal downtime solution for server planned and unplanned outages.

To use HyperSwap, you must have IBM PowerHA for i Enterprise Edition that is installed on your system. DS8000 HyperSwap with IASPs requires a cluster, and it uses PowerHA technologies.

5.5.2 DS8000 Full System HyperSwap technology

DS8000 Full System HyperSwap technology is an IBM System Storage technology that is integrated into PowerHA. Full System HyperSwap can be combined with Live Partition Mobility (LPM) to define an affinity between an IBM Power server where the IBM i LPAR is located and the storage server that is hosting the data.

5.5.3 PowerHA supported storage servers

PowerHA supports the following storage servers: DS8000 Storage family, IBM SVC, IBM FlashSystem servers, and the IBM Storwize models.

For the DS8000 family, IBM PowerHA also requires the installation of the DSCLI, which requires that at least Java 1.4 and option 35 (CCA Cryptographic Service Provider) are installed on each system or IBM i LPAR.

For SVC and Storwize solutions, IBM PowerHA also requires the installation of Portable Utilities for IBM i (5733-SC1).

5.6 High availability management

Planning, configuring, and managing a HA solution requires a set of management tools. With IBM i, several choices exist for HA management.

Depending on your needs and requirements and the OS version that you are using in your systems, HA management provides GUIs, commands, and APIs that can be used for creating and managing the HA environment. It is possible to use an IBM Business Partner application. Base IBM i V7R5 does not offer the PowerHA GUI integrated into IBM Navigator for IBM.

5.6.1 PowerHA SystemMirror for IBM i interfaces

PowerHA SystemMirror for IBM i is an end-to end HA offering. Combined with IASPs and HA Switchable resources (5770SS1 - Option 41), it enables a solution to be deployed by using IBM Storage Systems or internal disks.

PowerHA SystemMirror for IBM i provides a GUI that you use to configure and manage a HA solution, with the corresponding commands and APIs for functions that are related to HA technologies.

PowerHA GUI

On IBM i V7R4, a GUI is available. You can use the GUI to create and manage a cluster, CRGs, and device domains. Cluster administrative domains and IASPs use a single GUI.

On Base IBM i V7R5, this GUI is no longer available.

PowerHA commands

Use these commands to configure and manage your HA solution through a CLI. These PowerHA commands can be classified in the following categories:

- Cluster administrative domain commands
- Monitored resource entries commands
- Cluster commands
- Independent dis pools commands:
  - Copy description commands
  - Session commands
  - HyperSwap commands
  - HA configuration description commands

PowerHA APIs

With PowerHA APIs, you can work with different PowerHA versions, retrieve PowerHA related information, or implement IBM System Storage mirroring technologies and cross-site mirroring functions. Also, these APIs can be used by IBM i application providers or customers to enhance application availability.

- QhaChangeHAVersion API: Change the HA version.
- QhaListHAInfo API: List the HA information.
- QhaRetrieveHAInfo API: Retrieve the HA Information.
- QyasRtvInf API: Retrieve the ASP Copy information.

- QhaRetrieveConfigurationDesc API: Retrieve the HA configuration description.
- QhaRetrieveStatus API: Retrieve the HA status.

5.6.2 High availability functions in the base operating system

Some of the CL commands that are related to cluster management must remain in QSYS for debugging purposes or to perform certain operations like cluster objects deletion.

- DLTCRG: Delete a CRG.
- DMPCLUTRC: Dump a cluster trace.
- CHGCLURCY: Change a cluster recovery.
- STRCHTSVR: Start a clustered hash table server.
- ENDCHTSVR: End a clustered hash table server.

Also, you can write your own custom applications to manage and configure your cluster by using cluster APIs that are provided by Option 41 HA switchable resources.

5.7 Backup solutions in an HA environment

Performing a backup operation causes an unavailability of the IBM i system while it is being saved. To avoid this outage, there are some technologies that can be used to perform a backup.

5.7.1 FlashCopy

FlashCopy is an IBM System Storage technology that you use to take a snapshot of external disk units. In PowerHA solutions that use Metro Mirror or Global Mirror, FlashCopy technology helps in backup window reduction by taking a copy of data that then can be backed up to media without causing a production outage. To use the FlashCopy technology, a session must be created between the system and the external storage units.

This FlashCopy session can be used to start a backup LPAR with these disks attached and perform a backup operation by using Backup, Recovery, and Media Services (BRMS) or a **Save** menu.

Without any additional tools, all the steps must be done manually.

5.7.2 FSCF Toolkit

To automate the FlashCopy session, start the backup partition, and perform the backup by using BRMS, IBM Systems Lab Services created the Full System Copy Services Manager (FSFC) Toolkit. With this toolkit, you can perform a full system backup in a restricted state with a pause of less than 30 seconds on production, which is needed to quiesce the memory pages to disk before starting the FlashCopy session.

The FSFC for IBM i is unique because it does not require an IASP, but instead flashes the system ASP. FSFC provides extra functions to help automation and management of FlashCopy through a set of commands that you can use to create a point in time copy of a full system for performing backups.

Note: For more information about FSFC Toolkit, see PowerHA Tools for IBM i - Full System FlashoCopy.

5.7.3 PowerHA IASP Manager Toolkit

Written by IBM Systems Lab Services, this tool is designed for using PowerHA with IASP solutions. The tool provides simplicity and automation for PowerHA environments, and moves the management of the storage, HMC, and CSM functions to the IBM i server through CLI commands and automated scripts.

With this tool, a system backup can be performed as a fully automated FlashCopy process.

Note: For more information about IASP Manager Toolkit, see PowerHA Tools for IBM i - IASP Manager.

5.8 Third-party options for IBM i HA: logical replication

Although hardware forms of HA like PowerHA are now commonplace, there is a long-standing alternative, which is *logical replication*. Where PowerHA falls into the active/passive category with one partition in effect “owning” the database, logical replication replicates data from a SOURCE (production) to a TARGET (backup) partition, which means that in effect two copies of the data are available. Logical replication is designed around remote journaling, which is an efficient method of IBM i data replication for Db2, the Integrated File System (IFS), IBM MQ, and document library services file system (QDLS). This method of HA keeps the amount of data that is transferred to a minimum by replicating only changes, this method can be especially attractive for those organizations that are challenged by limited bandwidth.

Logical replication solutions do not require external storage or IASPs, although both storage solutions are supported. Often, the TARGET copy of the data is used for Business Intelligence purposes, so it is useful to keep query and reporting type workloads away from the more important production workload. Backups can also be conducted on the TARGET partition without downtime on the production partition. Furthermore, these third-party vendor solutions can co-exist and complement PowerHA by replicating objects that are not natively supported, or where limitations exist.

Software-based solutions can be used on-premises, in the cloud, or between the two because of their flexibility in providing synchronous or asynchronous methods of data transfer. Synchronous setups provide an acknowledgment before the next transaction is sent, and asynchronous setups provide a faster throughput and a greater tolerance to larger distances between the SOURCE and TARGET partitions.

For those organizations that want to use a cloud, software-based solutions can help with full migrations and provide a mechanism to use the cloud for a cost-effective method of DR. Both simulated role swaps and full role swaps are supported.

Software-based solutions are provided by Maxava, Precisely, and several other vendors, with products that are available for both small and enterprise environments alike.

5.9 Migrating and upgrading from previous releases

This section describes techniques for updating your PowerHA SystemMirror for IBM i to the latest level at the time of writing.

5.9.1 PowerHA SystemMirror for IBM i version support

A PowerHA version represents the level of a PowerHA SystemMirror for IBM i function that is available on a cluster. For example, Version 2.0 is a valid PowerHA version.

There are two PowerHA versions:

Potential PowerHA version Represents the version of PowerHA SystemMirror for IBM i that is installed on a node. The potential PowerHA is the highest PowerHA version that the node can support. The potential PowerHA version is updated when a new version of PowerHA SystemMirror for IBM i is installed.

Current PowerHA version Represents the version that is used for all cluster operations. This version is used for communications between the nodes in the cluster. To leverage all available new PowerHA functions, every node in the cluster must be at the latest potential PowerHA version, and the current PowerHA version must be adjusted to match.

Compatibility of PowerHA versions

PowerHA SystemMirror for IBM i supports up to a three-version difference among nodes. Nodes with a potential PowerHA version of at least equal to the current PowerHA version but not more than three versions higher than the current PowerHA version are compatible. For example, if the current PowerHA version is 2.1, nodes with a potential PowerHA version of at least Version 2.1 and less than Version 6.0 are compatible.

Setting the current PowerHA version

The PowerHA version is the version at which the nodes in the cluster that is managed by the PowerHA product are actively communicating with each other.

The PowerHA version values determine which functions can be used by the PowerHA product. The PowerHA version might require a certain cluster version to operate. For example, PowerHA Version 2.0 requires cluster Version 7.

The current PowerHA version is set when a cluster is created. If a cluster exists, the current PowerHA version is set to the lowest supported version.

Like cluster version, PowerHA has a current and potential version level. The current PowerHA version is the version at which the nodes in the cluster that are known by the PowerHA product are actively communicating with each other. The potential PowerHA version is the highest PowerHA version that the node can support. The PowerHA version cannot be changed until all PowerHA nodes are installed with a common potential PowerHA version. The potential PowerHA version can be between n and n+3. For example, the current PowerHA version is 2.0; NODE1 has a potential PowerHA version of 3.0; NODE2 has a potential PowerHA version of 4.0; and NODE3 has a potential PowerHA version of 5.0. All three nodes can support Version 3.0, so the current PowerHA version can be adjusted to 3.0.

Beginning with PowerHA Version 2.0, if a node with an incompatible potential PowerHA version is added to the cluster, the node successfully is added, but the node is considered “unknown” to PowerHA. A node is known to PowerHA if the node has the PowerHA product installed, and the potential PowerHA version is compatible with the current PowerHA version. If a node is unknown to PowerHA, certain product functions cannot be performed on the node.

For users of Cluster Administrative Domain, when the current PowerHA version is adjusted to 3.0 or higher, monitored resource entries in the cluster administrative domain might be affected. If all attributes that are supported at the previous cluster and PowerHA versions are being monitored for a MRE, then PowerHA automatically updates that MRE to monitor any new attributes that are supported at the current cluster version and current PowerHA version.

The current PowerHA version can be changed with the Change Cluster Version (**CHGCLUVER**) command. The Change Cluster Version (**CHGCLUVER**) command can be used only to adjust the version to the next higher cluster or PowerHA version. If you want to adjust the PowerHA version by two levels, the **CHGCLUVER** command must be run twice.

The current cluster version cannot be set higher than the lowest potential node version in the cluster. Likewise, the current PowerHA version cannot be set higher than the lowest potential PowerHA version of any node in the cluster. To view the potential node and PowerHA versions, use the Display Cluster Information (**DSPCLUINF**) command.

Since V7R3 TR8 or V7R4 TR3, PowerHA for IBM i with these levels automatically updates the PowerHA version after the upgrade or PTF is applied to all nodes within the environment, so the **CHGCLUVER** command is no required for enabling the new enhancements.

Adjusting the cluster version of a cluster

The cluster version defines the level at which all the nodes in the cluster are actively communicating with each other; the ability of a node to join the cluster; and the ability of the cluster to support new functions.

To change the cluster version, all nodes in the cluster must be at the same potential version. Then, the cluster version can be changed to match the potential version, and then the new function can be used. The version can be increased only by one, and it cannot be decremented without deleting the cluster and re-creating it as an earlier version. The current cluster version is set initially by the first node that is defined in the cluster. Subsequent nodes that are added to the cluster must be equal to the current cluster version or the next level version; otherwise, they cannot be added to the cluster.

If you are upgrading a node to a new release, you must ensure that the node has the appropriate cluster version. A cluster supports only a one-version difference. If all the nodes in the cluster are at the same release, upgrade to the new release before changing the cluster version to ensure that all functions that are associated with the new release are available.

Abbreviations and acronyms

AI	artificial intelligence
ARR	Automatic Repository Replacement
AWS	Amazon Web Services
BRMS	Backup, Recovery, and Media Services
CAA	Cluster Aware AIX
CBM	Cloud Backup Management
CCV	Cross-Cluster Verification
CG	Consistency Group
CoD	Capacity on Demand
CRG	cluster resource group
DLPAR	dynamic logical partitioning
DNP	Dynamic Node Priority
DR	disaster recovery
EPCoD	Enterprise Pool Capacity on Demand
ESB	Enterprise Service Bus
FC	Fibre Channel
FSFC	Full System Copy Services Manager
FW	firmware
GL	group leader
GLVM	Geographic Logical Volume Manager
GRS	Global Replication Service
HA	high availability or highly available
HACMP	High Availability Cluster Multi-Processing
HADR	high availability and disaster recovery
HMC	Hardware Management Console
HPC	high-performance computing
IASP	independent auxiliary storage pool
IBM	International Business Machines Corporation
IBM GDR	IBM Geographically Dispersed Resiliency
IFS	Integrated File System
IPAT	IP address takeover
IPv6	IP version 6
ISV	independent software vendor
JFS2	Enhanced Journaled File System
JOBD	job description
JSON	JavaScript Object Notation
LAN	local area network
LKU	Live Kernel Update
LPAR	logical partition or logical partitioning
LPM	Live Partition Mobility
LU	logical unit
LUN	logical unit number
LVM	Logical Volume Manager
MFA	multifactor authentication
MRE	monitored resource entry
NFS	Network File System
NRO	Network Recovery Objective
OCS	Operation Center Support
ODM	Object Data Manager
OS	operating system
PHYP	Power hypervisor
PKS	platform keystore
PRS	Platform Resource Scheduler
PVID	physical volume identifier
QDLS	document library services file system
RAS	reliability, availability, and serviceability
RG	resource group
RHEL	Red Hat Enterprise Linux
ROHA	Resource Optimized High Availability
RPO	recovery point objective
RPV	Remote Physical Volume
RSCT	Reliable Scalable Cluster Technology
RTO	recovery time objective
SAN	storage area network
SEA	Shared Ethernet Adapter
SLA	service-level agreement
SLIC	System Licensed Internal Code
SMTP	Simple Mail Transport Protocol
SMUI	SystemMirror User Interface
SPOF	single point of failure

SPOFs	single points of failure
SRDF	Symmetrix Remote Data Facility
SRR	Simplified Remote Restart
SSH	Secure Shell
SSP	Shared Storage Pool
SVC	IBM SAN Volume Controller
TR	Technology Release
UPS	uninterruptible power supply
V7R5M0	Version 7 Release 5
VIOS	Virtual I/O Server
VLAN	virtual local area network
VM	virtual machine
VMM	virtual memory manager
VMRM	VM Recovery Manager
VSP	Virtual Storage Platform

Related publications

The publications that are listed in this section are considered suitable for a more detailed description of the topics that are covered in this paper.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topics in this document. Some publications that are referenced in this list might be available in softcopy only.

- *Asynchronous Geographic Logical Volume Mirroring Best Practices for Cloud Deployment*, REDP-5665
- *Cloud Backup Management with PowerHA SystemMirror*, REDP-5651
- *Deploying PowerHA Solution with AIX HyperSwap*, REDP-4954
- *Guide to IBM PowerHA SystemMirror for AIX Version 7.1.3*, SG24-8167
- *High Availability and Disaster Recovery Options for IBM Power Cloud and On-Premises*, REDP-5656
- *High Availability for Oracle Database with IBM PowerHA SystemMirror and IBM Spectrum Virtualize HyperSwap*, REDP-5459
- *IBM PowerHA SystemMirror for AIX 7.1.3 Best Practices and Migration Guide*, SG24-8234
- *IBM PowerHA SystemMirror for AIX Cookbook*, SG24-7739
- *IBM PowerHA SystemMirror for i: Preparation (Volume 1 of 4)*, SG24-8400
- *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
- *IBM PowerHA SystemMirror Standard Edition 7.1.1 for AIX Update*, SG24-8030
- *IBM PowerHA SystemMirror V7.2 for IBM AIX Updates*, SG24-8278
- *IBM PowerHA SystemMirror V7.2.1 for IBM AIX Updates*, SG24-8372
- *IBM PowerHA SystemMirror V7.2.3 for IBM AIX and V7.22 for Linux*, SG24-8434
- *IBM PowerHA SystemMirror 7.1.2 Enterprise Edition for AIX*, SG24-8106
- *IBM SAN Volume Controller Stretched Cluster with PowerVM and PowerHA*, SG24-8142

You can search for, view, download, or order these documents and other Redbooks, Redpapers, web docs, drafts, and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites also are relevant as further information sources:

- Announcement OS V7R5:

  https://www.ibm.com/common/ssi/ShowDoc.wss?docURL=/common/ssi/rep_sm/1/877/ENUS5770-SS1/index.html

- *IBM i 7.5 Availability: High availability overview:*

  https://www.ibm.com/docs/en/ssw_ibm_i_75/pdf/rzarjpdf.pdf

- *IBM i 7.5 Availability: High availability technologies:*

  https://www.ibm.com/docs/en/ssw_ibm_i_75/pdf/rzauepdf.pdf

- *IBM i 7.5 Availability: Implementing high availability:*

  https://www.ibm.com/docs/en/ssw_ibm_i_75/pdf/rzaigpdf.pdf

- IBM i Documentation V7R3 Availability:

  https://www.ibm.com/docs/en/i/7.3?topic=availability

- IBM i Documentation V7R4 Availability:

  https://www.ibm.com/docs/en/i/7.4?topic=availability

- IBM i Documentation V7R5 Availability:

  https://www.ibm.com/docs/en/i/7.5?topic=availability

- IBM i Technology Updates:

  https://www.ibm.com/support/pages/node/1119129/

- Memo to Users V7R2:

  https://www.ibm.com/docs/en/i/7.2?topic=information-memo-users

- Memo to Users V7R3:

  https://www.ibm.com/docs/en/i/7.3?topic=information-memo-users

- Memo to Users V7R4:

  https://www.ibm.com/docs/en/i/7.4?topic=information-memo-users

- Memo to Users V7R5:

  https://www.ibm.com/docs/en/i/7.5?topic=documentation-memo-users

- Official PowerHA Wiki:

  https://ibm.biz/ibmi-powerha

- PowerHA SystemMirror Version 7.2.7 for AIX Enterprise Edition Notes:

  https://www.ibm.com/docs/en/powerha-aix/7.2?topic=notes-powerha-systemmirror-version-727-aix-enterprise-edition

- PowerHA SystemMirror Version 7.2.7 for AIX Standard Edition Release Notes:

  https://www.ibm.com/docs/en/powerha-aix/7.2?topic=notes-powerha-systemmirror-version-727-aix-standard-edition

- PowerHA Tools for IBM i - Full System FlashCopy:

  https://www.ibm.com/support/pages/node/1119435

- PowerHA Tools for IBM i - IASP Manager:

  https://www.ibm.com/support/pages/node/1126029

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

REDP-5694-00

ISBN 0738461113

Printed in U.S.A.