IBM HyperSwap and Multi-site HA/DR for IBM FlashSystem A9000 and A9000R

Francesco Anderloni
Bert Dufrasne
Roger Eriksson
Andrew Greenfield
Lisa Martinez
Stephen Solewin
Vadim Steckler

Storage

ibm.com/redbooks
International Technical Support Organization

IBM HyperSwap for IBM FlashSystem A9000 and A9000R

September 2018
Note: Before using this information and the product it supports, read the information in “Notices” on page vii.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notices</td>
<td>v</td>
</tr>
<tr>
<td>Trademarks</td>
<td>vi</td>
</tr>
<tr>
<td>Preface</td>
<td>ix</td>
</tr>
<tr>
<td>Authors</td>
<td>ix</td>
</tr>
<tr>
<td>Now you can become a published author, too!</td>
<td>xi</td>
</tr>
<tr>
<td>Comments welcome.</td>
<td>xi</td>
</tr>
<tr>
<td>Stay connected to IBM Redbooks</td>
<td>xi</td>
</tr>
<tr>
<td>Chapter 1. HyperSwap for IBM FlashSystem A9000 and A9000R</td>
<td>1</td>
</tr>
<tr>
<td>1.1 HyperSwap feature overview</td>
<td>2</td>
</tr>
<tr>
<td>1.1.1 Basic configuration</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2 HyperSwap as a critical feature for high availability</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Architecture, design, and components</td>
<td>5</td>
</tr>
<tr>
<td>1.2.1 HyperSwap volumes</td>
<td>5</td>
</tr>
<tr>
<td>1.2.2 Host paths</td>
<td>6</td>
</tr>
<tr>
<td>1.2.3 Failure domains</td>
<td>7</td>
</tr>
<tr>
<td>1.2.4 Connection with host stretch-cluster configurations</td>
<td>7</td>
</tr>
<tr>
<td>1.3 Independent QW</td>
<td>8</td>
</tr>
<tr>
<td>1.4 HyperSwap various topologies</td>
<td>10</td>
</tr>
<tr>
<td>1.4.1 Conventional topology</td>
<td>10</td>
</tr>
<tr>
<td>1.4.2 Symmetrical topology</td>
<td>10</td>
</tr>
<tr>
<td>1.4.3 Dedicated system topology</td>
<td>11</td>
</tr>
<tr>
<td>Chapter 2. Multi-site HA/DR solution for FlashSystem A9000 and A9000R</td>
<td>1</td>
</tr>
<tr>
<td>2.1 Multi-site HA/DR feature overview</td>
<td>2</td>
</tr>
<tr>
<td>2.2 Architecture, design, and components</td>
<td>3</td>
</tr>
<tr>
<td>2.2.1 Roles in a multi-site relationship</td>
<td>3</td>
</tr>
<tr>
<td>2.2.2 Standby asynchronous mirror</td>
<td>3</td>
</tr>
<tr>
<td>2.3 Multi-site HA/DR advantages</td>
<td>4</td>
</tr>
<tr>
<td>2.3.1 Flexibility</td>
<td>4</td>
</tr>
<tr>
<td>2.3.2 Non-disruptiveness</td>
<td>4</td>
</tr>
<tr>
<td>2.3.3 Ease of use</td>
<td>4</td>
</tr>
<tr>
<td>Chapter 3. Prerequisites</td>
<td>9</td>
</tr>
<tr>
<td>3.1 Storage systems and Quorum Witness prerequisites</td>
<td>10</td>
</tr>
<tr>
<td>3.1.1 FlashSystem A9000 and A9000R prerequisites</td>
<td>10</td>
</tr>
<tr>
<td>3.1.2 Quorum Witness prerequisites</td>
<td>11</td>
</tr>
<tr>
<td>3.2 Software levels</td>
<td>12</td>
</tr>
<tr>
<td>3.2.1 The Host Attachment Kit</td>
<td>12</td>
</tr>
<tr>
<td>3.2.2 SAN boot.</td>
<td>12</td>
</tr>
<tr>
<td>3.2.3 Quorum Witness software levels</td>
<td>12</td>
</tr>
<tr>
<td>3.3 Network requirement</td>
<td>13</td>
</tr>
<tr>
<td>3.3.1 Storage systems connectivity</td>
<td>13</td>
</tr>
<tr>
<td>3.3.2 Quorum Witness network requirements</td>
<td>13</td>
</tr>
<tr>
<td>3.4 Scalability</td>
<td>14</td>
</tr>
<tr>
<td>3.4.1 FlashSystem A9000 and A9000R scalability</td>
<td>14</td>
</tr>
<tr>
<td>3.4.2 Quorum Witness scalability</td>
<td>14</td>
</tr>
</tbody>
</table>
Chapter 4. HyperSwap implementation and usage ........................................................................ 15
  4.1 HyperSwap implementation process overview ................................................................. 16
  4.2 Quorum Witness setup ...................................................................................................... 16
    4.2.1 Quorum Witness download and installation ............................................................. 17
  4.3 Configuring the Quorum Witness on storage systems ...................................................... 21
    4.3.1 FlashSystem A9000 and A9000R HyperSwap configuration ................................. 21
  4.4 HyperSwap volume and Consistency Groups ................................................................. 28
    4.4.1 HyperSwap volume creation and activation .............................................................. 28
    4.4.2 Converting existing mapped volume to a HyperSwap volume ................................... 30
    4.4.3 Convert a synchronously mirrored volume to HyperSwap volume ............................ 34
    4.4.4 Convert a HyperSwap relationship to a synchronous mirror relationship ............... 36
    4.4.5 Define a Consistency Group as a HyperSwap Consistency Group ............................ 38
    4.4.6 Convert a synchronous mirror CG to a HyperSwap CG ........................................ 40
    4.4.7 Convert a HyperSwap CG to a synchronous mirror CG ........................................... 42
  4.5 Checking HyperSwap status ............................................................................................. 43
  4.6 HyperSwap snapshots ....................................................................................................... 45
    4.6.1 Creating HyperSwap snapshots .................................................................................. 45

Chapter 5. HyperSwap scenarios .............................................................................................. 49
  5.1 Failures on the storage system serving the primary copy .................................................. 50
    5.1.1 System A failure scenario ......................................................................................... 51
    5.1.2 Data service failure .................................................................................................... 53
  5.2 Connection failure between host and storage systems ...................................................... 53
    5.2.1 FC path failure between host and storage system .................................................... 54
    5.2.2 FC connectivity loss between host and System A and between System A and System B .................................................................................................................................................. 57
  5.3 Quorum Witness failure scenarios .................................................................................... 61
    5.3.1 Quorum Witness failure ............................................................................................ 62
    5.3.2 Ethernet connection loss between system A and QW ................................................. 66
    5.3.3 QW connectivity failure to System A, followed by mirroring link failure .................. 70
    5.3.4 Restoring or reinstalling the Quorum Witness .......................................................... 76
  5.4 Failback scenarios ............................................................................................................. 76
    5.4.1 Recovery from a failover where both volumes have primary role ............................... 77
    5.4.2 Recovery from an error during resynchronization .................................................... 83

Chapter 6. Multi-site HA/DR implementation and usage .......................................................... 89
  6.1 Multi-site HA/DR implementation process overview ....................................................... 90
    6.1.1 Assembly from scratch ............................................................................................... 91
    6.1.2 Assembly from existing Synchronous relation ........................................................... 91
    6.1.3 Assembly from existing HyperSwap ......................................................................... 93
    6.1.4 Assembly from existing Asynchronous relation ....................................................... 94
    6.1.5 Assembly with CG ..................................................................................................... 103
    6.1.6 Leaving CG with active HyperSwap or asynchronous relation ................................. 106
    6.2 Data migration ............................................................................................................... 108

Chapter 7. Multi-site scenarios .................................................................................................. 115
  7.1 Failure and recovery of the storage system serving the primary volume or CG ............... 116
    7.1.1 System A failure scenario ......................................................................................... 117
    7.1.2 Multi-site relationship recovery after System A failure ........................................... 119
    7.1.3 Host paths and IO behavior during the outage and recovery ....................................... 124
  7.2 Primary volume or CG loses target .................................................................................... 128
    7.2.1 Losing Secondary volume or CG .............................................................................. 128
    7.2.2 Losing Tertiary volume or CG .................................................................................. 130
  7.3 DR test at secondary volume site ....................................................................................... 132
7.4 Primary and Secondary volume or CG failure

7.4.1 Systems A and B failure scenario

7.4.2 Multi-site relationship recovery after Systems A and B failure

7.5 Quorum Witness failures

7.5.1 Loss of connection between the Primary system and the Quorum Witness

7.5.2 Loss of connection between both Primary and Secondary systems and the Quorum Witness

Chapter 8. Using VMware Site Recovery Manager and HyperSwap

8.1 Prerequisites

8.2 Installing IBM Spectrum Accelerate Family SRA

8.3 Configuring tags

8.4 Configuring a VMware Storage Policy

8.5 Configuring SRM

8.5.1 Create storage policy based protection group

8.5.2 Create a recovery plan

8.5.3 Test a recovery plan

8.5.4 Fail over a recovery plan

8.5.5 Reprotect virtual machines after recovery

Chapter 9. Microsoft Multi-Site Failover Clustering and HyperSwap

9.1 Microsoft Multi-Site Failover Clustering configuration

9.1.1 File Share Witness or Disk Witness

9.2 Planned and unplanned failovers

9.2.1 Planned failover

9.2.2 All Fibre Channel connections lost on the owner node

9.2.3 Site failure

Chapter 10. Volume Shadow Copy Service and HyperSwap Snapshots

10.1 VSS product and components

10.2 VSS function

10.3 IBM Spectrum Accelerate family VSS Provider (xProv)

10.3.1 Install the VSS Provider

10.3.2 Configure VSS Provider

10.3.3 Diskshadow command line utility

10.3.4 Create HyperSwap VSS snapshot

Related publications

IBM Redbooks

Other publications

Online resources

Help from IBM
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.

<table>
<thead>
<tr>
<th>AIX®</th>
<th>IBM Spectrum™</th>
<th>Redbooks (logo) ®</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS8000®</td>
<td>IBM Spectrum Accelerate™</td>
<td>Storwize®</td>
</tr>
<tr>
<td>FlashCopy®</td>
<td>IBM Spectrum Protect™</td>
<td>System Storage®</td>
</tr>
<tr>
<td>HyperSwap®</td>
<td>IBM Spectrum Virtualize™</td>
<td>XIV®</td>
</tr>
<tr>
<td>IBM®</td>
<td>Redbooks®</td>
<td></td>
</tr>
<tr>
<td>IBM FlashSystem®</td>
<td>Redpaper™</td>
<td></td>
</tr>
</tbody>
</table>

The following terms are trademarks of other companies:

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

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

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

IBM® HyperSwap® is the high availability (HA) solution that provides continuous data availability in case of hardware failure, power failure, connectivity failure, or disasters. The HyperSwap capability is available for IBM FlashSystem® A9000 and IBM FlashSystem A9000R, starting with software version 12.2.1.

Version 12.3 introduces a function that combines HyperSwap and Asynchronous replication, creating a solution that entails both High Availability (HA) and Disaster Recovery (DR). One side of the HyperSwap pair has an active async link to the third system, and the other side has a standby link. Known as Multi-site HA/DR, this configuration provides HyperSwap active-active high availability, while keeping data mirrored to a third copy to ensure two levels of business continuity.

This IBM Redpaper™ publication gives a broad understanding of the architecture, design, and implementation of HyperSwap and Multi-site HA/DR solution. It also discusses and illustrates various use cases pertaining to their use and functionality.

This paper is intended for those who want to deploy solutions that take advantage of HyperSwap and Multi-site HA/DR for FlashSystem A9000 and A9000R.

Authors

This paper was produced by a team of specialists from around the world working for the International Technical Support Organization, at the IBM European Storage Competence Center (ESCC) in Kelsterbach, Germany.

Francesco Anderloni is an Infrastructure Specialist working for the IBM Global Technology Services organization in IBM Italy. Since he joined IBM in 2015, he has been working in the Storage Management area, delivering IBM Storage solutions to clients in Italy and Europe. He is a member of the Young Technical Expert Council in IBM Italy. Francesco holds a Bachelor's Degree in Information Engineering and a Master's Degree in Computer Science, both from the University of Padova, Italy.

Bert Dufrasne is an IBM Certified Consulting IT Specialist and Project Leader for IBM System Storage® disk products at the International Technical Support Organization (ITSO). He has worked at IBM in various IT areas. He has authored many IBM Redbooks® publications and has also developed and taught technical workshops. Before joining the ITSO, he worked for IBM Global Services as an Application Architect. He holds a Master's degree in Electrical Engineering.

Roger Eriksson works at IBM Systems Lab Services Nordic, based in Stockholm, Sweden. He is a Senior Accredited IBM Product Service Professional. Roger has over 25 years of experience working on IBM servers and storage, including Enterprise and Midrange disk, NAS, SAN, IBM x86 and IBM Power. He has done consulting, proof of concepts and education, mainly with the Spectrum Accelerate and Spectrum Virtualize product line, since December 2008. He also works with the Swedish technical sales team for storage solutions on a daily basis. Working with both clients and various IBM teams worldwide is a normal days work. He holds a Technical Collage Graduation in Mechanical Engineering.
Andrew Greenfield is an IBM Global XIV and Flash Solution Engineer who is based in Phoenix, Arizona. He holds numerous technical certifications from Cisco, Microsoft, and IBM. He is also responsible for many of the photos and videos that are featured in this book and at http://www.IBM.com. Andrew brings over 24 years of data center experience with Fortune 100 companies to the team. He graduated Magna cum Laude, Honors, from the University of Michigan, Ann Arbor. Andrew has also written other IBM FlashSystem and XIV Gen3 IBM Redbooks publications.

Lisa Martinez has been working in the Washington Systems Center—Storage Team as a storage specialist since January 2012. Her focus has been with pre-sales support for DS8000 and the IBM Spectrum Accelerate family of storage products (FlashSystem A9000 and A9000R, XIV and IBM Spectrum Accelerate). She is also the lead instructor for FlashSystem A9000 and A9000R, XIV and Spectrum Accelerate customer based workshops. Prior experience includes roles as a storage architect in the Specialty Services Area in GTS and test architect in disk storage focusing on system level test for XiV for three years, and copy services for DS8000. Lisa holds degrees in Computer Science from New Mexico Highlands University and Electrical Engineering from the University of New Mexico.

Stephen Solewin is an IBM Corporate Solutions Architect based in Tucson, Arizona. He has over 20 years of experience in working on IBM storage, including enterprise and midrange disk, Linear Tape-Open (LTO) drives and libraries, SAN, storage virtualization, and storage software. Steve is a global resource, working with customers, IBM Business Partners, and fellow IBMers worldwide. He has been working on the XIV product line since 2008 and FlashSystem since 2013. He holds a BS degree in electrical engineering from the University of Arizona, where he graduated with honors.

Special thanks to Yossi Siles, Rivka Matosevitch, and Iftach Landau for their support and advice.

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

Anat Bar-Natan-Kigel
Kobi Beifus
Craig Gordon
Detlef Helmbrecht
John Hyams
Markus Oscheka
Marie Romero
Vadim Steckler
Max Vidgerovich
Ralf Wohlfart
IBM
Now you can become a published author, too!

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

Find out more about the residency program, browse the residency index, and apply online at: ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our 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, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

► Find us on Facebook:
  http://www.facebook.com/IBMRedbooks

► Follow us on Twitter:
  http://twitter.com/ibmredbooks

► Look for us on LinkedIn:
  http://www.linkedin.com/groups?home=&gid=2130806

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

► Stay current on recent Redbooks publications with RSS Feeds:
  http://www.redbooks.ibm.com/rss.html
Chapter 1. HyperSwap for IBM FlashSystem A9000 and A9000R

This chapter provides a high-level, functional overview of the IBM HyperSwap feature that is available on IBM FlashSystem A9000 and A9000R systems, starting with system software version 12.1.

This chapter covers the following topics:
- HyperSwap feature overview
- Architecture, design, and components
- Independent QW
- HyperSwap various topologies
1.1 HyperSwap feature overview

HyperSwap is an IBM trademark that names the capability of active-active storage deployments with automatic failover. It is available with various IBM product lines, but its architecture and implementation are different in each case. This overview specifically applies to HyperSwap for FlashSystem A9000 and A9000R.

The HyperSwap feature of FlashSystem A9000 and A9000R, also referred to as transparent failover, delivers always-on, high availability (HA) storage service for storage volumes or Consistency Groups in a production environment. It is based on an active-active, cross-system, and cross-datacenter configuration and does not require additional licensing or special hardware.

HyperSwap builds upon the synchronous mirroring functionality that is already included with FlashSystem A9000 and A9000R, together with other advanced replication and stretched-cluster features.

HyperSwap relies on Asymmetrical Logical Unit Access (ALUA) support to indicate to the host the preferred paths to the storage system, and minimize latency.

HyperSwap volumes can autonomously and transparently switch between primary and secondary roles, based on the volume failover state. In effect, from a host perspective, the pair of mirrored volumes on both mirrored systems constitute a single HyperSwap volume, also referred to as a stretched volume.

1.1.1 Basic configuration

The basic HyperSwap configuration is shown in Figure 1-1.

![HyperSwap volume on peer systems serving host](image)

The basic HyperSwap configuration consists at a minimum of the following components:

- **Two storage systems**: It can be any combination of FlashSystem A9000 or FlashSystem A9000R. The two systems must be able to interconnect for synchronous mirroring over Fibre Channel.
Application host: The application host must have read and write access to both storage systems. Host I/Os can be over iSCSI or Fibre Channel.

Quorum Witness: As in other high availability solutions, HyperSwap requires a quorum witness component. The Quorum Witness (QW) is included with the HyperSwap solution. Its role is to monitor both storage systems health and to arbitrate which storage system owns the primary volume. The QW is ideally located at a third, separate, physical site with IP connectivity to both storage systems.

Note: The basic configuration shown in Figure 1-1 on page 2 represents what is known as a uniform configuration where the host can access both the primary and secondary volumes. The uniform configuration is the preferred practice to protect a host from data access problems.

In a non-uniform configuration, the storage high availability relies on the server detection and failover of the application to the server with access to active storage. A non-uniform configuration can be used when the host is part of a cluster, can fail over to another host in the cluster, and the other host is connected to the peer system. This configuration is less costly from network perspective. However, it relies on the host failover, which in most cases would not be necessary in a uniform configuration.

1.1.2 HyperSwap as a critical feature for high availability

The automatic and transparent failover at the volume or Consistency Group level is critical to high availability in partial or complete system failures, as well as in site disaster scenarios, either in the same datacenter or between metro-distant datacenters serving single or clustered hosts.

HyperSwap protected workloads are not disrupted through most storage, network, server, application, and site failures and disasters. Enterprise IT systems can automatically fail over within seconds, without human intervention and in a straightforward way using a metro-distance stretch-cluster, to protect against loss of access to mission-critical data.

HyperSwap is, therefore, a critical feature for high availability, as further described in the following sections:

- Always-on, zero-downtime volume operation
- Simplicity
- Robustness
- Data protection
- Typical high availability use cases for HyperSwap

Always-on, zero-downtime volume operation

The ability to attain always-on flash storage systems is critical to applications (such as time-sensitive financial applications) that require constant operation with availability on a 24 by 7 basis. The active-active pairing, per volume or Consistency Group enables zero downtime operation.

Each individual FlashSystem A9000 or A9000R system has a rating of five nines high availability, without HyperSwap. The possibility that two totally separate systems will fail at once, is equivalent to a failure of a system with 10 nines availability rating. To get a sense of the rarity, a system with nine nines is expected to have a total of 31.5569 milliseconds downtime a year.
However, the overall high-availability (number of nines) of the solution also depends on the reliability of the applications, hosts, networks, power systems, cooling systems, and separation of sites. Generally, the high availability of a HyperSwap solution is as good as the weakest link in the chain. When all the components are redundant and appropriately separated among fault-domains, the overall availability from an application perspective can go above and beyond the five nines, and HyperSwap is made to enable such levels of reliability\textsuperscript{1}.

**Simplicity**

Using active-active systems is a way to greatly simplify failure management. Automation is provided mostly by the solution and does not require complex scripting or procedures. This aspect is especially important for smaller organizations that cannot deal with failover automation complexity.

**Robustness**

HyperSwap behavior is designed to address the impact of failures and disasters. The HyperSwap function can identify the following failures and recover from them automatically:

- Replication is down
- The QW is down
- The QW connectivity is down

**Data protection**

With HyperSwap on IBM FlashSystem A9000 and A9000R:

- Protected hosts and applications continue to benefit from non-disruptive storage service during any of the following failures:
  - Host connectivity to any one of the peer systems is down, for any reason.
  - One of the peer systems does not respond to the host, for any reason.
  - One of the peer systems is completely down, for any reason.
  - A host application or a whole host that is part of a server high availability cluster, fails-over to a remote host that is already connected to the peer HyperSwap volume.
- VMware virtual machines can be migrated non-disruptively using VMware Storage vMotion across vCenter servers and across data centers, supporting VMware vSphere Metro Storage Cluster (vMSC) configurations, as well as VMware Site Recovery Manager (SRM) 6.1 or later.

**Typical high availability use cases for HyperSwap**

HyperSwap on IBM FlashSystem A9000 and A9000R is mostly useful in the following circumstances:

- Critical applications must run on an all-flash storage system and require continuous operation.
- Requirements mandate active-active all-flash storage capability.
- When used with VMware Site Recovery Manager 6.1 (or later), HyperSwap enables zero-downtime with faster recovery for VMware ESXi clusters.

Regulations or policies require system administrators to be off-site during weekends and holidays. Fully automated disaster recovery (DR) might be fast enough to compensate for administrator absence, but could be too complex or fragile for operation.

\textsuperscript{1} The actual level of availability for applications that use HyperSwap in customer environments requires solution-specific expert assessment.
1.2 Architecture, design, and components

This section provides some insight into the architecture, design and components of the HyperSwap implementation for FlashSystem A9000 and A9000R.

1.2.1 HyperSwap volumes

HyperSwap high availability is based on dual storage systems with active-active pairing per volume, or per Consistency Group. It is important to note that a HyperSwap relationship is not defined at the storage system level. The HyperSwap function is between peer volumes or peer Consistency Groups on two separate FlashSystem A9000 or A9000R systems.

HyperSwap volumes have a copy on one storage system at one site and a copy on the other storage system at another site. Data that is written to the volume is automatically sent to both copies; if one site is no longer available, the other site can provide access to the volume.

Each volume or Consistency Group pair is using synchronous replication to keep both systems updated at all times. When certain conditions apply, an autonomous and completely transparent failover from a volume to its peer is performed, so that host applications experience no downtime.

As previously indicated, HyperSwap for FlashSystem A9000 and A9000R builds upon the synchronous mirroring functionality. As is the case with mirroring, one volume is designated as the primary and its peer is designated as the secondary. With HyperSwap, volumes can automatically and transparently switch between primary and secondary roles.

Primary and secondary volumes in a HyperSwap relationship are read and write enabled for host I/O. In effect, the pair of volumes on both mirrored systems have identical SCSI attributes and, as such, act as a single HyperSwap volume, also referred to as a stretched volume.

As shown in Figure 1-2, each HyperSwap-enabled volume exists simultaneously on two IBM FlashSystem A9000 or A9000R peer systems (depending on the specific deployment), in an active-active, synchronous replication relationship over Fibre Channel (FC).

![HyperSwap volume in synchronous replication over FC](image)

The pair of volumes in a HyperSwap relationship have a unique an identical SCSI identity in Network Addressing Authority (NAA) format and the same characteristics relative to I/Os (size, lock state, SCSI reservations). Each storage system with volumes in a HyperSwap configuration has its own volume mapped to the application host, but the application sees them as one LUN. A HyperSwap stretch volume behaves as a single volume in all SCSI aspects. HyperSwap guarantees that the host can use the stretch volume as a single LUN.
Storage administrators can non-disruptively and without pausing synchronization, convert synchronously mirrored volumes to HyperSwap volumes, and vice versa. HyperSwap volumes can also be set up using offline initialization.

1.2.2 Host paths

For reliability host operating systems include a multipath storage driver that allows multiple paths to a storage volume. With recent operating systems, those paths are optimized using Asymmetric Logical Unit Access (ALUA) support from the multipath driver. ALUA, also known as Target Port Groups Support (TPGS), consists of SCSI concepts and commands that establish path preferences for storage devices.

In simple terms, ALUA allows a storage system to indicate to an operating system the state of the port group. The states include both priority (preferred or non-preferred) and functionality (active or unavailable). When everything is healthy, the port group state conveys only the preference, but in failure scenarios, it also tells the host which ports cannot be used for I/O.

The purpose of the primary secondary designation is to optimize latency: the primary volume must be co-located with the hosts that generate most of the I/O. The HyperSwap function in FlashSystem A9000 and A9000R marks paths to the system that currently owns the primary volume as active/preferred, while paths to the system that currently owns the secondary volume are marked as active/non-preferred. Consequently, most of the write I/Os will go directly to the primary volume because it is the preferred path.

**Note:** The host experiences a failover as a change in the state of its paths. The change is therefore handled by the host multipath driver with no impact to the application (except some delay) and no manual intervention.

Non-preferred paths are used by hosts to load-balance or mitigate disconnects. Occasional write I/Os going over the active, but not preferred path to the secondary volume are redirected to the primary. In this case the secondary volume acts as a proxy of the primary volume. Refer to Figure 1-3 for a details about the transaction.

![Figure 1-3 Write I/O to secondary](image-url)
Read I/Os requests on the non-preferred path are normally served directly by the secondary. However if the relationship was just activated and the secondary is not yet fully synchronized, read requests are also redirected to the primary volume until synchronization is complete.

### 1.2.3 Failure domains

For best protection against different types of possible failures, each of the two storage systems and the QW must be located in distinct failure domains. The failure domains must be defined by the failures that must be overcome. For example, three physical sites can be considered as three failure domains with regard to failures that affect the physical site but not in case of a major power failure when the three physical sites are connected to the same power grid.

Typically, you want to be protected from any type of disaster that might either affect both storage systems at the same time or that would simultaneously disable one of the storage systems and the quorum witness application.

Keep in mind, however, that the physical sites must still be within distances that allow synchronous mirroring between the storage systems, and acceptable latency. See Chapter 3, “Prerequisites” on page 9 for details.

### 1.2.4 Connection with host stretch-cluster configurations

HyperSwap in stretch-cluster configurations include the following properties, as illustrated in Figure 1-4:

- Protected hosts are connected to both systems locally and remotely.
- I/O path priorities are assigned for minimum latency, by default. The two systems assign a preferred or non-preferred Asymmetric Logical Unit Access (ALUA) state to the port groups, which are translated by the host multipath driver to path priorities, resulting in optimized I/O latency.
- Multiple HyperSwap volumes can exist on each pair of peer systems.
1.3 Independent QW

An independent QW software component is used as part of the HyperSwap solution in order to enable transparent failover and facilitate coordination between two storage systems with volumes in a HyperSwap relationship. The QW role is to continuously monitor the systems health by using dedicated heartbeat messages and, when needed, to serve as a tie-breaker or ruler in possible split-brain scenarios.

The most fundamental function of the QW is to determine upon failure, per HyperSwap volume or Consistency Group (CG), which system should own the primary volume or CG and which system should own the secondary volume or CG.

Whenever a failure is detected at the system level, the QW is used immediately by the storage systems to resolve any potential contention and to determine which system should own each primary volume or CG. The QW component is, therefore, critical for the complete always-on HyperSwap functionality.

The two systems send keepalive messages to report their health to the QW, as illustrated in Figure 1-5. The systems also retrieve the information pertaining to the health of their peer systems from the QW. The communication with the QW is performed over Ethernet using a TCP/IP. A system can detect the failure of its peer system through the mirroring connectivity between them. If the peer failure is verified through the QW information, the system then takes over as owner of primary volumes or Consistency Groups, provided that it has a synchronized copy of the volume or Consistency Group data.

Some rare failure scenario might result in both systems attempting to take over the primary volumes. The QW is used to tie-break these cases and make sure a single system wins and becomes the primary for the volume or CG.

The automatic recovery ensures that, upon failure, the protected host can access at least one of the volume instances.
The QW node consists of a small-footprint software application, typically deployed separately at a third site.

**Tip:** The QW is preferably deployed as a highly-available virtual machine on a VMware vSphere 6.1 (or later) cluster, using VMware HA.

Figure 1-6 illustrates a typical configuration of hosts, storage systems, and a QW node in a HyperSwap high availability solution. Typically, the hosts and the storage systems with primary volumes are located at the same site, and the QW is deployed at a separate third site.
1.4 HyperSwap various topologies

HyperSwap on IBM FlashSystem A9000 and A9000R supports various topologies. We review in this section, the most typical ones.

1.4.1 Conventional topology

In a conventional topology, as shown in Figure 1-7, primary volumes or Consistency Groups are located in one of the systems, and the secondary volumes or Consistency Groups are located in the other system.

![Figure 1-7 Conventional HyperSwap topology](image)

1.4.2 Symmetrical topology

In a symmetrical topology, as shown in Figure 1-8, the systems have both primary and secondary volumes or Consistency Groups. This configuration can be used to split production workload over the two systems and thus balance the I/O utilization of the configuration.

![Figure 1-8 Symmetrical topology](image)
1.4.3 Dedicated system topology

In a dedicated system topology, as shown in Figure 1-9, a single storage system at a third site is dedicated to serving two primary systems simultaneously, and is initially configured to contain only the secondary volumes.

Figure 1-9  Dedicated HyperSwap system for secondary volumes topology
Multi-site HA/DR solution for FlashSystem A9000 and A9000R

This chapter covers the following topics:

- Multi-site HA/DR feature overview
- Architecture, design, and components
- Multi-site HA/DR advantages
2.1 Multi-site HA/DR feature overview

Introduced with the IBM FlashSystem A9000 and A9000R storage system software version 12.3.0, the Multi-site HA/DR feature allows the deployment of well established high availability (HA) and disaster recovery (DR) solutions over multiple sites, keeping three concurrent copies of data.

As shown in Figure 2-1, a Multi-site HA/DR configuration is composed of three IBM FlashSystem A9000 or A9000R systems in a multi-site topology. It comprises:

- One HyperSwap relationship between volumes on systems A and B
- One active asynchronous mirroring relationship between volumes on systems A and C
- Optionally, one standby asynchronous mirroring relationship between volumes on systems B and C

This configuration provides HyperSwap active-active high availability, while keeping data mirrored to a third site to ensure two levels of business continuity.

System B holds the most recent snapshot and last replicated snapshots automatically created by the mirroring relationship with volumes on System A. Upon a HyperSwap failover, the asynchronous replication of volumes from System A to System C stops, and the standby asynchronous relation between corresponding volumes on System B and System C is automatically activated. At this point, the MRS and LRS on System B are used to initialize the asynchronous replication with System C from exactly the same point where the System A to System C replication stopped.

The Multi-site HA/DR feature is able to operate with volumes and consistency groups.

Any storage system in a multi-site relation can have multiple multi-site relations with volumes and consistency groups (CG) in different roles.
2.2 Architecture, design, and components

A multi-site configuration is typically defined by extending a HyperSwap relationship or an asynchronous mirror. In other words, defining the multi-site relationship assumes that at least there already exists a fully initialized two-way HyperSwap or asynchronous mirroring relationship.

The multi-site relationship runs on system A (see Figure 2-1) by activating the A-B and A-C relationships. If one of these is already active, no further changes to that existing relation are required.

**Important:** Running the same multi-site configuration on system B or C is not allowed.

In a multi-site configuration, the asynchronous replication can be over Fibre Channel (FC) or iSCSI, while the HyperSwap connection must be FC only.

2.2.1 Roles in a multi-site relationship

The Multi-site HA/DR feature introduces some new terms in addition to the familiar HyperSwap and asynchronous mirroring terminology.

As shown in Figure 2-1, the volume/CG roles are defined as follows:

- **Primary volume**
  The primary volume in a HyperSwap relationship with the secondary volume, and in an asynchronous mirroring relationship with the tertiary volume. Normally, the primary volume resides on system A.

- **Secondary volume**
  The secondary volume in a HyperSwap relationship between A and B. This volume also acts as the primary volume in the optional standby asynchronous mirroring relationship between B and C. Normally, the secondary volume resides on system B.

- **Tertiary volume**
  The secondary volume in two asynchronous mirroring relationships, between A and C (active) and B and C (standby). Normally, the tertiary volume resides on system C.

In a regular, properly operating multi-site relationship, the role of each volume is as indicated above.

**Important:** As a counterpart to "Primary", "Secondary", and "Tertiary", the input and output syntax of CLI commands uses the legacy terminology of "Master", "sMaster", and "Slave". This inconsistency is a necessary compromise, required to avoid changes to older CLI commands that are in customer use, and also to keep the CLI terminology consistent across the board.

2.2.2 Standby asynchronous mirror

The standby asynchronous mirror between systems B and C can be defined either in advance, at the time of the multi-site relationship assembly, or when needed for data recovery.

Defining the standby mirror in advance requires that the target connectivity between the systems B and C (or at least its definitions) be in place when the multi-site relationship is
assembled. When defined, the B-C asynchronous mirror remains on standby under normal conditions and only becomes active, by request, when a disaster occurs on system A, making the primary volume inaccessible and triggering a failover in the A-B HyperSwap relationship.

The preferred practice is to define the standby asynchronous relationship in advance.

2.3 Multi-site HA/DR advantages

As implemented in IBM FlashSystem A9000 and A9000R, the Multi-site HA/DR solution offers the following advantages, as discussed hereafter.

2.3.1 Flexibility

Any of the three peer systems participating in a multi-site relationship can be either BM FlashSystem A9000 and A9000R. System models do not have to be identical. In other words, a multi-site relationship peer system can be model 415 or 425.

The Multi-site HA/DR feature does not require separate licensing.

2.3.2 Non-disruptiveness

The Multi-site HA/DR solution is designed to maintain non-disruptive host data access and recover data in case of a disaster.

The administrator assembles it from existing relationships. Any existing two-way relationship (HyperSwap or asynchronous) can be extended to a multi-site relationship without disrupting the existing relation.

2.3.3 Ease of use

Multi-site HA/DR is easy to configure and monitor, because its administration and management are based on existing features.

IBM Hyper-Scale Manager provides special management for the Multi-site HA/DR solution.
Prerequisites

This chapter covers the prerequisites and dependencies for implementing the IBM HyperSwap with the IBM FlashSystem A9000 and A9000R systems. The corresponding topics include the solution infrastructure, minimum code version, the Quorum Witness specifications, hardware and software requirements for the storage systems and Quorum Witness, as well as scalability considerations.

This chapter includes the following topics:

- Storage systems and Quorum Witness prerequisites
- Software levels
- Network requirement
- Scalability
3.1 Storage systems and Quorum Witness prerequisites

This section covers the requirements and prerequisites for FlashSystem A9000 and A9000R and the Quorum Witness (QW) to use the IBM HyperSwap feature.

3.1.1 FlashSystem A9000 and A9000R prerequisites

For any volume or Consistency Group, HyperSwap is implemented only between two FlashSystem A9000 and A9000R systems (any combination of the two). It includes the following requirements specific to these storage systems:

- Minimum software level to support HyperSwap is Version 12.1 and Hyper-Scale Manager V5.2 or later is required.
- For Multi-site HA/DR, any FlashSystem A9000 or A9000R participating in the multi-site solution must all be running at least Version 12.3 of the storage system software. To manage a multi-site environment, Hyper-Scale Manager V5.5 or later is required.
- When defining hosts, make sure that the correct type is selected. If you use Windows 2008, HP-UX or z/VM host, you must change the type to match your host type. For all other hosts, such as AIX, Linux, Solaris, VMware, and Windows (except Windows 2008), the Default (or All Others) option is correct. Note that starting with Host Attachment Kit (HAK) version 2.7.0, HP-UX and Solaris are no longer supported.
- PKCS12 certificates must be installed and enabled for the QW service, with the following different certificates of importance:
  - The QW certificate that is generated by the QW during QW installation must be copied to the local machine where you currently use Hyper-Scale Manager (see step 8 on page 20 in Chapter 4, "HyperSwap implementation and usage" on page 15). You also need to install this certificate on each FlashSystem A9000 or A9000R system using the corresponding QW (see 4.3, "Configuring the Quorum Witness on storage systems" on page 21).
  - On FlashSystem A9000 or A9000R systems where the HyperSwap volumes reside, the system certificate must be assigned to the QUORUM service:
    - For systems shipped with a minimum of Version 12.1 code, the certificate that was preinstalled by manufacturing is already assigned to the QUORUM service.
    - For older systems, you must assigned the certificate to the QUORUM service after the code is upgraded.

In any case, use the XCLI `pki_list` command to check which services are already enabled as follows:

```
A9000R>> pki_list
Name Fingerprint Has signed certificate Services
XIV c5dfc4f842946b669314e544d189c6f7 yes XCLI,CIM,IPSEC
```

Then, if QUORUM is not listed among the services, use the `pki_update` command to use the same certificate (XIV, in this example) for the QUORUM service as well:

```
A9000R>> pki_update name=XIV services=XCLI,CIM,IPSEC,QUORUM
Command executed successfully.
```

To check whether the QUORUM service was added, issue the `pki_list` command again:

```
A9000R>> pki_list
Name Fingerprint Has signed certificate Services
XIV c5dfc4f842946b669314e544d189c6f7 yes XCLI,CIM,IPSEC,QUORUM
```
A minimum of four Fibre Channel ports must be used for replication between the two systems, set up as initiator and target on each system.

A maximum of two QW servers per storage system.

All management ports (three ports per system, thus a total of six) must be connected and configured.

Management ports for the active QW must be connected and configured.

**Note:** Always refer to the latest storage system release notes to check for HyperSwap support. You can find release notes in [IBM Knowledge Center](https://www.ibm.com/support/knowledgecenter).

### 3.1.2 Quorum Witness prerequisites

The QW is used to enable a non-disruptive, automatic failover in HyperSwap. Without a functioning QW, automatic failover cannot be performed. Its role in the solution is to track and report on the health of both storage systems and to serve as a tie-breaker in split-brain scenarios.

The QW application can be installed on a dedicated server or a virtual machine (VM) with the following minimum hardware requirements:

- RHEL 6.8 or 7.3, CentOS 6.8 or 7.3
- 64-bit dual-core CPU
- 4 GB RAM
- Hard drive requirements:
  - SAS/SATA interface
  - at least 7200 RPM
  - RAID protection
  - 40 GB free space
- 1 Gbps broadband network connection

Because the solution is implemented with an application for the QW, the user installing the application must have relevant permissions on the host (physical or virtual) to install the RPM and BIN files.

Installing the QW application on a highly available VM will ensure the QW will still be available in the event of a hardware problem on the server or VM. However, for the best performance, use a dedicated server.

A maximum of 12 storage systems can be configured per QW.

For full details about which ports to open and installing the application, refer to the *IBM Spectrum Accelerate Family HyperSwap Quorum Witness User Guide*, SC27-4631, available on IBM Knowledge Center.
3.2 Software levels

This section covers the different software levels required to implement HyperSwap and Multi-site HA/DR.

3.2.1 The Host Attachment Kit

The minimum Host Attachment Kit (HAK) version of IBM FlashSystem A9000 or A9000R for HyperSwap is 2.8.

For information about the HAK installation, review IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System: Host Attachment and Interoperability, SG24-8368. In addition, for VM solutions, refer to Using the IBM Spectrum Accelerate Family in VMware Environments: IBM XIV, IBM FlashSystem A9000 and IBM FlashSystem A9000R, and IBM Spectrum Accelerate, REDP-5425.

You can download and install the HAK from Fix Central. For further information about using the HAK, refer to the IBM Storage Host Attachment Kit welcome page in IBM Knowledge Center.

3.2.2 SAN boot

By using SAN boot make sure that in the adapter BIOS targets (WWPNs) from both storage systems are defined. The adapter BIOS may be changed during boot time or by using tools from the HBA vendor running on the operating system.

3.2.3 Quorum Witness software levels

You can install QW software on a bare metal server using compatible versions of Red Hat Enterprise Linux (RHEL), CentOS, or as a VMware virtual machine.

Verify that the dependencies mentioned in IBM Spectrum Accelerate Family HyperSwap Quorum Witness User Guide, SC27-4631, are installed prior to installing the QW application itself.

These dependencies as mentioned previously are also supplied in the installation package together with the QW application BIN file and will be deployed prior to the application installation.
3.3 Network requirement

This section lists the configurations that have to be enabled or changed in both storage systems or in the network infrastructure to use a HyperSwap with FlashSystem A9000 and A9000R (any combination of the two).

3.3.1 Storage systems connectivity

The following storage systems connectivity requirements must be met, for HyperSwap:

- Fibre Channel connectivity of 4, 8, or 16 GB between the two FlashSystem A9000 and A9000R systems, for mirror connectivity.
- A minimum of four Fibre Channel ports for each primary and secondary system is required, for mirror connectivity.
- All Fibre Channel ports of each storage system that have volumes in HyperSwap relationships, must be in the same zone or zones.
- Users should have the ability to control the zone settings of the target and initiator FlashSystem A9000 and A9000R systems for failover situations.
- At least two host ports of either iSCSI or Fibre Channel on each target and initiator systems should be available for host application connectivity. In case of Fibre Channel connectivity, Fibre Channel ports on both storage systems should be set as target ports, each connected and zoned with the host application server.
- HyperSwap distances are limited by synchronous mirroring distances (typically up to 100 km apart) depending on the available bandwidth and impact of latency on the host. If the application is latency sensitive, you might want to further limit the distance to a maximum of 75 km.

For more details, see *IBM FlashSystem A9000 and A9000R Business Continuity Solutions*, REDP-5401.

For a Multi-site deployment:

- The distance between systems follow the same limitations that apply to HyperSwap and asynchronous mirroring.
- The asynchronous replication can be over Fibre Channel (FC) or iSCSI, while the HyperSwap connection must be FC only.

3.3.2 Quorum Witness network requirements

The QW network requirements are as follows:

- At least one network connection of one Gigabit configured and connected to be able to reach all management ports of both storage systems.
- It requires Secure Sockets Layer (SSL) certification to communicate with the storage systems.
- The connection must have ports TCP 8460 (for QW API communication), TCP 8461 (for retrieving logs), and TCP 8462 (for retrieving Nginx statistics) unrestricted.
- Maximum latency allowed between the QW and FlashSystem A9000 or A9000R is 750 ms.
3.4 Scalability

This section covers the scalability of HyperSwap.

3.4.1 FlashSystem A9000 and A9000R scalability

As with replication in general for the IBM Spectrum Accelerate Family, a single FlashSystem A9000 or A9000R initiator system can have up to 10 target systems. These target systems could be used for mirroring, Hyper-Scale Mobility (FlashSystem A9000 only) as well as for HyperSwap configurations. It is important to remember that HyperSwap can be configured only between two systems for any single volume or Consistency Group; however, a source system can have HyperSwap configured to multiple target systems using different volumes or Consistency Groups.

FlashSystem A9000 and A9000R systems have a maximum number of mirrors that can be configured. This number now includes pairs configured in a HyperSwap relationship due to using synchronous mirroring between the two systems. The maximum number of mirroring relations at the time of writing this document is 1,536. Refer to the latest document release notes for any updates to this number.

**Important:** Remember that each asynchronous mirror counts for three mirrors (primary volume snapshot, most recent snapshot and last replicated snapshot). If you have any combination of HyperSwap with asynchronous mirroring, be sure to plan accordingly. Refer to *IBM FlashSystem A9000 and A9000R Business Continuity Solutions*, REDP-5401 for more information about maximum mirrors.

3.4.2 Quorum Witness scalability

Although it is recommended to have the QW installed on a highly available virtual machine, there can be only a single QW defined as the active quorum for a system. This must also be the same QW defined on both storage systems configured for HyperSwap.

A single QW server can be defined on up to 12 FlashSystem A9000 and A9000R systems that will be running a HyperSwap configuration.

A second QW can be configured for each FlashSystem A9000 and A9000R that is part of a HyperSwap configuration, however this will only be used for redundancy purposes, as a backup system.

To maximize the protection, in a multi-site HA/DR deployment it is better that the three systems, and the Quorum witness be deployed in four completely-separate failure domains (typically, separate sites).
Chapter 4. HyperSwap implementation and usage

This chapter describes how to implement and use IBM HyperSwap for FlashSystem A9000 and FlashSystem A9000R. It covers the following topics:

- HyperSwap implementation process overview
- Quorum Witness setup
- Configuring the Quorum Witness on storage systems
- HyperSwap volume and Consistency Groups
- Checking HyperSwap status
- HyperSwap snapshots

Note: Most illustrations in this chapter are based on IBM Hyper-Scale Manager Version 5.5. Version 5.2 or later is required for system code level 12.1, which enables the HyperSwap feature.

Note that Version 5.5 is required for system code level 12.3 which enables the Multi-site feature. See Chapter 6, “Multi-site HA/DR implementation and usage” on page 89.
4.1 HyperSwap implementation process overview

To ensure smooth implementation, make sure to first review Chapter 3, “Prerequisites” on page 9.

Implementing HyperSwap consists of the following steps:
1. IBM FlashSystem A9000 and A9000R HyperSwap initial configuration starts with the installation of the external Quorum Witness application. The Quorum Witness is installed on a dedicated server or as a VMware virtual machine, preferably at a separate site. A separate site is better because it creates another, isolated failure domain.
2. The Quorum Witness is connected to both IBM FlashSystem A9000 or A9000R with volumes to be used in the HyperSwap solution. Each storage system can be configured with one Quorum Witness instance, identified by its unique ID.
3. The volumes from both systems used in HyperSwap relationship must be set as HyperSwap volumes.
4. HyperSwap volumes on each storage system must be mapped to the same application host server.

4.2 Quorum Witness setup

The server on which Quorum Witness (QW) application will be installed must be configured with all the latest updated and prerequisites.

Use the following command to install the required Linux packages:

```
yum install --y <Explicit RPM name> [...<Explicit RPM name>]
```

**Note:** The following example uses RHEL-7.3. For other versions of RHEL, refer to IBM Spectrum Accelerate Family HyperSwap Quorum Witness User Guide.

For RHEL-7 install the following RPMs:

- erlang-18.1-1.el7.centos.x86_64.rpm
- jemalloc-3.6.0-1.el7.x86_64.rpm
- nginx-1.10.3-1.el7.nginx.x86_64.rpm
- uuid-1.6.2-26.el7.x86_64.rpm
- postgresql92-libs-9.2.14-1PGDG.rhel7.x86_64.rpm
- postgresql92-9.2.14-1PGDG.rhel7.x86_64.rpm
- postgresql92-contrib-9.2.14-1PGDG.rhel7.x86_64.rpm
- postgresql92-server-9.2.14-1PGDG.rhel7.x86_64.rpm
- rabbitmq-server-3.6.0-1.noarch.rpm
- redis-3.2.5-1.el7.x86_64.rpm
- libxslt-1.1.28-5.el7.x86_64 (Only for CentOS)
- bzip2

**Tip:** RabbitMQ code can be downloaded from the RabbitMQ website.
4.2.1 Quorum Witness download and installation

Follow these steps to install QW:

1. Open TCP ports 8460 and 8641, as shown in Example 4-1. For retrieving nginx statistics, you also need to open TCP port 8462.

   **Note:** If you are using a different firewall software, refer to your firewall system documentation for specific instructions about how to open TCP ports.

   **Example 4-1  Opening the firewall ports on RHEL 7.3**

   [root@QW2 rhe17]# firewall-cmd --permanent --zone=trusted --add-interface=lo
   success
   [root@QW2 rhe17]# firewall-cmd --permanent --add-port=8460/tcp
   success
   [root@QW2 rhe17]# firewall-cmd --permanent --add-port=8461/tcp
   success
   [root@QW2 rhe17]# firewall-cmd --reload
   success

2. If the rabbitmq-server service is reported as not running, restart it, as shown in Example 4-2.

   **Example 4-2  Restarting rabbitmq-server**

   [root@QW2 ~]# systemctl restart rabbitmq-server
   [root@QW2 ~]# service rabbitmq-server status
   Status of node rabbit@QW2 ...
   (pid,18910),
   {running_applications,{{rabbit,"RabbitMQ","3.6.0"},
       {mnesia,"MNESIA CXC 138 12","4.14.3"},
       {os_mon,"CPO CXC 138 46","2.4.2"},
       {ranch,"Socket acceptor pool for TCP protocols.","1.2.1"},
       {xmerl,"XML parser","1.3.13"},
       {rabbit_common,[],"3.6.0"},
       {sasl,"SASL CXC 138 11","3.0.3"},
       {stdlib,"ERTS CXC 138 10","3.3"},
       {kernel,"ERTS CXC 138 10","5.2"}},
   {os,[unix,linux]},
   {erlang_version,"Erlang/OTP 19 [erts-8.3] [source] [64-bit] [async-threads:64]
    [hipe] [kernel-poll:true]\n"},
   {memory,{{total,41791760},
     {connection_readers,0},
     {connection_writers,0},
     {connection_channels,0},
     {connection_other,0},
     {queue_procs,2688},
     {queue_slave_procs,0},
     {plugins,0},
     {other_proc,19185560},
     {mnesia,57976},
     {mgmt_db,0},
     {msg_index,32376},
     {other_ets,863976},
     {...}}}
3. Download the installation package to a local folder on the Linux host that will be used as Quorum Witness server. You can download the installation package from IBM Fix Central.

4. Extract the installation package file using the following command:

```
# tar -xzvf ibm_quorum_witness-1.0.0-<build number>-x86_64.tar.gz
```

5. Verify that you have relevant permissions on the host.

6. Enter `./ibm_quorum_witness-1.0.0--<build number>.bin` to start the installation, as shown in Example 4-3. Note that you might need to install bzip2 if it is not already present on your system.

```
Example 4-3   Installing via the .bin file

[root@QW2 ~]# ./ibm_quorum_witness-1.0.0-1751-x86_64.bin
Verifying dependencies...
Starting installer, please wait...
International License Agreement for Non-Warranted Programs

Part 1 - General Terms
....
```

7. Review and accept the license agreement which is displayed after you run the installation file, as shown in Example 4-4.

```
Example 4-4   Accepting the license agreement

Press Enter to continue viewing the license agreement, or, Enter "1" to accept the agreement, "2" to decline it or "99" to go back to the previous screen, "3" Print, "4" Read non-IBM terms.

1

After installation, the Quorum Witness application starts automatically.
Refer to Example 4-5 for details.

Example 4-5  Troubleshooting ibm_quorum_witness restart problems

[root@QW2 ~]# systemctl status ibm_quorum_witness.service
    ibm_quorum_witness.service - Quorum Witness
      Loaded: loaded (/usr/lib/systemd/system/ibm_quorum_witness.service; enabled; vendor preset: disabled)
      Active: failed (Result: exit-code) since Wed 2017-04-26 11:55:30 CEST; 1min 36s ago
            Process: 20225 ExecStart=/opt/ibm/ibm_quorum_witness/bin/ibm_quorum_witness start (code=exited, status=4)
    Apr 26 11:55:30 QW2.local systemd[1]: Starting Quorum Witness...
    Apr 26 11:55:30 QW2.local ibm_quorum_witness[20225]: Error: nginx service is not running.
    Apr 26 11:55:30 QW2.local ibm_quorum_witness[20225]: Verify that SELinux is configured to allow Quorum Witness to bind to the TCP ports 8460 8461
    Apr 26 11:55:30 QW2.local ibm_quorum_witness[20225]: Refer to the user guide for details.
    Apr 26 11:55:30 QW2.local systemd[1]: ibm_quorum_witness.service: control process exited, code=exited status=4
    Apr 26 11:55:30 QW2.local systemd[1]: Failed to start Quorum Witness.
    Apr 26 11:55:30 QW2.local systemd[1]: ibm_quorum_witness.service entered failed state.
    Apr 26 11:55:30 QW2.local systemd[1]: ibm_quorum_witness.service failed.

[root@QW2 ~]# service nginx start
Redirecting to /bin/systemctl start nginx.service
[root@QW2 ~]# service nginx status
    nginx.service - nginx - high performance web server
      Loaded: loaded (/usr/lib/systemd/system/nginx.service; enabled; vendor preset: disabled)
      Active: active (running) since Wed 2017-04-26 12:04:55 CEST; 3s ago
            Docs: http://nginx.org/en/docs/
            Process: 20458 ExecStart=/usr/sbin/nginx -c /etc/nginx/nginx.conf (code=exited, status=0/SUCCESS)
            Process: 20457 ExecStartPre=/usr/sbin/nginx -t -c /etc/nginx/nginx.conf (code=exited, status=0/SUCCESS)
            Main PID: 20461 (nginx)
            CGroup: /system.slice/nginx.service

Note: If you are using SELinux, the program activation can fail. In this case allow nginx service to bind to network interfaces and connect to the QW socket.

setsebool -P nis_enabled 1 and semodule -i /opt/ibm/ibm_quorum_witness/conf.d/selinux/qw_rhel7.pp

Verify that the policy has been applied correctly by running:
cat /opt/ibm/ibm_quorum_witness/conf.d/selinux/qw_rhel7.te

Restart the QW service:
service ibm_quorum_witness restart
8. Copy the certificate file (qw.crt) from the
/opt.ibm/ibm_quorum_witness/settings/ssl_cert directory to the local workstation
where you currently use Hyper-Scale Manager. That certificate will later be added to both
A9000/A9000R systems participating in the HyperSwap relationship (see 4.3,
"Configuring the Quorum Witness on storage systems" on page 21).
4.3 Configuring the Quorum Witness on storage systems

After you install the Quorum Witness service on the Linux server, you must define it on the FlashSystem A9000 and A9000R that will be involved in the HyperSwap configuration. You must define volumes as HyperSwap volumes and map them to the host.

4.3.1 FlashSystem A9000 and A9000R HyperSwap configuration

Defining the Quorum Witness server on the storage systems can be done with the Hyper-Scale Manager GUI or using the XCLI.

Configuration using the Hyper-Scale Manager GUI

Complete the following steps for each storage system:
1. Navigate to the Systems view, as shown in Figure 4-1.

![Figure 4-1 Navigating to the Systems view](image)
2. Right-click the first FlashSystem A9000 or A9000R to be used in the HyperSwap relationship. From the drop-down menu, select **Quorum Witness** and then **Define Quorum Witness**, as shown in Figure 4-2.

![Figure 4-2 Define a Quorum Witness](image)

3. The System Quorum Witness panel displays, as shown in Figure 4-3. You must enter the Quorum Witness name, the IP address of the Quorum Witness server, and the port number (default is 8460).

![Figure 4-3 Quorum Witness definition form](image)
4. Select **Browse** to specify the mandatory Certificate file name, as shown in Figure 4-4.

```
Figure 4-4   Inputting QW definitions
```

5. Point to the QW Certificate file, as shown in Figure 4-5, and click **Open**.

```
Figure 4-5   Pointing to the qw.cert exported from the Quorum Witness
```

6. Click **Apply**, as shown in Figure 4-6.

```
Figure 4-6   Adding the certificate to the Quorum Witness
```
7. The panel then shows the **Activating** status as illustrated in Figure 4-7.

![Figure 4-7 Quorum Witness in Activating state](image1)

It finally goes to **Active** status, as shown in Figure 3-8.

![Figure 4-8 Quorum Witness in Active state](image2)

8. Repeat these steps on the second FlashSystem A9000 or A9000R.

**Important:** Make sure that you applied the FlashSystem A9000 or A9000R certificate to the QUORUM service. This process is normally the case on systems that shipped with software Version 12.1 or later. For systems that were upgraded from a lower version, use the `pki_update` command to add the QUORUM service. Refer to 3.1.1, “FlashSystem A9000 and A9000R prerequisites” on page 10 for more information.
For verification, follow these steps:

1. Go to the SYSTEMS & DOMAINS VIEWS, and then select **Quorum Witnesses**, as shown in Figure 4-9.

![Figure 4-9 Navigating to the Quorum Witnesses](image)

Active Quorum Witness Status should be **OK** and Quorum Witness should reflect the defined Quorum Witness Name, as shown in Figure 4-10.

![Figure 4-10 Showing Quorum Witness status](image)

To get the same information from the XCLI, use the `quorum_witness_list` command.

2. To attach target to QW, select the FlashSystem A9000 or A9000R to own the primary volume in the relationship. (You can change this option later by performing a Switch Role action.) On the right panel, select the Targets arm, as shown in Figure 4-11.

![Figure 4-11 Selecting the Targets arm](image)
3. In the target object window frame, select the menu box in the top right corner (the three small horizontal lines), and select **Attach to Quorum Witness**, as shown in Figure 4-12.

**Note:** It is necessary to run through the steps to attach the Quorum Witness.

4. Verify that the information presented in the panel is correct, and then click **Apply**, as shown in Figure 4-13.
Configuration using the XCLI

Refer to Example 4-6 for an illustration of how to configure the Quorum Witness by using the XCLI.

Example 4-6  XCLI Configuring the Quorum witness in XCLI

Primary System:
A9000>>quorum_witness_define address="9.155.117.24" port="8460" name="ITSO_QW1" certificate="-----BEGIN CERTIFICATE-----MIIC6jCCAdKgAwIBAgIJAIIwLcIiIwaoIANBgkqhkiG9w0BAQFAAOCAQ8MB0GA1UdDwEB/wQEAwIB  
activate="yes"
Command executed successfully.

Secondary System:
A9000R>>quorum_witness_define address="9.155.117.24" port="8460" name="ITSO_QW1" certificate="-----BEGIN CERTIFICATE-----MIIC6jCCAdKgAwIBAgIJAIIwLcIiIwaoIANBgkqhkiG9w0BAQFAAOCAQ8MB0GA1UdDwEB/wQEAwIB  
activate="yes"
Command executed successfully.

See Example 4-7 for an illustration of how to define Quorum Witness on the storage systems.

Example 4-7  Defining the Quorum Witness in XCLI

Primary System:
A9000>>target_add_quorum_witness quorum_witness="ITSO_QW1" target="A9000R"
Command executed successfully.

Secondary System:
A9000R>>target_add_quorum_witness quorum_witness="ITSO_QW1" target="A9000R"
Command executed successfully.
4.4 HyperSwap volume and Consistency Groups

This section discusses the creation, activation, and other operations relevant to HyperSwap volumes and Consistency Groups, such as converting synchronously mirrored volumes to HyperSwap volumes or vice versa.

4.4.1 HyperSwap volume creation and activation

You can create and activate HyperSwap volume from the GUI or the XCLI.

Using the GUI

To create a new HyperSwap volume, start by creating a new volume and assigning it to a pool, as usual. After you create the volume, follow these steps to make it a HyperSwap volume:

1. Navigate to the Volumes view in the Hyper-Scale Manager UI, for the FlashSystem A9000 or A9000R the new volume was created on.
2. Right-click the volume name and select Replication, and then Define/View Replication relation, as shown in Figure 4-14.

![Figure 4-14 Selecting volume to define HyperSwap](image-url)
3. Select the HyperSwap ADD button.

4. Select the system and the pool with the secondary volumes in that system, making sure to select **Activate on creation**, then click **Apply**, as shown in Figure 4-15.

![Figure 4-15  Volume HyperSwap definition](image)

Alternatively, you can use **offline init** to populate the volume before the mirror is activated, as shown in Figure 4-16.

![Figure 4-16  Offline Init configuration](image)
The GUI momentarily shows the I/O blocked on the system that owns the secondary volume. Shortly after, the volume shows as converted to HyperSwap, as illustrated in Figure 4-17.

![Figure 4-17 Volume HyperSwap completion](image)

**Using XCLI configuration**
Use the commands shown in Example 4-8 create and activate HyperSwap volumes, using the XCLI.

**Example 4-8 XCLI commands for HyperSwap volume definition**

```
A9000>>ha_create vol=ITSO_HA_VOLUME create_slave=yes remote_pool=ITSO_HA
target=A9000R

Command executed successfully.

A9000>>ha_activate vol=ITSO_HA_VOLUME

Command executed successfully.
```

### 4.4.2 Converting existing mapped volume to a HyperSwap volume

**Note:** The following illustrations are represented with a pre-existing volume (ITSO_Existing_vol).

The volume exists on one FlashSystem A9000 or A9000R system that is mapped to a host, and serving I/O, as shown in Figure 4-18.

![Figure 4-18 Volume with active IOPS](image)
The following steps are required to convert the volume to a HyperSwap volume:

1. Navigate to the Volumes view in the Hyper-Scale Manager GUI for the A9000 or A9000R on which the volume resides.

2. Right-click the volume name and select **HyperSwap** then **Define HyperSwap**, as shown in Figure 4-19.

![Figure 4-19 Defining HyperSwap](image)

3. In the panel displayed, select the FlashSystem A9000 or A9000R system that owns the secondary volume and the corresponding pool on that system, making sure to select **Activate on creation** and then click **Apply**, as shown in Figure 4-20.

![Figure 4-20 Defining the secondary volume](image)

**Note:** The synchronizing state as depicted in Figure 4-21 will take time to complete depending on the amount of data, as well as the number and speed of the mirroring connections.
The Availability Status will change to Synchronizing, as shown in Figure 4-21.

![Figure 4-21 Volume in Synchronizing state](image1)

When completed, the volume will show as a HyperSwap volume, as shown in Figure 4-22.

![Figure 4-22 Volume synchronization complete](image2)

4. Map the secondary volume created on the target system to the same host on which the primary data resides.

Steps 1 through 4 can also be completed from the XCLI, as shown in Example 4-9.

**Example 4-9  XCLI commands for these steps**

```
A9000>>ha_create vol="ITSO_Existing_vol" remote_pool="ITSO_HA" create_slave="yes" target="A9000R"
A9000>>ha_activate vol="ITSO_Existing_vol" target="A9000R"
A9000R>>map_vol vol="ITSO_Existing_vol" lun="1" host="ITSO_W2K12R2_01"
```
5. Use the `xiv_fc_admin -R` or `xiv_iscsi-admin -R` command, to perform a Fibre Channel or iSCSI rescan on the host to discover the additional new paths.

Notice the HyperSwap column indicating the system identified the volume as a HyperSwap volume, as shown in Figure 4-23.

![Figure 4-23 Hyperswap indicator is set to Yes for the volume](image)

6. Identify the connected paths as well as which are optimized and which are unoptimized using the `mpclaim -s -d` command.

**Note:** Example 4-10 on page 33 assumes a Windows host. Use the MPIO Disk number (xxx) in the `mpclaim -s -d xxx` command.

**Example 4-10 Determining optimized and unoptimized paths**

C:\Users\Administrator>mpclaim -s -d

For more information about a particular disk, use 'mpclaim -s -d # where # is the MPIO disk number.

**Example 4-10 Determining optimized and unoptimized paths**

```
MPIO Disk    System Disk  LB Policy    DSM Name
-------------------------------------------------------------------------------
MPIO Disk3   Disk 4       RRWS         Microsoft DSM

C:\Users\Administrator>mpclaim -s -d 3

MPIO Disk3: 08 Paths, Round Robin with Subset, Implicit Only
Controlling DSM: Microsoft DSM
SN: 6001738CCCE0567300000036288
Supported Load Balance Policies: FOO RRWS LQD WP LB
Path ID          State              SCSI Address      Weight
-----------------------------------------------------------------------
0000000077070003 Active/Unoptimized 007|000|003|001   0
TPG_State : Active/Unoptimized, TPG_Id: 1, : 4097

0000000077060003 Active/Unoptimized 006|000|003|001   0
TPG_State : Active/Unoptimized, TPG_Id: 1, : 4097

0000000077070002 Active/Unoptimized 007|000|002|001   0
* TPG_State : Active/Unoptimized , TPG_Id: 0, : 768

0000000077070001 Active/Unoptimized 007|000|001|001   0
* TPG_State : Active/Unoptimized , TPG_Id: 0, : 512
```
4.4.3 Convert a synchronously mirrored volume to HyperSwap volume

To convert a mirrored volume into a HyperSwap volume, the secondary volume of the mirror relationship has to be unmapped. After you unmap the secondary volume in the relationship, use the following steps to convert the mirrored volume to a HyperSwap volume:

**Note:** The following example illustrations were completed using a volume (ITSO_MIRROR_VOLUME) mapped to a Windows 2012 R2 host with I/O running. *Only synchronously* mirrored volumes can be converted to a HyperSwap relationship.

1. Navigate to the **Remote Views** view and select **Replication Details**, as shown in Figure 4-24.

2. Highlight the volume listed as primary in the mirror relationship, as shown in Figure 4-25.
3. Right click and select **Convert Type → Convert to HyperSwap**, as shown in Figure 4-26.

![Figure 4-26 Converting to HyperSwap](image)

4. Click **Apply**, as shown in Figure 4-27.

![Figure 4-27 Applying the conversion](image)

5. After the process completes successfully, the volumes are in a HyperSwap relationship, as shown in Figure 4-28.

![Figure 4-28 Replication updated to HyperSwap](image)

Figure 4-29 shows the relationship displayed in the Replication Details view with HyperSwap indicated under the High Availability column.

![Figure 4-29 Volume view now shows HyperSwap relationship](image)

For an example of how to convert a mirrored volume into a HyperSwap volume using the XCLI, refer to Example 4-11.
Example 4-11  XCLI to convert mirrored relation to HyperSwap

A9000>mirror_convert_into_ha vol="ITSO_MIRROR_VOLUME"

A9000R>>map_vol vol="ITSO_MIRROR_VOLUME" lun="2" host="ITSO_W2K12R2_01"

6. Map the secondary volume to the host and rescan as seen in Example 4-12.

Example 4-12  Rescanning the host after conversion

C:\Users\Administrator>xiv_fc_admin -R

4.4.4 Convert a HyperSwap relationship to a synchronous mirror relationship

A HyperSwap volume relationship can be converted to a synchronous mirror relationship. The following steps can be taken to make the conversion:

1. Unmap the secondary copy of the volume from the host.
2. In the Volumes view, select the primary copy of the volume, or the copy that is still mapped to the host.
3. Right click the primary volume and select HyperSwap then Convert HyperSwap to Mirror, as shown in Figure 4-30.

Note: The following example illustrations were completed using a volume (ITSO_MIRROR_VOLUME) which was mapped as a HyperSwap volume to a Windows 2012 R2 host with I/O running.

HyperSwap relationships can only be converted to a synchronous mirroring relationship.

Figure 4-30  Selecting Convert HyperSwap to Mirror option
4. Click **Apply**, as shown in Figure 4-31.

![Figure 4-31  Applying the conversion](image)

The Hub now reflects that the volume was successfully converted to a mirrored volume, as shown in Example 4-32.

![Figure 4-32  Verification that volume is now a Sync Mirror](image)

To convert a HyperSwap relationship to a synchronous mirror relationship in XCLI, refer to Example 4-13.

**Example 4-13  XCLI commands to convert from a HyperSwap relation to a mirrored relation**

```
A9000R>>unmap_vol vol="ITSO_MIRROR_VOLUME" host="ITSO_W2K12R2_01"

A9000>>ha_convert_into_mirror vol="ITSO_MIRROR_VOLUME"
```
4.4.5 Define a Consistency Group as a HyperSwap Consistency Group

HyperSwap Consistency Groups (CG) uses one CG on the source system and one CG on the destination (target) system. Both CGs must be empty, meaning they cannot contain any volume, before defining the CG as a HyperSwap CG.

The following steps can be performed to create a HyperSwap CG:

1. Create a CG on both storage systems.
2. Right click the primary Consistency Group and select HyperSwap → Define/View HyperSwap, as shown in Figure 4-33.

![Figure 4-33 Defining a new CG as a HyperSwap CG](image)

3. Define the secondary Consistency Group on the other system and click Apply, as shown in Figure 4-34.

![Figure 4-34 Defining the secondary system and CG](image)
The Hub shows that the CG is now a HyperSwap CG as illustrated in Figure 4-35.

![Figure 4-35  CG shows as a HyperSwap CG](image)

4. Navigate to the **Volumes** View, select the volumes, and right-click the list of volumes to be added to the CG. Select **Consistency Group** then **Move to Group**, as shown in Figure 4-36.

   **Note:** The volume or volumes to be added to the CG must already be a HyperSwap volume or volumes.

![Figure 4-36  Selecting volumes to add to the HyperSwap CG](image)

5. Select the appropriate CG, and click **Apply**, as shown in Figure 4-37.

![Figure 4-37  Selecting the CG to add the volumes to](image)
The CG now shows as Synchronized and displays an Availability status of HyperSwap, as shown in Figure 4-38.

![Figure 4-38  CG shows Synchronized and HyperSwap](image)

To configure a HyperSwap CG in XCLI, refer to Example 4-14.

**Example 4-14  XCLI commands for creating a HyperSwap Consistency Group**

```bash
A9000>>ha_activate cg="ITSO_HA_CG_P" target="A9000R"
A9000>>ha_create slave_cg="ITSO_HA_CG_S" cg="ITSO_HA_CG_P" target="A9000R"
A9000>>ha_create vol="ITSO_HA_VOLUME1" remote_pool="ITSO_HA_POOL1" create_slave="yes" target="A9000R"
A9000>>ha_activate vol="ITSO_HA_VOLUME1" target="A9000R"
A9000>>ha_create vol="ITSO_HA_VOLUME2" remote_pool="ITSO_HA_POOL1" create_slave="yes" target="A9000R"
A9000>>ha_activate vol="ITSO_HA_VOLUME2" target="A9000R"
A9000>>cg_add_vol vol="ITSO_HA_VOLUME2" cg="ITSO_HA_CG_P"
A9000>>cg_add_vol vol="ITSO_HA_VOLUME1" cg="ITSO_HA_CG_P"
```

4.4.6 Convert a synchronous mirror CG to a HyperSwap CG

Only synchronously mirrored CGs can be converted to a HyperSwap CG. The following steps are necessary to convert a Sync CG to a HyperSwap CG:

1. Navigate to the **Remote Views** view and select **Replication Details**, as shown in Figure 4-39.

![Figure 4-39  Navigating to the Replication Details view](image)
2. In the Replication Details view, select the primary CG. Right-click the primary CG (or under the Actions menu once the CG is selected) and select Convert Type, then Convert Sync Mirror to HyperSwap, as shown in Figure 4-40.

![Figure 4-40 Selecting convert to HyperSwap option](image)

3. Click Apply, as shown in Figure 4-41.

![Figure 4-41 Applying the conversion of the CG to a HyperSwap CG](image)

The Hub now shows that the CG was successfully converted to a HyperSwap CG, as illustrated in Figure 4-42.

![Figure 4-42 CG successfully converted](image)

To convert a synchronous mirror CG to a HyperSwap CG in XCLI, refer to Example 4-15.

**Example 4-15 XCLI commands to convert a Sync CG to a HyperSwap CG**

```
A9000>> mirror_convert_into_ha cg="ITSO_HA(CG)P"
```
4.4.7 Convert a HyperSwap CG to a synchronous mirror CG

HyperSwap CGs can only be converted to a synchronously mirrored CG. The following steps can be used to perform the conversion:

1. Navigate to the Remote Views view and select Replication Details, as shown in Figure 4-43.

   ![Figure 4-43 Navigating to the Replication Details view](image)

2. In the Consistency Groups view, select the primary CG. Right click the primary CG and select HyperSwap then Convert HyperSwap to Sync Mirror, as shown in Figure 4-44.

   ![Figure 4-44 Selecting Convert HyperSwap to Sync Mirror](image)

3. Click Apply, as shown in Figure 4-45.

   ![Figure 4-45 Applying the CG conversion](image)

The Hub reflects that the HyperSwap CG was successfully converted to a Sync Mirrored CG, as shown in Figure 4-46.
4.5 Checking HyperSwap status

The current status of the QW, as well as the status of the HyperSwap volumes and CGs can be viewed both through the XCLI and the GUI.

To view the QW status through the XCLI use the `quorum_witness_list` command, as depicted in Example 4-17.

**Example 4-17  Viewing the QW list**

<table>
<thead>
<tr>
<th>Name</th>
<th>ID</th>
<th>Address</th>
<th>Port</th>
<th>State</th>
<th>Connection</th>
<th>External Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Db</td>
<td>Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITSO_QW1</td>
<td>OK 11bb74a4f4654459a6c510f2e57c8813</td>
<td>9.155.117.24</td>
<td>8460</td>
<td>Activated</td>
<td>Up</td>
<td>myquorum</td>
</tr>
</tbody>
</table>

Secondary System:

A9000R>>quorum_witness_list

<table>
<thead>
<tr>
<th>Name</th>
<th>ID</th>
<th>Address</th>
<th>Port</th>
<th>State</th>
<th>Connection</th>
<th>External Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Db</td>
<td>Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITSO_QW1</td>
<td>OK 11bb74a4f4654459a6c510f2e57c8813</td>
<td>9.155.117.24</td>
<td>8460</td>
<td>Activated</td>
<td>Up</td>
<td>myquorum</td>
</tr>
</tbody>
</table>

To view the HyperSwap volumes and CGs through the XCLI use the `ha_list` command, as illustrated in Example 4-18.

**Example 4-18  Viewing HyperSwap volumes and CGs**
### Primary System: A9000

<table>
<thead>
<tr>
<th>Name</th>
<th>HA Object</th>
<th>Role</th>
<th>Remote System</th>
<th>Active</th>
<th>Status</th>
<th>Link Up</th>
<th>Automatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITSO_HA_VOLUME</td>
<td>Volume</td>
<td>Master</td>
<td>A9000R</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>N/A</td>
</tr>
<tr>
<td>ITSO_HA_AIX_DHE_001</td>
<td>Volume</td>
<td>Slave</td>
<td>A9000R</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>Active</td>
</tr>
<tr>
<td>ITSO_HA_AIX_DHE_002</td>
<td>Volume</td>
<td>Master</td>
<td>A9000R</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>N/A</td>
</tr>
<tr>
<td>ITSO_Existing_vol</td>
<td>Volume</td>
<td>Master</td>
<td>A9000R</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>N/A</td>
</tr>
<tr>
<td>ITSO_HA_VMware</td>
<td>Volume</td>
<td>Master</td>
<td>A9000R</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>N/A</td>
</tr>
<tr>
<td>ITSO_HA_CG_P</td>
<td>CG</td>
<td>Master</td>
<td>A9000R</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>N/A</td>
</tr>
<tr>
<td>ITSO_HA_VOLUME1</td>
<td>Volume</td>
<td>Master</td>
<td>A9000R</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>N/A</td>
</tr>
<tr>
<td>ITSO_HA_VOLUME2</td>
<td>Volume</td>
<td>Master</td>
<td>A9000R</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>N/A</td>
</tr>
</tbody>
</table>

### Secondary System: A9000R

<table>
<thead>
<tr>
<th>Name</th>
<th>HA Object</th>
<th>Role</th>
<th>Remote System</th>
<th>Active</th>
<th>Status</th>
<th>Link Up</th>
<th>Automatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITSO_HA_VOLUME</td>
<td>Volume</td>
<td>Slave</td>
<td>A9000</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>Active</td>
</tr>
<tr>
<td>ITSO_HA_AIX_DHE_001</td>
<td>Volume</td>
<td>Master</td>
<td>A9000</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>N/A</td>
</tr>
<tr>
<td>ITSO_HA_AIX_DHE_002</td>
<td>Volume</td>
<td>Slave</td>
<td>A9000</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>Active</td>
</tr>
<tr>
<td>ITSO_Existing_vol</td>
<td>Volume</td>
<td>Slave</td>
<td>A9000</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>Active</td>
</tr>
<tr>
<td>ITSO_HA_VMware</td>
<td>Volume</td>
<td>Slave</td>
<td>A9000</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>Active</td>
</tr>
<tr>
<td>ITSO_HA_CG_S</td>
<td>CG</td>
<td>Slave</td>
<td>A9000</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>Active</td>
</tr>
<tr>
<td>ITSO_HA_VOLUME1</td>
<td>Volume</td>
<td>Slave</td>
<td>A9000</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>N/A</td>
</tr>
<tr>
<td>ITSO_HA_VOLUME2</td>
<td>Volume</td>
<td>Slave</td>
<td>A9000</td>
<td>yes</td>
<td>Synchronized</td>
<td>yes</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Using the Hyper-Scale Manager GUI, the volumes and CGs can be viewed by navigating to the Remote Views view and selecting Replication Details, as shown in Figure 4-47.
4.6 HyperSwap snapshots

As is the case with synchronous mirroring, there are three types of snapshots that can be created on a HyperSwap volume or Consistency Group: local, mirrored, and internal snapshot also designated as Last Consistent Snapshot (LCS).

Local snapshots on a HyperSwap volume behave exactly like they do for a regular volume. There are no special limitations; For example it is possible to create a snapshot on the secondary volume, even when that volume is not synchronized.

Mirrored snapshots for volumes in a HyperSwap relationship are created using the `ha_create_snapshot` command. This command is similar to `mirror_create_snapshot` and has similar behavior and limitations. Both commands create an equivalent snapshot on both peers at the same point-in-time, meaning that the contents of the two snapshots at the time of creation are identical. Those snapshots can be created only when the peers are connected and synchronized.

After you create the snapshots, they are independent and behave like local snapshots. On each peer, the snapshot can be unlocked, updated, deleted, and restored independently. Also snapshot deletion as a result of capacity exhaustion is independent for each peer.

The LCS is a special snapshot created on a consistent secondary volume before a resynchronization is started.

The advantage of creating mirrored snapshots for HyperSwap is that it allows you to restart a temporarily removed HyperSwap relationship, by restoring the HyperSwap volumes from their snapshots, by using as HyperSwap/Mirror with Offline Init option. The synchronization process will only need to transfer data that is different between primary and secondary, rather than initially transferring the whole primary volume contents.

**Note:** For general use of snapshots, refer to *IBM Hyper-Scale Manager for IBM Spectrum Accelerate Family: IBM XIV, IBM FlashSystem A9000 and A9000R, and IBM Spectrum Accelerate*, SG24-8376.

4.6.1 Creating HyperSwap snapshots

To create HyperSwap snapshots using the Hyper-Scale Manager GUI, proceed as follows:

1. Navigate to the **REMOTE VIEWS** view and select **Replication Details**, as shown in Figure 4-48.

![Figure 4-48](image)
2. Right-click the primary HyperSwap volume and select **Snapshot → Create Replicated Snapshot**, as shown in Figure 4-49.

![Create HyperSwap snapshot for HyperSwap volumes](image1)

*Figure 4-49  Create HyperSwap snapshot for HyperSwap volumes*

3. Optionally, change the names of snapshots by pressing the snapshot name and edit, as shown in Figure 4-50, and then click **Apply**.

```
Note: The default names of both snapshots are <name_of_volume>.mirror_snapshot_.
```

![Create HyperSwap snapshot box view](image2)

*Figure 4-50  Create HyperSwap snapshot box view*

4. After the process completes successfully, the snapshots are shown in **POOLS & VOLUMES VIEWS → Snapshots View**, as shown in Figure 4-51.

![Snapshots view](image3)

*Figure 4-51  Snapshots view*

When using XCLI, open an XCLI session for the system that contains the desired HyperSwap primary volume and run the following command:
**ha_create_snapshot** vol=<volume_to_snapshot> delete_priority="<1 for Last deleted until 4 for First>" name="<name_of_snapshot_for_primary>"

Refer to Example 4-19 for an illustration.

**Example 4-19  Create HyperSwap snapshot in xcli mode**

```
A9000R>>ha_create_snapshot vol=ITSO_HA_VOLUME delete_priority="4"
name="ITSO_HA_VOLUME.my_xcli_mirrored_snapshot"
Command executed successfully.
A9000R>>
```

The snapshot can be shown using **snapshot_list vol=<name_of_vol>** as illustrated in Figure 4-52.

```
<table>
<thead>
<tr>
<th>Name</th>
<th>Size (GB)</th>
<th>Master Name</th>
<th>Consistency Group</th>
<th>Pool</th>
<th>Creator</th>
<th>Written (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITSO_HA_VOLUME_mirror_snapshot_Primary</td>
<td>103</td>
<td>ITSO_HA_VOLUME</td>
<td>ITSO_HA_POOL</td>
<td>admin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITSO_HA_VOLUME_snapshot</td>
<td>103</td>
<td>ITSO_HA_VOLUME</td>
<td>ITSO_HA_POOL</td>
<td>admin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITSO_HA_VOLUME.my_xcli_snapshot</td>
<td>103</td>
<td>ITSO_HA_VOLUME</td>
<td>ITSO_HA_POOL</td>
<td>admin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITSO_HA_VOLUME.my_xcli_mirrored_snapshot</td>
<td>103</td>
<td>ITSO_HA_VOLUME</td>
<td>ITSO_HA_POOL</td>
<td>admin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

**Figure 4-52  Snapshot list in XCLI mode**

**Note:** For HyperSwap Consistency Group snapshots, perform the same operations for the desired Consistency Group, rather than for a volume.

Refer also to Chapter 10, “Volume Shadow Copy Service and HyperSwap Snapshots” on page 189 for a discussion on combining Microsoft Virtual Shadow Copy Services (VSS) with Snapshots.
HyperSwap scenarios

This chapter describes different failure scenarios for IBM FlashSystem A9000 and A9000R with volumes or Consistency Groups in a HyperSwap relationship. The following scenarios are discussed:

- Failures on the storage system serving the primary copy
- Connection failure between host and storage systems
- Quorum Witness failure scenarios
- Failback scenarios

The material in this chapter, is based on an environment that uses one FlashSystem A9000 and one FlashSystem A9000R. The scenarios assume that the primary copy of a HyperSwap volume is on System A at Site A, as shown in Figure 5-1 on page 50.
5.1 Failures on the storage system serving the primary copy

This section discusses the following failure scenarios on the storage system serving I/Os for the primary volume:

- System A is down
- Data service failure on System A, such as no storage space left for writing on System A

Figure 5-1 shows System A (A9000) and System B (A9000R) with one volume in HyperSwap relationship. The host and the primary copy of the volume are at Site A.

![Diagram of IBM FlashSystem HyperSwap Volume](image-url)
5.1.1 System A failure scenario

Figure 5-2 depicts a failure of System A at Site A, which could result from a loss of power affecting System A:

- Host A at Site A and System B (A9000R) at site B have lost their Fibre Channel (FC) connections to System A at Site A.
- The Quorum Witness at site Q has lost its Ethernet connection to System A at Site A.

Host A has still access to the HyperSwap volume because all read and write I/Os can still be served from System B (A9000R). System B has no FC connection to System A and the Quorum Witness has no Ethernet connection to System A.
As a result of HyperSwap automatic failover, System B now serves read and write I/O to host A and volumes on System B take on the primary role, as can be seen from the Hyper-Scale Manager GUI shown in Figure 5-3.

After System A becomes operational again, all paths and connections are reestablished.

At that point, the HyperSwap volume is not synchronized. The HyperSwap volume ITSO_RH_HA_004 has its primary copy on System B (A9000R). I/O is served to the host from System B. The copy of this HyperSwap volume on System A (A9000) is in I/O blocked state because it is not synchronized with the copy of this HyperSwap volume on System B.

On System A, the copy of this HyperSwap volume is still in primary role because at the time System B started serving I/O, System A was down and did not get the information about the new role, as shown in Figure 5-4.

Section 5.4, “Failback scenarios” on page 76 describes the process to recover the I/O blocked volumes on System A and to get back to the Automatic Failover state.
5.1.2 Data service failure

This section illustrates the data services failure scenario. This type of failure can occur when no space is left on System A for writing a physical block. This situation is known as an out of partitions (OOP) situation.

In an OOP situation, the HyperSwap relationship will fail over to Site B and I/O will be served from System B at Site B as described in 5.1.1, “System A failure scenario” on page 51.

Figure 5-5 shows the dashboard information of System A in the OOP state.

![Figure 5-5 Dashboard OOP information](image)

5.2 Connection failure between host and storage systems

This section illustrates connection failure scenarios, either a Fibre Channel (FC) connection loss from System A or loss of the Ethernet connection to the Quorum Witness (QW). Figure 5-1 shows two IBM FlashSystem A9000 and A9000R with a volume in a HyperSwap relation. The hosts and the primary copy of the volume are at site A. The primary copy is serving I/O to host A.

The different scenarios are:

- Failure of Fibre Channel path between the host and the storage system
- Loss of the FC connection between System A and the host, then loss of the FC connection between System A and System B
5.2.1 FC path failure between host and storage system

This scenario considers a FC connectivity loss between the Host A and System A at site A as illustrated in Figure 5-6.

![Diagram of FC connection failure between Host A and System A](image)

*Figure 5-6  FC connection failure between Host A and System A*
Figure 5-7, shows the Hyper-Scale Manager GUI representation of the host connection with both System A (A9000) and System B (A9000R). To get to this view, select **Hosts & Clusters Views → Hosts**, then selected the corresponding host and go to **Properties → Connectivity → View Connectivity Details**.

**Figure 5-7  Host Connectivity when connected to both storage systems.**

Figure 5-8, shows the Hyper-Scale Manager GUI view when connectivity between Host A and System A (A9000) is broken.

**Figure 5-8  FC connection loss between Host and A9000**
When FC connectivity between Host A and System A (owning the primary volume) is broken, Host A still has access to the HyperSwap volume because read and write I/Os can still be served from System B, as shown in Figure 5-9.

The data synchronization between System B and System A, before a destage of data to storage, happens in the following sequence:

1. Host A sends write I/O to System B.
2. System B sends writes to System A.
3. System A caches the data and syncs the data with System B.
4. System B sends an acknowledgement of the writes to System A.
5. System A sends an acknowledgement to System B.
6. System B sends an acknowledgement to the host.

![Figure 5-9](image.png)

*Figure 5-9  Host now reads or writes data to System B which syncs data with System A*

When host connectivity with System A comes back, host I/Os are automatically served from System A without any additional process required to synchronize data between System A and System B.
5.2.2 FC connectivity loss between host and System A and between System A and System B

Figure 5-10 depicts a connectivity failure between Host A and System A (A9000), along with a connectivity failure between System A (A9000) and System B (A9000R), which also results in replication failure between System A and System B.

![Diagram showing connectivity failure between Host A and System A and between System A and System B.]

Figure 5-10  Connectivity failure between Host A and System A and between System A and System B
Figure 5-11 shows the Hyper-Scale Manager GUI representation of active FC connectivity between System A and System B. To get to this view, select Systems & Domains Views, then select the storage system and go to System Connectivity → Targets → View Connectivity → Connectivity Details.

The connectivity pane is displayed, as shown on the left side in Figure 5-11. The view on the right side of Figure 5-11 shows the GUI representation of inactive connectivity between System A and System B when all the links are down.

The following sequence of events occurs when connectivity between Host A and System A as well as connectivity between System A and System B has failed:

1. Host A tries to divert I/O traffic to System B.
2. Quorum Witness is still up and running, so System A and System B poll the Quorum Witness to check if the other storage system (System B or System A) is up and running.
3. Because Quorum Witness can still talk to both System A and System B, it keeps the volume on System A in primary role and places System B in I/O blocked status, as shown in Figure 5-12.

Now when the connectivity between Host A and System A is restored, a manual failback can be done to System A and I/Os can be served again from System A.
Figure 5-13 shows the GUI view when the connectivity between Host A and System A is restored while the connectivity between System A and System B is still not restored.

![Figure 5-13](image)

**Figure 5-13** Host Connectivity with System A restored: IOPS happening to System A

When the connectivity between System A and System B is restored, data synchronization between System A and System B occurs first and HyperSwap is then re-enabled.

Figure 5-14 shows the GUI view when the connectivity between System A and System B has been restored and data is being synchronized between both storage systems.

![Figure 5-14](image)

**Figure 5-14** Connectivity between System A and System B is established
After the volume synchronization has completed, the GUI shows that the high availability is reestablished and an automatic failover is now possible again, as shown in Figure 5-15.

![HyperSwap GUI showing the high availability is reestablished and automatic failover is now possible again](image)

**Figure 5-15  Automatic failover re-established**

### 5.3 Quorum Witness failure scenarios

This section covers several failure scenarios related to the Quorum Witness (QW) availability:

- QW failure
- Loss of the Ethernet connection between System A and the QW
- QW connectivity failure to System A, followed by mirroring link failure
- QW down, restore from backup is possible
- QW down, new installation is needed

The host and the primary copy of the volume are at site A. The primary copy is serving I/O to host A. The QW on the third site is connected to System A on site A and System B on site B.
### 5.3.1 Quorum Witness failure

Figure 5-16 shows a QW failure scenario that might occur as a result of the QW server issue, or network problems, or even disaster at Site Q where the QW server resides.

When a QW is no longer available, then the HyperSwap automatic failover capability is disabled. However, because connectivity between System A and System B is normally still in place, replication between Site A and Site B is still operational and a manual failover (switch role) between System A and System B volumes remains possible.

![Figure 5-16 Quorum Witness failure](image)

Figure 5-17 is the Hyper-Scale Manager GUI view indicating that QW connections to both System A (A9000) and System B (A9000R) are down.

To get to this view, select **Systems & Domains Views → Quorum Witnesses**.

![Figure 5-17 Quorum Witness down](image)
Figure 5-18 shows the Hyper-Scale manager GUI view of the Quorum witness connectivity status on System A.

Go to Remote Views → Mirrored/HyperSwap Volumes (Availability) and select the concerned volume to view the Volume Availability details, as illustrated in Figure 5-19. Notice that it shows only QW as not available but System A and System B are still connected and replication between them is active and synchronized.
As shown in Figure 5-20, I/O is served from System A (A9000) and the data is being replicated to System B (A9000R).

Figure 5-20  Hyper-Scale Manager GUI view of I/Os being served and replicated

Although automatic failover is disabled in case of a QW failure, it is still possible to do a manual failover using the Switch Roles option, as illustrated in Figure 5-21.

Figure 5-21  Switching roles of Storage systems in HyperSwap manually
After you select the **Switch Roles** option, confirm the action as illustrated in Figure 5-22. Then click **Apply**.

![Switch Roles](image)

Figure 5-22  Prompt to confirm role switch for HyperSwap (Primary to Secondary)

When the switch role option is applied, system B (A9000R) is now serving I/Os to the host, as shown in Figure 5-23.

![Site B serving the I/O](image)

Figure 5-23  Site B serving the I/O
Figure 5-24 shows that after the switch role is applied, I/Os from the host would be served from System B and the data would now be synchronously replicated to System A.

When the QW starts working again and connects with the System A and System B, the automatic failover capability of the HyperSwap function is restored.

### 5.3.2 Ethernet connection loss between system A and QW

A loss of communication between a storage system and the Quorum witness affects the health of the HyperSwap relationships stemming from and to that system, and compromise their normal operation in case of subsequent failures.

Automatic failover is compromised depending on the role of System A in the HyperSwap relationship. If A is primary, then the HyperSwap relationship is in “orange” state, meaning that automatic failover is still possible in case of failure of A or loss of communication between system A and system B (system B, will hold the quorum at that point). On the other hand, if A has secondary role in the relationship, then the HyperSwap state turns to “red”, to indicate that the automatic failover is not possible (if B were to fail, A would not hold the Quorum and therefore could not automatically assume the Primary role). Manual failover would still be possible.

Both scenarios are shown: In the two HyperSwap relationships shown in Figure 5-25, System A is designated respectively as Primary (Figure 5-26) and Secondary (Figure 5-27 on page 67).
Chapter 5. HyperSwap scenarios

When System A loses connection to the Quorum Witness, its status is shown as Down in the Systems & Domains Views → Quorum Witnesses panel (Figure 5-28) and details (Figure 5-29).
Corresponding entries are logged in the event list, reachable by the Systems & Domains Views → Events, as shown in Figure 5-30.

The status of the HyperSwap relationships changes to reflect the failure, as shown in Figure 5-31. Notice the difference of behavior in the relationship where System A has the Primary role (Figure 5-32, Figure 5-33 on page 69) versus the one where it has Secondary role (Figure 5-34 on page 69, Figure 5-35 on page 69).
Chapter 5. HyperSwap scenarios

Figure 5-33  In the relationship where System A has Primary role, Automatic Failover reports potential problems, as shown in the relationship details

Figure 5-34  Status of the relationship where System A has Secondary role

Figure 5-35  In the relationship where System A has Secondary role, automatic failover is disabled

In this last case, however, a manual failover is still possible if needed, and can be performed by selecting Failover/Recovery Switch Roles in the relationship Actions menu, as shown in Figure 5-36.
After confirming the role switch, the relationship enters the state depicted in Figure 5-37 on page 70, where Automatic Failover capability is now restored (as System B is now Secondary and therefore able to hold the QW and become Primary in case of failure on System A).

When the connection between System A and the Quorum Witness is restored, all the HyperSwap relationships become healthy again, and both automatic and manual failover are possible.

**5.3.3 QW connectivity failure to System A, followed by mirroring link failure**

Figure 5-38 depicts a connectivity failure between Quorum Witness and System A (A9000) along with a failure in the mirroring links between System A and System B. This scenario describes what happens when each failure occurs in succession (what is often described as a **rolling disaster**).

We assume that the HyperSwap configuration was fine prior to the failures. The failure affecting the QW connectivity to the System A could occur due to a hardware failure on the
server (the Ethernet card failed), a problem in the network preventing the QW from communicating with the system or potentially a failure on the system itself that affects the Ethernet connectivity.

Figure 5-38  QW connectivity failure with failed mirroring links

This scenario does not impact host I/O to System A and System B as this is a Fibre Channel connection. However, depending on how the mirroring links failed, the host could also be affected with failed paths. In this example, we assume that the host paths are not affected to either system when the mirroring links fail.

As described in the previous section, upon failure of the Quorum Witness connection different behaviors occur in the HyperSwap relationship depending on the role of System A. Figure 5-39 shows the same examples, with two relationships in which System A has a different role.

Figure 5-39  Two HyperSwap relationships between System A and System B, in which System A has different designated roles.

After System A loses its connection to the QW, the state of the two relationships changes as already shown (Figure 5-40 on page 72):
Figure 5-40   State of the two HyperSwap relationships after the connectivity failure between System A and the Quorum Witness

When the connectivity between System A and System B is interrupted, the situation illustrated in Figure 5-41 occurs.

Figure 5-41   State of the two HyperSwap relationships after the connectivity failure between System A and System B

In the HyperSwap relationship where System A had the Primary role, an automatic failover was performed (because System B holding the QW). System A, missing connectivity with both the QW and its partner system, found itself in compromised state and therefore blocked IO on its copy of the volume. The HyperSwap relationship enters the compromised state, but host IOs are still served through System B, as shown in Figure 5-42.

Figure 5-42   System B assumes Primary role after the Automatic Failover; IOs are blocked on System A for this volume

In the HyperSwap relationship where System A had the Secondary role, System A found itself isolated from both its partner system and the QW, therefore blocking IO on its copy of the volume to avoid possible data corruption; System B, recognizing the failure, held its Primary role, continuing to serve host IOs (Figure 5-43 on page 73).
In the other relationship, where System A was Designated Primary but System B took over the Primary role after the Automatic Failover, there is now a role conflict (Figure 5-46 and Figure 5-47 on page 74):

When all failures are subsequently fixed, again different behaviors occur. The relationship where System B was Primary and System A was Secondary automatically enters the Healthy state (after a brief resynchronization window), as shown in Figure 5-44 and Figure 5-45:
Figure 5-47  Relationship details shows the role conflict after the connectivity recovery

Note that IOs are still blocked on the copy of the volume on System A, therefore data is not being compromised.

To return to the original state, we first need to resolve the role conflict; since B has served IOs during the outage, we need to bring back the modified data to System A. Therefore, the first step is to temporarily change the role of System A to Secondary, through the relationships Actions menu (Failover/Recovery → Change Role), as shown in Figure 5-48:

Figure 5-48  Performing a manual change role via the relationship’s Actions menu

Confirm the role change (Figure 5-49 on page 74), then activate the relationship to start resynchronizing data from System B to System A (Figure 5-50 on page 75).

Figure 5-49  Confirming the manual role change
When the relationship is synchronized, it becomes healthy again (Figure 5-51). At this point, we can perform a manual role switch to restore the initial configuration where System A is Primary and System B is Secondary. This is performed with the **Configuration → Switch Roles** option in the relationship **Actions** menu, as shown in Figure 5-52 on page 75.

Confirm the switch (Figure 5-53) to return to the initial state (Figure 5-54).
5.3.4 Restoring or reinstalling the Quorum Witness

After a Quorum Witness (QW) failure you can restore the QW from a backup. Ensure that the following settings have not changed after the QW backup was taken:

- QW IP address
- QW Hostname
- QW certificate, in the qw.cert file

After restoring the QW the automatic failover capability becomes possible again.

Alternatively, you can reinstall using the QW saved configuration information. The following information is needed to reinstall the QW as it was:

- QW original IP address
- QW original Hostname
- QW original certificate files, the qw.cert and qw.key files

The qw.cert and the qw.key files are located in the /opt/ibm/ibm_quorum_witness/settings/ssl_cert directory on the QW server.

Follow these steps to re-install the QW:

1. Install the QW on a server with the original IP address and host name.
2. Copy the previously saved qw.cert and qw.key files to following folder on the QW server:
   ```bash
   /opt/ibm/ibm_quorum_witness/settings/ssl_cert
   ```
3. Restart the `nginx` service, using the `service nginx restart` command.
4. Activate the Quorum Witness using the Hyper-Scale Manager.

After restoring the QW automatic failover capability becomes possible again.

5.4 Failback scenarios

After a failover, and when all the systems and connections have been reestablished, it is usually desirable to go back to the original configuration where the volume on System A at Site A has the primary role.

You also might need to recover from specific error situations. This section describes the following failback scenarios:
 Recovery from a failover where both volumes have primary role
 Recovery from an error during resynchronization

5.4.1 Recovery from a failover where both volumes have primary role

After a failover to the secondary volume of a HyperSwap relationship, the high availability of the HyperSwap volume has to be reestablished. This is only possible when the failed components are operational again. Because of the failover, the original primary volume on System A does not serve I/Os. It is in a blocked state, but still has the primary role.

The original secondary volume on System B is now serving the I/O and has the primary role. After the failed components are operational again the role of the original primary volume has to be manually changed to secondary.

Figure 5-55 shows the HyperSwap volume state after System A at site A is operational again. The HyperSwap peer volumes are not synchronized.

Two steps are necessary to reestablish the high availability of the HyperSwap volume:
1. Set the correct role for the volume with blocked I/O.
2. Activate HyperSwap, which will automatically start the re-synchronization.

Optionally, change the Primary and Secondary roles so that System A will serve read and write I/O to host A after a switch role on the HyperSwap volumes.

**Note:** Synchronizing the volumes and activating HyperSwap is a manual process which has to be carefully planned and executed.
Reestablishing the high availability after a failover

The HyperSwap pane in the Hyper-Scale Manager GUI indicates the current state of a volume, as shown in Figure 5-4 on page 52. The detailed information shown in Figure 5-56 indicates that the system connectivity and the Quorum Witness are without any failure.

Hovering over the failed components lists the necessary actions to restore the high availability of the volume, as shown in Figure 5-57.

Select **Click here to change the roles of the volumes** to open the Change Role dialog.

The volume on System A at Site A has to become *secondary*, because the volume on System B at Site B is the current *primary*. 

---

**Figure 5-56** HyperSwap detailed availability status

**Figure 5-57** Necessary actions to restore HyperSwap high availability
Mark **Change to Secondary** and click **Apply**, as shown in Figure 5-58.

![Figure 5-58 Changing the role of an I/O blocked volume](image)

Figure 5-58 shows the result of changing the role.

![Figure 5-59 Correct role of an I/O blocked volume](image)

Figure 5-59 shows the result of changing the role.
Hovering over the failed components displays the second and last necessary actions to restore the high availability (automatic failover) of the volume, as shown in Figure 5-60.

![There's no automatic failover on the Volume](image)

**Figure 5-60  Last necessary action to restore HyperSwap volume high availability**

Select **Click here to activate HyperSwap**, as shown in Figure 5-60 to open the Change Activation State dialog.

Mark **Activation State: Active** and click **Apply**, as shown in Figure 5-61.

![Change Activation State](image)

**Figure 5-61  Activate dialog box**
Figure 5-62 shows the result of activating HyperSwap. The volumes on System A and System B will be automatically synchronized. The figure shows a value of 82%, which means that 82% of the HyperSwap volume are currently synchronized.

A snapshot of the secondary volume is taken before starting synchronization from the primary volume to the secondary volume. The snapshot represents the last consistent copy of the secondary volume before it is synchronized. See 5.4.2, “Recovery from an error during resynchronization” on page 83 for an example of when that snapshot could be used. After successful synchronization of the secondary volume, the snapshot will be deleted.

After the volumes are synchronized, the GUI indicates that the high availability is reestablished and an automatic failover is now possible again, as shown in Example 5-63.

The high availability of the HyperSwap volume is now reestablished but I/O is served from System B (A9000R) volume which is now the primary. Before the failover I/O was served from System A.
Switching the I/O back to the original primary (volume on System A)

Figure 5-63 on page 81 shows that the primary volume serving I/O is on System B (A9000R). This is not the original configuration that existed before the failover.

To fail back to the original configuration where the volume on System A was in primary role, use the **Switch Role** action for the HyperSwap volume. Select the HyperSwap volume, right-click, and select **HyperSwap → Switch Role**, as shown in Figure 5-64.

In the **Switch Role** dialog for the selected HyperSwap volume, click **Apply** to start the switch role process, as shown in Figure 5-65.
The primary HyperSwap volume is now back on System A, as shown in Figure 5-66. I/O is now served again from the original site, as it was before the failover occurred.

5.4.2 Recovery from an error during resynchronization

This section describes the recovery from an error during the resynchronization of the HyperSwap volumes following a failover. The steps to start the re-synchronisation of the secondary volume are described in 5.4.1, “Recovery from a failover where both volumes have primary role” on page 77. The status during synchronisation of the secondary volume is shown in Figure 5-62 on page 81.

If System B with the current primary volume has an outage during the re-synchronisation, no I/O access from the host is possible. The secondary volume (currently on System A) is in I/O blocked state and the primary volume is not available, as shown in Figure 5-67.

Because System B is not available, no information from the volume on System B can be displayed.

The secondary volume contains data from different points in time:

- Data from the point system A stopped
- Data partially updated during the synchronisation
Two scenarios are possible to get the volume in a storage consistent status:

- Waiting until System B is available again
- Using the secondary volume for further host access

### Waiting until System B is available again

The synchronisation of the secondary volume will start automatically after System B is available again. After synchronisation, the automatic failover capability of the volume is re-established. The administrator can now use the *switch role* function to choose the site where the primary volume should reside.

### Using the secondary volume for further host access

The last consistent status of the secondary volume is kept in the snapshot taken just before the synchronisation started. Reverting the volume to this snapshot will present a volume to the host, which is consistent from the storage side. Figure 5-68 shows this last consistent snapshot. This snapshot is automatically generated and marked as *internal*.

![Figure 5-68  Last consistent snapshot](image)

Figure 5-68 shows the details of the internal last consistent snapshot.

![Figure 5-69  Details on the internal last consistent snapshot](image)

The following steps are needed to revert the secondary volume to this last consistent snapshot and to serve I/O to the host:

- Unmap the volume from the host
- Change role from secondary to primary
- Deactivate HyperSwap for this volume
Chapter 5. HyperSwap scenarios

- Restore the last consistent snapshot
- Map the volume from the host
- Delete snapshot

These steps are described hereafter.

The volume can be unmapped from the host using the Hyper-Scale Manager GUI.

Figure 5-70 shows how to change the role of the volume to primary. If the volume were still mapped to the host, the host could write and read inconsistent data on the volume. A note in the change role window is describing this situation. The volume should only be accessible by the host after the last consistent snapshot is restored.

**Note:** The volume has to be unmapped from the host before switching its role to primary.
Figure 5-71 show the result of the **switch role** command.

![Figure 5-71 Result of the switch role command](image)

The **switch role** command implies a renaming of the internal last consistent snapshot. It is not internal anymore and can be used by the administrator:

- **Previous name**: `last-consistent-<Volume Name>`
- **New name**: `external-last-consistent-<Volume Name>`

In our example `<Volume Name>` is `ITSO_RH_HA_003`. The HyperSwap relationship has to be cancelled, as shown in Figure 5-72, to be able to revert the volume to the last consistent snapshot.

![Figure 5-72 Deactivating HyperSwap](image)
To restore the last consistent snapshot select **Snapshots → Restore from Snapshot** from the **Actions Menu** and choose the **external-last-consistent-<Volume Name>** snapshot, as shown in Figure 5-73.

![Figure 5-73  Restoring the last consistent snapshot](image_url)

Click **Apply** to restore the snapshot. The Hyper-Scale Manager GUI indicates that the operation was successful, as shown in Figure 5-74.

![Figure 5-74  Successful snapshot restore](image_url)

The volume is now consistent from the storage side and can be mapped to a host. The last consistent snapshot can now be deleted.
Chapter 6. Multi-site HA/DR implementation and usage

This chapter describes how to implement and use the Multi-site HA/DR feature for FlashSystem A9000 and FlashSystem A9000R.
6.1 Multi-site HA/DR implementation process overview

To ensure smooth implementation, make sure to first review Chapter 3, “Prerequisites” on page 9.

The scenarios are based on the configuration depicted in Figure 6-1.

![Figure 6-1 Basic Multi-site HA/DR configuration](image)

### Multi-site HA/DR CLI commands

For reference, Table 6-1 lists the CLI commands used in the implementation scenarios.

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>multisite_define</td>
<td>Provided that a HyperSwap relation is established between the Primary and Secondary volumes, and an asynchronous mirror is defined between the Primary and Tertiary volumes, issuing this command on the Primary volume creates a named Multi-site relationship for a volume or consistency group. If the command completes successfully, the Multi-site relationship will be recognized on all the systems involved.</td>
</tr>
<tr>
<td>multisite_activate_async_mirror</td>
<td>When issued on the Primary volume or Consistency Group, this command activates an asynchronous mirror as part of a Multi-site relationship.</td>
</tr>
</tbody>
</table>
6.1.1 Assembly from scratch

Creating a new Multi-site HA/DR configuration from scratch implies that you have a set of newly created or existing volumes or Consistency Groups that are not currently replicated to other systems. In this case, the suggested implementation strategy is to set up a HyperSwap relationship for these volumes or Consistency Groups (refer to section 4.1, “HyperSwap implementation process overview” on page 16 for details) and then extend it to a multi-site relationship, as described in section 6.1.2, “Assembly from existing Synchronous relation”.

6.1.2 Assembly from existing Synchronous relation

Creating a Multi-site HA/DR relationship from an existing synchronous mirror is a relatively simple process. Select the volumes in the Remote Views > Replication Details table and select Convert Type > Convert Sync Mirror to HyperSwap under the Actions menu as shown in Figure 6-2.

![Figure 6-2 Select volumes to convert to HyperSwap](image)

Confirm the action as shown in Figure 6-3. Note that this is a non-disruptive process for I/O to the selected volumes.

---

**Table: Command Description**

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>multisite_register_standby_mirror</td>
<td>Registers a Standby mirror in a multi-site relationship. When applied to a Consistency Group, this command registers the Standby relation on every volume in the Consistency Group.</td>
</tr>
<tr>
<td>multisite_list</td>
<td>Lists the configuration and status of existing multi-site relationships.</td>
</tr>
<tr>
<td>multisite_change_role</td>
<td>When issued for a non-operational relationship, this command changes the role of a multi-site relationship peer.</td>
</tr>
<tr>
<td>multisite_switch_roles</td>
<td>When issued on the Primary volume or Consistency Group, this command switches the roles between the Primary and the Secondary volumes or Consistency Groups.</td>
</tr>
<tr>
<td>multisite_delete</td>
<td>Deletes an existing multi-site relationship. The two-way relationships that compose the multi-site relationship are not affected.</td>
</tr>
</tbody>
</table>
Figure 6-3  Confirm the convert action

Once confirmed, the system will run through the conversion process. This process happens quickly, as shown in Figure 6-4. More information on the process can be found in Chapter 4, “HyperSwap implementation and usage” on page 15.

Figure 6-4  Active HyperSwap relationship

At this point, the conversion to Multi-site is possible as described in 6.1.3, “Assembly from existing HyperSwap”.

92  IBM HyperSwap for IBM FlashSystem A9000 and A9000R
6.1.3 Assembly from existing HyperSwap

If an existing HyperSwap relationship will be the base for setting up a Multi-site configuration, the steps to extend are easily done using the Hyper-Scale Manager (HSM) GUI or the command line.

Configuration using the Hyper-Scale Manager GUI

With the HSM GUI, it is just a matter of selecting the HyperSwap primary and clicking **Extend to Multi-site** button as shown by the red circle in Figure 6-4 on page 92. The panel shown in Figure 6-5 displays and you need to enter information about the tertiary, along with options to create the standby asynchronous relationship or not.

![Figure 6-5  Extend HyperSwap to Multi-site](image)

Configuration using the CLI

There are 5 steps needed to create a Multi-site assembly from an existing HyperSwap configuration with the command line. These steps are outlined below with illustration shown in Example 6-1.

1. On system A: create asynchronous mirror from A to C
2. On system A: issue `multisite_define`
3. On system B: create asynchronous mirror from B to C
4. On system A: activate multi-site
5. On system B: register standby mirror for multi-site
Example 6-1  Steps to extend existing HyperSwap to Multi-site via CLI

Step 1
A9000-Site-A>> mirror_create rpo=300 schedule=xiv_gui_schedule_1537991049849 vol=ITSO_HS_CLI_003 create_slave=yes part_of_multisite=yes slave_vol=ITSO_HS_CLI_003 type=ASYNC_INTERVAL remote_schedule=xiv_gui_schedule_1537991049849 init_type=ONLINE remote_rpo=300 target=A9000-Site-B remote_pool=ITSO_list2
Command executed successfully.

Step 2
A9000-Site-A>> multisite_define vol=ITSO_HS_CLI_003
Command executed successfully.

Step 3
A9000-Site-B>> mirror_create rpo=300 schedule=xiv_gui_schedule_1538677624024 vol=ITSO_HS_CLI_003 create_slave=no part_of_multisite=yes slave_vol=ITSO_HS_CLI_003 type=ASYNC_INTERVAL remote_schedule=xiv_gui_schedule_1537991049849 init_type=ONLINE remote_rpo=300 target=A9000-Site-C remote_pool=ITSO_list2
Command executed successfully.

Step 4
A9000-Site-A>> multisite_activate_async_mirror vol=ITSO_HS_CLI_003
Command executed successfully.

Step 5
A9000-Site-B>> multisite_register_standby_mirror vol=ITSO_HS_CLI_003
Command executed successfully.

Tip: When extending to Multi-site using the command line, you can use an existing schedule that meets the RPO and interval needed (found by issuing schedule_list) or create a new schedule with the schedule_create CLI command.

6.1.4 Assembly from existing Asynchronous relation

This section describes how to assemble a multi-site relationship from an existing asynchronous relationship between two volumes or consistency groups (CG). This task can be fulfilled with Hyper-Scale Manager (HSM) GUI or using the CLI.

Configuration using the Hyper-Scale Manager GUI

Complete the following steps on the system that owns the Primary volume or CG in the Asynchronous relationship (system A):

1. Navigate to the Volumes (or Consistency Groups) view and locate the required asynchronous relationship, as shown in Figure 6-6.

![Figure 6-6  Select an Asynchronous relationship](image)
2. Navigate to the selected relationship’s Replication Properties, as shown in Figure 6-7:

![Figure 6-7 Replication properties of Asynchronous relationship between volumes](image)

For consistency groups, this panel appears, as shown in Figure 6-8.

![Figure 6-8 Replication properties of Asynchronous relationship between CGs](image)

3. Select **Extend to Multi-site**. The **Extend to Multi-site** panel displays, as shown in Figure 6-9. You must create the secondary volume in the HyperSwap relationship that will be part of the new Multi-site relationship. Select the system and the pool that will own the secondary volume. The volume name is identical to the primary volume's name and cannot be changed.

You can also decide whether to activate Multi-site on creation or not by checking the appropriate option. In addition, you can choose whether to create a standby mirror as part of the new multi-site relationship. For the purposes of this document, both of these options are kept as default.
For consistency groups, this panel appears, as shown in Figure 6-10.
4. Select **Apply and Close** to initiate the multi-site relationship assembly. The standby mirror creation is completed only after the secondary and tertiary volumes (or CGs) are fully synchronized. Until then, the multi-site relationship status will remain **Compromised**, as shown in Figure 6-11.

![Figure 6-11](image1.png)

*Figure 6-11  Multi-site in Compromised status due to synchronization in progress (volumes)*

For consistency groups, this panel appears, as shown in Figure 6-12.

![Figure 6-12](image2.png)

*Figure 6-12  Multi-site in Compromised status due to synchronization in progress (CGs)*
5. When the synchronization is complete, the multi-site relationship status changes to **Operational**, as shown in Figure 6-13.

![Figure 6-13 Multi-site in Operational status (volumes)](image)

For consistency groups, this panel appears, as shown in Figure 6-14.

![Figure 6-14 Multi-site in Operational status (CGs)](image)
6. The newly created multi-site relationship appears in the Volumes view, as shown in Figure 6-15.

![Figure 6-15](image1)

**Figure 6-15  New multi-site relationship in Volumes view**

For CGs, newly created multi-site relationship appears in the Consistency Groups view, as shown in Figure 6-16.

![Figure 6-16](image2)

**Figure 6-16  New multi-site relationship in Consistency Groups view**

### Assembly from existing asynchronous relationship using the CLI

The assembly process to create a multi-site relationship starting from an existing asynchronous relationship using the CLI is similar to the one already shown for the GUI, but it is broken down into more granular steps in order to create all the required components that will eventually form the multi-site relationship.

In this scenario, the starting point is either an existing asynchronous relationship between two volumes or two Consistency Groups (holding multiple volumes) that you want to extend to multi-site by including a HyperSwap leg between systems A and B and a Standby Asynchronous leg between systems B and C.

We can check the mirroring details on the system hosting the Primary volumes (we assume this is system A in our example from now on, as depicted in Figure 6-1) by using the `mirror_list` command. Example 6-2 shows the output for a single relationship between volumes, Example 6-3 for mirroring between Consistency Groups across systems:

**Example 6-2  Mirroring relationship details for a single volume**

```
A9000R-Site_A>>mirror_list
Name     Mirror Type     Mirror Object  Remote System  Remote Peer
ITSO_AS_Volume_1 async_interval  Volume         A9000R-Site_C  ITSO_AS_Volume_1_DR
```

In this example, the Primary volume `ITSO_AS_Volume_1` on system A is in an asynchronous mirroring relation with the Tertiary volume `ITSO_AS_Volume_1_DR` on system C.
In this example, the Primary volumes in Consistency Group ITSO_AS.CG_A on System A are asynchronously mirrored to their Tertiary counterparts in Consistency Group ITSO_AS.CG_C on system C.

1. The first step is the creation of the Secondary volume(s) on system B that will receive data from the HyperSwap leg of the multi-site relationship.

   For each volume you are operating on (either in a single relationship or as member of a Consistency Group), create a volume on system B with the same name and size as the Primary volume; Example 6-4 shows the volume creation.

   **Example 6-4  Creating the Secondary volume**

   A9000-Site_B>>vol_create vol=ITSO_AS_Volume_1 pool=ITSO_B size=10
   Command executed successfully.

   If you are working with Consistency Groups, then you also need to create a new Consistency Group on the system that will host the Secondary volumes in the HyperSwap level of the multi-site relationship, as shown in Example 6-5:

   **Example 6-5  Creating the Secondary Consistency Group**

   A9000-Site_B>>cg_create cg=ITSO_AS_CG_B pool=ITSO_B
   Command executed successfully.

   Next, you need to create a HyperSwap relationship between the Consistency Group containing the Primary volumes and the newly created one on system B that will contain the Secondary volumes.

2. On System A, issue the `ha_create` command, specifying the Primary and Secondary Consistency Group names with the `cg` and `slave_cg` parameters respectively, as shown in Example 6-6. If you are working with a single relationship you can skip this step entirely.

   **Example 6-6**

   A9000R-Site_A>>ha_create cg="ITSO_AS_CG_A" slave_cg="ITSO_AS_CG_B" part_of_multisite="yes" target="A9000-Site_B"
   Command executed successfully.

3. The next step is the creation of the HyperSwap leg of the multi-site relationship.

   On the system hosting the Primary volume(s), issue the `ha_create` command as shown in Example 6-7, specifying the volume name, the target system, and the pool. Notice the use of the `part_of_multisite` parameter. If you are working with Consistency Groups, then you need to perform this for each pair of Primary-Secondary volumes involved. The Secondary volumes will automatically be added to the Secondary Consistency Group.

   **Example 6-7  Creating the HyperSwap leg of the multi-site relation**

   A9000R-Site_A>>ha_create vol="ITSO_AS_Volume_1" remote_pool="ITSO_B" create_slave="no" part_of_multisite="yes" init_type="ONLINE" target="A9000-Site_B"
   Command executed successfully.
4. On system A, define the new multi-site relation with the multisite_define command. Follow Example 6-8 for a single relationship and Example 6-9 for a Consistency Group.

Example 6-8  Defining the new multi-site relation between volumes

```
A9000R-Site_A>>multisite_define vol="ITSO_AS_Volume_1"
Command executed successfully.
```

Example 6-9  Defining the new multi-site relation between Consistency Groups

```
A9000R-Site_A>>multisite_define cg="ITSO_AS_CG_A"
Command executed successfully.
```

5. The next step requires the creation of a sync job schedule on both the System hosting the Secondary volume(s) (system B) and the one hosting the Tertiary volume(s) (system C). The schedules will be used by the Standby Asynchronous relation(s) that we are going to create next. Example 6-10 and Example 6-11 show the relevant steps. The schedule name and interval must be the same across both systems.

Example 6-10  Creating a sync schedule for the Standby Asynchronous relation on the system hosting the Secondary volume

```
A9000-Site_B>>schedule_create schedule="ITSO_AS_Schedule_BC" interval="00:01:40"
Command executed successfully.
```

Example 6-11  Creating a sync schedule for the Standby Asynchronous relation on the system hosting the Secondary volume

```
A9000R-Site_C>>schedule_create schedule="ITSO_AS_Schedule_BC" interval="00:01:40"
Command executed successfully.
```

**Note:** You can only choose between a few predetermined interval values. The interval value must be smaller than your desired RPO. For more information on this topic, refer to *IBM FlashSystem A9000 Version 12.3.0 Command-Line Interface (CLI) Reference Guide*, SC27-8559-09.

6. On System B, establish the Standby Asynchronous relationships between the Consistency Group holding the Secondary Volumes and the one on System C holding the Tertiary volumes. The target Consistency Group must be the one that is already in use in the Active Asynchronous relationship between system A and system C, as shown in Example 6-3 on page 100. Specify your desired RPO and make sure to use the schedules defined on both systems in the previous step.

If you are working with a single relationship you can skip this step.

Example 6-12

```
A9000-Site_B>>mirror_create slave_cg="ITSO_AS_CG_C" rpo="300" schedule="ITSO_AS_Schedule_BC" cg="ITSO_AS_CG_B" part_of_multisite="yes" type="ASYNC_INTERVAL" remote_schedule="ITSO_AS_Schedule_BC" remote_rpo="300" target="A9000R-Site_C"
Command executed successfully.
```
7. The Standby Asynchronous relation(s) between the Secondary and the Tertiary volumes can now be created. On the System hosting the Secondary volume, issue the `mirror_create` command, specifying your desired RPO, both the local and remote sync schedule defined in step 5, the source volume defined in step 1 on page 100 (Secondary) and the target volume (Tertiary). Refer to Example 6-13 for the full list of parameters that need to be used in this step.

Example 6-13   Creating the Asynchronous mirror between the Secondary and Tertiary volumes

```
A9000-Site_B>>mirror_create rpo="300" schedule="ITSO_AS_Schedule_BC"
    vol="ITSO_AS_VOLUME_1" create_slave="no" part_of_multisite="yes"
    slave_vol="ITSO_AS_VOLUME_1_DR" type="ASYNC_INTERVAL"
    remote_schedule="ITSO_AS_Schedule_BC" init_type="ONLINE"
    remote_rpo="300" target="A9000R-Site_C"
Command executed successfully.
```

8. On the system hosting the Secondary volume, issue the `multisite_register_standby_mirror` specifying the name of the Secondary volume (for a single relation, Example 6-14) or the name of the Secondary Consistency Group (Example 6-15):

Example 6-14   Register the multi-site Standby mirror volumes relationship

```
A9000-Site_B>>multisite_register_standby_mirror vol="ITSO_AS_VOLUME_1"
Command executed successfully.
```

Example 6-15   Register the multi-site Standby mirror Consistency Group relationship

```
A9000-Site_B>>multisite_register_standby_mirror cg="ITSO_AS_CG_B"
Command executed successfully.
```

9. Finally, activate the HyperSwap relationship(s) between the Primary and Secondary volumes (for a single volume, Example 6-16) or the Primary and Secondary Consistency Groups (Example 6-17) by issuing the `ha_activate` command on system A:

Example 6-16   Enabling the HyperSwap leg of the multi-site relationship for a single volume

```
A9000R-Site_A>>ha_activate vol="ITSO_AS_VOLUME_1" target="A9000 Demo"
Command executed successfully.
```

Example 6-17   Enabling the HyperSwap leg of the multi-site relationship for a Consistency Groups

```
A9000R-Site_A>>ha_activate cg="ITSO_AS_CG_A" target="A9000-Site_B"
Command executed successfully.
```

You can verify the state of the newly created multi-site relation with the `multisite_list` command, as shown in Example 6-18 for a single relationship and in Example 6-19 for a Consistency Group.

Example 6-18   Listing multi-site relations details for a single volume

```
A9000R-Site_A>>multisite_list
Name       State       Multisite Standby    Master   SMaster   Slave
ITSO_AS_VOLUME_1 Operational Up Local A9000-Site_B A9000R-Site_C
```
Example 6-19  Listing multi-site relations details for Consistency Groups

<table>
<thead>
<tr>
<th>Name</th>
<th>State</th>
<th>MultisiteStandby</th>
<th>Master</th>
<th>SMaster</th>
<th>Slave</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITSO_AS(CG)</td>
<td>Operational</td>
<td>Up</td>
<td>Local</td>
<td>A9000-Site_B</td>
<td>A9000R-Site_C</td>
</tr>
</tbody>
</table>

6.1.5 Assembly with CG

Creating a Multi-site HA/DR configuration can be done using a brand new empty consistency group. The process is similar to creating a mirrored consistency group for HyperSwap or replication as well as converting an existing mirrored or HyperSwap consistency group to Multi-site. This section reviews both options.

New Consistency Group

The steps involved in creating a new Multi-site relationship for a consistency group (CG) are similar to the steps required for a new HyperSwap or Asynchronous relationship. You need to start with an empty CG and configure either an Asynchronous or HyperSwap relationship. Once the empty CG Multi-site setup has reached an operational state, you can add Multi-site volumes to the Multi-site CG.

Existing Consistency Group in a HyperSwap relationship

Creating a Multi-site relationship with an existing populated CG in a HyperSwap relationship has some additional steps on top of a new CG. In this case Site A is the HyperSwap primary, Site B is the HyperSwap secondary. We are starting with a HyperSwap CG called ITSO_CG_HS.

Important: Ensure that a Network Time Protocol (NTP) server is configured on all systems involved in Multi-site to avoid issues with potential failures due to time differences.

Extend to Multi-site in the Hyper-Scale Manager

The steps to move this to a Multi-site relationship are as follows:

1. Create a local consistency group at Site C as shown is Figure 6-17.

Figure 6-17  Create a new consistency group on C
2. Under the Remote Views > Replication Details view, select the existing HyperSwap consistency (in this case ITSO_CG_HS) and select **Extend to Multi-site**. This option is shown in Figure 6-18.

![Figure 6-18 Extend to Multi-site option in the HSM GUI](image)

3. At this point the asynchronous replication links, both active and standby, are created. The active asynchronous link will start with initialization and then Multi-site will move to an operational state. These states are shown in Figure 6-19.

![Figure 6-19 Multi-site changing states](image)

**Extend to Multi-site using the CLI**

The process for creating a Multi-site consistency group requires a few extra steps that are done in the background when using the HSM GUI. The following steps describe what needs to be done when using the CLI and are followed by an illustration of the CLI commands shown in Example 6-20.

1. Create a consistency group on site C.
2. Create a schedule, or find an existing schedule that meets the requirement on site A and B (the HyperSwap pair). We have shown the example of creating the schedule.
3. Create the asynchronous consistency group mirror between sites A and C (this will become the active link).
4. Create the asynchronous mirror between the volumes that will be in the consistency group on A and C.

5. Define the Multi-site relationship for the consistency group on A.

6. Create the standby asynchronous consistency group mirror between B and C.

7. Register the standby asynchronous mirror on A.

8. Activate the Multi-site asynchronous consistency group relationship on A.

Once the empty Multi-site consistency group has reached an operational state, the Multi-site volumes can be added to the consistency group using the command `cg_add_vol`.

```
Example 6-20  Setup Multi-site with command line

A9000R-Site_C>>cg_create cg=ITSO_HS_CLI pool=ITSO_C
Command executed successfully

A9000R-Site_A>>schedule_create schedule=xiv_gui_schedule_1538587261921
interval=00:01:40
Command executed successfully.

A9000-Site_B>>schedule_create schedule=xiv_gui_schedule_1538587261921
interval=00:01:40
Command executed successfully.

A9000R-Site_A>>mirror_create slave_cg=ITSO_HS_CLI rpo=300
schedule=xiv_gui_schedule_1538587261921 cg=ITSO_HS_CLI part_of_multisite=yes
type=ASYNC_INTERVAL remote_rpo=300 remote_schedule=xiv_gui_schedule_1538587261921
target=A9000R-Site_C
Command executed successfully

A9000R-Site_A>>mirror_create rpo=300 schedule=xiv_gui_schedule_1538587261921
vol=ITSO_HS_CLI_003 remote_pool=ITSO_C create_slave=yes part_of_multisite=yes
slave_vol=ITSO_HS_CLI_003 type=ASYNC_INTERVAL
remote_schedule=xiv_gui_schedule_1538587261921 init_type=ONLINE remote_rpo=300
target=A9000R-Site_C
Command executed successfully

A9000R-Site_A>>multisite_define cg=ITSO_HS_CLI
Command executed successfully

A9000-Site_B>>mirror_create slave_cg=ITSO_HS_CLI rpo=300
schedule=xiv_gui_schedule_1538587261921 cg=ITSO_HS_CLI part_of_multisite=yes
type=ASYNC_INTERVAL remote_rpo=300 remote_schedule=xiv_gui_schedule_1538587261921
target=A9000R-Site_C
Command executed successfully

A9000R-Site_A>>multisite_register_standby_mirror cg=ITSO_HS_CLI
Command executed successfully

A9000R-Site_A>>multisite_activate_async_mirror cg=ITSO_HS_CLI
Command executed successfully

A9000R-Site_A>>cg_add_vol cg=ITSO_CG vol=ITSO_CG_MS_001
Command executed successfully
```
When adding a mirrored volume to a mirrored consistency group with either the HSM GUI or command line, you may need to retry the operation if a sync job is in progress, making the last sync time different. Simply retry the command. An example of this failure with the command line is shown in Example 6-21.

Example 6-21 Adding a mirrored volume to a CG may fail

```
A9000R-Site_A>> cg_add_vol cg=ITSO_CG vol=ITSO_CG_MS_001
Command executed successfully
```

6.1.6 Leaving CG with active HyperSwap or asynchronous relation

It is possible to remove an active Multi-site volume from a Multi-site consistency group and maintain the active relationships.

Using the HSM GUI

To remove a Multi-site volume from a Multi-site consistency group is a simple process. Select the volume(s) in the Volumes view (or link to them from the consistency group as shown in Figure 6-20 and then select the volume(s) to remove.

**Figure 6-20** Navigate to the volumes in the consistency group

Once the volume is selected, right click (or open the actions menu) and select **Consistency Group > Remove from Group**, as shown in Figure 6-21.

**Figure 6-21** Select Remove from Group in the Actions menu

**Important**: Without registering the standby mirror, there will not be an automatic activation in the case of a HyperSwap failover.
You are prompted to confirm the removal of the volume(s) from the consistency group. Click Apply to confirm, as shown in Figure 6-22. This process is the same as used for a replicated or simplex volume(s) and consistency group.

![Figure 6-22 Confirm removal from consistency group](image)

Once this is done, you will be able to see both the Multi-site consistency group and volume(s) in the Replication Details View. This is shown in Figure 6-23.

![Figure 6-23 Multi-site relationship persists for volume removed from CG](image)

### Using command line

The command for removing a Multi-site volume from a Multi-site consistency group is the same as removing a simplex volume from a consistency group. See Example 6-22.

**Example 6-22  Remove a Multi-site volume from Multi-site consistency group**

```
A9000R-Site_A>> cg_remove_vol vol=ITS_O_CG_HS_004  
Command executed successfully
```
6.2 Data migration

One possible scenario enabled by the use of Multi-site relationships is to perform Data Migration. Before the introduction of Multi-site, the existing Data Migration solutions for IBM FlashSystem A9000 and A9000R all involved a maximum of 2 systems - meaning that any existing replication relationship involving the volume(s) to be migrated had to be necessarily deleted beforehand. This situation is especially annoying in the case of the migration of a Volume which already has an Asynchronous mirroring relationship towards a DR system - migration efforts require interruption of the DR mirroring for the whole process, which can take quite some time.

Multi-site relationships allow for 3 Systems to replicate data from the same volume, therefore enabling the migration of a Volume from one System to another, while keeping data replicated asynchronously to a third System.

Figure 6-24 on page 108 shows the initial configuration. In this example, we want to migrate a volume from System A to System B, online - i.e. without causing loss of access to the host(s) using it - and retaining Asynchronous DR mirroring towards System C.

In order to proceed, the three Systems involved should conform to the Multi-site implementation requirements. For details, refer to Chapter 2, “Multi-site HA/DR solution for FlashSystem A9000 and A9000R” on page 1.

Figure 6-24  Initial configuration for the Data Migration scenario

The first step in the Data Migration procedure is to extend the existing Asynchronous mirroring relationship between System A and System C to a Multi-site relationship including our target System, B. To do this, follow the procedure in 6.1.4, “Assembly from existing Asynchronous relation” on page 94.
The Secondary volume in the Asynchronous relationship between A and C now becomes the Tertiary volume in the Multi-site relationship; the Primary volume will retain its role, while a new Secondary volume will be created on System B. After the extension, data from the Primary volume will be replicated to the Secondary via the HyperSwap leg of the Multi-site relationship, while a new Standby Asynchronous relationship will be created between B and C. At the end of this step, the configuration will resemble the one shown in Figure 6-25.

Host IO will still be served by the Primary volume on A but, after mapping the Secondary volume in B to the host(s), additional Unoptimized paths will be available.

**Important:** When mapping the Secondary volume(s), make sure to use the same LUN ID assignment(s) as already in use by the corresponding Primary volume(s), to avoid host access errors.

![Figure 6-25 Configuration after the extension to a Multi-site relationship](image)

Wait for the Multi-site relationship to be fully Synchronized and Operational, as shown in Figure 6-26.
Primary and Secondary roles can now be switched to indicate to the host(s) to start using paths towards System B to service IOs, and activating the B-C Asynchronous relationship. To do so, start by deactivating the A-C (Active) Asynchronous relationship between A and C; from the Actions menu, select Configuration → Change Activation (Figure 6-27).

Select Activation State: Inactive and confirm by clicking Apply as shown in Figure 6-28.

Switch the Primary and Secondary roles by selecting Failover/Recovery → Switch Roles from the Actions menu, as shown in Figure 6-29.

Make sure to select Switch the designations as well in the subsequent dialog, then confirm by clicking Apply (Figure 6-30).
Finally, activate the B to C Asynchronous relationship to bring the Multi-site relationship back to the *Operational* state.

![Change Activation State](image)

*Figure 6-31  Activating the Asynchronous relationship between System B and System C*

The configuration is now as shown in Figure 6-32.

![Configuration Diagram](image)

*Figure 6-32  Multi-site configuration after the roles switch*

Figure 6-33 shows the HSM GUI representation.

![GUI Representation](image)

*Figure 6-33  The Multi-site configuration as shown in the Hyper-Scale Manager GUI*
With the IO now being serviced by the target System B, we can dismantle the Multi-site relationship to remove the volume from System A.

Start by deactivating both the HyperSwap and the Active Asynchronous legs of the Multi-site relationship, as shown in Figure 6-34 and Figure 6-35.

![Figure 6-34 Deactivating the HyperSwap leg of the Multi-site relationship](image)

![Figure 6-35 Deactivating the Active Asynchronous leg of the Multi-site relationship](image)

Next step is to unmap the migrated volume(s) from the now designated Secondary system (A, in our example). See Figure 6-36.

![Figure 6-36 Unmap migrated volumes from the Secondary system](image)

The Multi-site relationship can now be dismantled; from the Actions menu, select **Delete → Convert Multi-site to Async Mirror**, as shown in Figure 6-37.

![Figure 6-37 Action menu to convert the Multi-site relationship to Async Mirror](image)
Review the summary and confirm by clicking **Apply** (Figure 6-38).

![Convert Multi-site to Async Mirror]

*The following Multi-site relations will be converted to Async mirror relations.*

- ITSO_FA_Volume_2
- ITSO_FA_Volume_2
- ITSO_FA_Volume_2
- ITSO_FA_Volume_2
- ITSO_FA_Volume_2
- ITSO_FA_Volume_2

![Figure 6-38  Multi-site to Async Mirror convert summary]

After this last step, both the HyperSwap and the Standby Asynchronous leg of the Multi-site relationship will be deleted, effectively leaving us with a standalone Asynchronous relationship between System B and System C. Activate this relationship as shown in Figure 6-39 to reach the final configuration, depicted in Figure 6-40 on page 113.

![Change Activation State]

*Activation State: Inactive*

![Figure 6-39  Activate the Asynchronous relationship between B and C]

![Figure 6-40  Final configuration]
Figure 6-41 shows the HSM GUI representation.

<table>
<thead>
<tr>
<th>Local Peer</th>
<th>System</th>
<th>Remote Peer</th>
<th>Remote System</th>
<th>Synchronization</th>
<th>Local Role</th>
<th>Multi-site Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDisk_F2_Volume_2</td>
<td>MDisk_F1_Site_B</td>
<td>MDisk_F2_Volume_2</td>
<td>MDisk_F1_Site_C</td>
<td>BPO OK</td>
<td>Primary</td>
<td></td>
</tr>
<tr>
<td>MDisk_F2_Volume_2</td>
<td>MDisk_F2_Site_C</td>
<td>MDisk_F2_Volume_2</td>
<td>MDisk_F1_Site_B</td>
<td>BPO OK</td>
<td>Secondary</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 6-41  Final relationship details as shown in the Hyper-Scale Manager GUI*
Multi-site scenarios

This chapter describes the following failure scenarios and recovery steps (failback):

- Failure and recovery of the storage system serving the primary volume or CG
- Primary volume or CG loses target
- DR test at secondary volume site
- Primary and Secondary volume or CG failure
- Quorum Witness failures
7.1 Failure and recovery of the storage system serving the primary volume or CG

This section discusses the failure of the storage system serving I/Os for the Primary volume or Consistency Group, in a Multi-site relationship.

Figure 7-1 shows a Multi-site configuration with the Primary volume or Consistency Group on System A.

In this example, the host sees the volume through paths from System A and System B, but IO is served only through the preferred paths from System A. Data is replicated synchronously from system A to system B through the HyperSwap leg of the Multi-Site relationship, and is replicated asynchronously from A to C through the active Asynchronous leg. The Asynchronous B to C relationship is in Standby.

Figure 7-2 shows the replication details of a Consistency Group in a Multi-site relationship in a steady state.
### 7.1.1 System A failure scenario

Figure 7-3 depicts a failure of System A. As a result:
- The host and System B lose connection to System A.
- The asynchronous replication from System A to System C is disrupted.

![Multi-site configuration after the failure of System A](image)

In the Hyper-Scale Manager GUI (Figure 7-4), System A disappears from the replication relationship details, and an automatic failover is performed to the Secondary volume on System B.

![Replication relations details after System A failure](image)

Now the multi-site configuration status is compromised, because System B is acting not as designated, and System A is temporarily not being monitored (Figure 7-5):

![Multi-site relationship status after failover on B](image)
The status of the HyperSwap relationship between the Primary and Secondary volumes is represented in greater detail in the HyperSwap tab of the replication relation properties, as shown in Figure 7-6.

Upon the failure of System A, the Standby asynchronous mirror between Systems B and C was automatically activated. Its details now appear in the Async tab of the replication relation properties, as shown in Figure 7-7.
If System A were up, it would be possible to define a Standby asynchronous mirror between it and System C, but since it is down, this option is not available, as shown in the **Standby** tab of the replication relation properties (Figure 7-8).

![Figure 7-8  Standby asynchronous mirror status after failover on System B](image)

### 7.1.2 Multi-site relationship recovery after System A failure

When System A is operational again, the original Primary volume is still not serving I/Os and is in a blocked state.

The original Secondary volume on System B is serving the I/O and has the primary role, as can be observed in the replication relations details.

The HyperSwap peer volumes are not synchronized. Consequently, the multi-site relationship is still compromised, as shown in the **Multi-site** tab (Figure 7-9).

![Figure 7-9  Replication relations details after restoring System A](image)
Multi-site relationship recovery from the GUI

To re-establish the high availability, the role of the original Primary volume has to first be manually changed to Secondary, via the replication Actions menu (Failover/Recovery → Change Role) and then selecting Change to Secondary, as shown in Figure 7-10.

Figure 7-10  Changing role of the original Primary volume to Secondary

Then HyperSwap must be activated, as shown in Figure 7-11.

Figure 7-11  HyperSwap activation

Figure 7-12 shows the result of activating HyperSwap. The volumes on System A and System B are being automatically synchronized.

Figure 7-12  Synchronizing the HyperSwap volume
After the volumes are synchronized, the GUI indicates that the high availability is re-established; an automatic failover is possible, and the multi-site relationship is operational, as shown in Figure 7-13.

![Figure 7-13  Re-synchronized HyperSwap volume](image)

The high availability of the HyperSwap volume is now re-established, but I/O is served from System B, which now holds the Primary volume. Figure 7-14 depicts this transitional state.

![Figure 7-14  State of the Multi-site relationship after A is recovered and High Availability reinstated](image)
To failback to the original configuration that existed before the failover, the asynchronous mirror between Systems B and C must first be de-activated, as shown in Figure 7-15.

Figure 7-15  De-activating asynchronous mirror between Systems B and C

Then, switch the roles of the HyperSwap peers, as shown in Figure 7-16.

Figure 7-16  Switching the roles of the Primary and Secondary volumes

Finally, activate the asynchronous mirroring relationship between Systems A and C. After a brief period of RPO lagging, the Primary-Tertiary asynchronous mirroring is reactivated, and the multi-site relationship is fully restored.

Multi-site relationship recovery from the CLI

The recovery procedure can alternatively be carried out by issuing CLI commands to the three individual systems.

Just like in the GUI procedure, the first step is to re-establish the high availability, by changing the role of the original Primary volume/Consistency Group to Secondary. This step can be done by issuing the `multisite_change_role` command on the system hosting the original Primary volume or Consistency Group, as shown in Example 7-1:

Example 7-1  Changing Multi-site roles with the CLI

```
A9000R-Site_A>>multisite_change_role cg=ITSO_FA_CG_A new_role=SMaster
```

Warning:  Are you sure you want to change the Multi-site role? y/n: y
Command executed successfully.

In order to resynchronize the volume data between Systems A and B, the HyperSwap relationship needs to be reactivated using the `ha_activate` command on the System currently acting as primary (B, in our case) - see Example 7-2.

Example 7-2  Activating the HyperSwap leg of the Multi-site relationship via CLI

```
A9000-Site_B>>ha_activate cg=ITSO_FA_CG_B
```

Command executed successfully.
Wait until the status of the HyperSwap relationship become *Synchronized* again. You can check the status with the `ha_list` command, as shown in Example 7-3.

**Example 7-3  Displaying HyperSwap relationship details via CLI**

```
A9000-Site_B>>ha_list cg=ITSO_FA_CG_B
Name    HA Object  Role   Remote System  Active  Status
ITSO_FA_CG_B  CG       Master  A9000R-Site_A  no    Synchronized
```

Temporarily disable the Asynchronous relationship between the current Primary and the Tertiary volume or Consistency Group by issuing the `mirror_deactivate` command on System B, as seen in Example 7-4.

**Example 7-4  Deactivating an Asynchronous relationship via CLI**

```
A9000-Site_B>>mirror_deactivate cg=ITSO_FA_CG_B target=A9000R-Site_C
Warning: Are you sure you want to deactivate mirroring? y/n: y
Command executed successfully.
```

At this point, it is possible to perform a switch role in order to reinstate the original Primary-Secondary roles; use the `multisite_switch_roles` command on the current Primary system, as shown in Example 7-5.

**Example 7-5  Switching roles in a Multi-site relationship via CLI**

```
A9000-Site_B>>multisite_switch_roles  cg=ITSO_FA_CG_B
Warning: Are you sure you want to switch the roles in this relation? y/n: y
Command executed successfully.
```

Activate the Asynchronous relationship between the systems now hosting the Primary and Tertiary volumes/Consistency Groups with the `multisite_activate_async_mirror` command, as shown in Example 7-6.

**Example 7-6  Activate the Active Asynchronous leg of a Multi-site relationship via CLI**

```
A9000R-Site_A>>multisite_activate_async_mirror cg=ITSO_FA_CG_A
Command executed successfully.
```

Wait until the status of the Asynchronous relationship reports RPO OK (you can use the `mirror_list` to verify that, as shown in Example 7-7).

**Example 7-7  Displaying Asynchronous relationship details via CLI**

```
A9000R-Site_A>>mirror_list cg=ITSO_FA_CG_A
Name        Mirror Type   Role    Remote System  Remote Peer   Status
ITSO_FA_CG_A async_interval  Master  A9000R-Site_C  ITSO_FA_CG_C  RPO OK
```

Finally, check that the Multi-site relationship exits the *Compromised* state and reports as *Operational*, with the `multisite_list` command, as illustrated in Example 7-8.

**Example 7-8  Displaying Multi-site relationship details via CLI**

```
A9000R-Site_A>>multisite_list cg=ITSO_FA_CG_A
Name   Multisite Object  Multisite ID   Role   State
ITSO_FA_CG_A   CG   0142C95000000009   SMaster   Operational
```
7.1.3 Host paths and IO behavior during the outage and recovery

The Multi-site implementation leverages HyperSwap to guarantee high availability and prevent loss of access to volumes at the production sites. On top of that, an Active Asynchronous mirror remains active even when losing either the Primary or Secondary system, therefore consistency of data at the DR site is also guaranteed.

Example 7-9 on page 124 shows the paths state on a VMware ESXi 6.0 hosts for a volume in a Consistency Group that is in a Multi-site relationship. Exactly like for a regular HyperSwap configuration, paths towards the Primary volumes (the ones in Target Port Group 0) are preferred/optimized (Array Priority: 1, TPG_state=AO, active optimized), while paths towards the Secondary system (Target Port Group 1) are available but not preferred (Array Priority: 0, TPG_state=ANO, active not optimized).

Example 7-9  Path details for a volume in a Multi-site relationship on an ESXi host

```
[root@A9Demoesx0c1:~] esxcli storage nmp path list --device
naa.6001738c7c8053980000000000ecf70 | grep -E 'fc.|Prio|Storage'

fc.200000090faa30c8e:10000090faa30c8e-fc.5001738056750000:5001738056750112-naa.6001
738c7c805398000000000000e7f0
  Array Priority: 0
  Storage Array Type Path Config:
  {TPG_id=1,TPG_state=ANO,RTP_id=4129,RTP_health=UP}
fc.200000090faa30c8e:10000090faa30c8e-fc.5001738056750000:5001738056750132-naa.6001
738c7c805398000000000000e7f0
  Array Priority: 0
  Storage Array Type Path Config:
  {TPG_id=1,TPG_state=ANO,RTP_id=12321,RTP_health=UP}
fc.200000090faa30c8d:10000090faa30c8d-fc.5001738056750000:5001738056750110-naa.6001
738c7c805398000000000000e7f0
  Array Priority: 0
  Storage Array Type Path Config:
  {TPG_id=1,TPG_state=ANO,RTP_id=4097,RTP_health=UP}
fc.200000090faa30c8d:10000090faa30c8d-fc.5001738056750000:5001738056750122-naa.6001
738c7c805398000000000000e7f0
  Array Priority: 0
  Storage Array Type Path Config:
  {TPG_id=1,TPG_state=ANO,RTP_id=8225,RTP_health=UP}
fc.200000090faa30c8d:10000090faa30c8d-fc.5001738056750000:5001738056750112-naa.6001
738c7c805398000000000000e7f0
  Array Priority: 0
  Storage Array Type Path Config:
  {TPG_id=1,TPG_state=ANO,RTP_id=8193,RTP_health=UP}
fc.200000090faa30c8d:10000090faa30c8d-fc.5001738056750000:5001738056750130-naa.6001
738c7c805398000000000000e7f0
  Array Priority: 0
  Storage Array Type Path Config:
  {TPG_id=1,TPG_state=ANO,RTP_id=12289,RTP_health=UP}
```

IBM HyperSwap for IBM FlashSystem A9000 and A9000R
In Example 7-9, the System hosting the Primary volumes (A) in the Consistency Group is an A9000R zoned to host with 8 ports across two fabrics (8 paths total), and the System hosting the Secondary volumes (B) is an A9000 with 6 ports zoned across two fabrics (6 paths total). In a normal state, all the paths are reported as UP.

When System A fails, the host will start reporting errors on the paths leading to it, finally marking them as DOWN and continuing IO on the paths towards System B, now selecting them as preferred. Example 7-10 (truncated for brevity) shows the paths state in this configuration.

Example 7-10  Path details after the failure of System A

```
[root@A9Demoesxdc1:~] esxcli storage nmp path list --device-naa.6001738c7c8053980000000000000000ecf70 | grep -E 'fc.|Prio|Storage'
fcc.20000090faa30c8e:10000090faa30c8e-fc.5001738053980000:5001738053980142-naa.6001738c7c8053980000000000000000ecf70
  Array Priority: 1
  Storage Array Type Path Config:
  {TPG_id=0,TPG_state=AO,RTP_id=512,RTP_health=UP}
fcc.20000090faa30c8d:10000090faa30c8d-fc.5001738053980000:5001738053980130-naa.6001738c7c8053980000000000000000ecf70
  Array Priority: 1
  Storage Array Type Path Config:
  {TPG_id=0,TPG_state=AO,RTP_id=1026,RTP_health=UP}
fcc.20000090faa30c8e:10000090faa30c8e-fc.5001738053980000:5001738053980112-naa.6001738c7c8053980000000000000000ecf70
  Array Priority: 1
  Storage Array Type Path Config:
  {TPG_id=0,TPG_state=AO,RTP_id=258,RTP_health=UP}
fcc.20000090faa30c8d:10000090faa30c8d-fc.5001738053980000:5001738053980140-naa.6001738c7c8053980000000000000000ecf70
  Array Priority: 1
  Storage Array Type Path Config:
  {TPG_id=0,TPG_state=AO,RTP_id=256,RTP_health=UP}
fcc.20000090faa30c8e:10000090faa30c8e-fc.5001738053980000:5001738053980122-naa.6001738c7c8053980000000000000000ecf70
  Array Priority: 1
  Storage Array Type Path Config:
  {TPG_id=0,TPG_state=AO,RTP_id=1024,RTP_health=UP}
```
When System A is available again, paths towards A are reported as UP again, but the paths to System B remains preferred because its designation is still Primary (Example 7-11).

Example 7-11   Paths state after A is recovered, with B still Primary

```
[root@A9Demoesxdc1:~] esxcli storage nmp path list --device naa.6001738c7c80539800000000000ecf70 | grep -E 'fc.|Prio|Storage'
fc.20000090faa30c8e:10000090faa30c8e-fc.5001738056750000:5001738056750112-naa.6001 738c7c805398000000000000ecf70
Array Priority: 1
Storage Array Type Path Config:
{TPG_id=0,TPG_state=UNAVAIL,RTP_id=12321,RTP_health=UP}
f
```
Finally, when the Multi-site relationship is brought back to its original state - with A Primary and B Secondary - the path state will return to its original configuration as shown in Example 7-9 on page 124.

Figure 7-17 is a representation of the IO activity on the Systems involved in the Multi-site relationship.

In the initial phase (steady state, between points 1 and 2 in the diagram) System A (Primary) receives read and write operations from the host; System B only receives the written data, as Secondary System in the HyperSwap leg of the Multi-site relationship. When System A fails (point 2), all IOs from the host are diverted to System B, which continues to serve the host while A is down (time interval between point 2 and 3).

When System A is recovered and then designated as Secondary (point 3), the data modified on B during A’s outage is resynchronized (as indicated by the increasing IOs flowing from B to A in the interval between point 3 and 4).

Finally, when the roles are reverted as they originally were (point 4), the IO distribution returns as it was in the initial steady state phase.
It is important to notice that, while all of these developments were involving System A and System B, System C had a steady flow of replication IOs coming from either systems, because there was always one Active Asynchronous relationship from the current Primary System.

### 7.2 Primary volume or CG loses target

This section illustrates the scenarios in which System A loses either the Fibre Channel (FC) connection to System B, or the iSCSI connection to System C.

#### 7.2.1 Losing Secondary volume or CG

This scenario considers an FC connectivity loss between Systems A and B, which as a result inhibits the automatic failover capability.

The following sequence of events occurs when connectivity between System A and System B has failed:

1. Quorum Witness is up and running, so System A and System B poll the Quorum Witness to check if the other storage system (System B or System A) is up and running.
2. Because Quorum Witness can still talk to both System A and System B, it keeps the volume on System A in primary role and places System B in I/O blocked status, as shown in Figure 7-18.
3. The Multi-site relationship status changes to Compromised, as shown in Figure 7-18.

![Figure 7-18 System B I/O Blocked](image)

The HyperSwap relationship details can be viewed in the HyperSwap tab of the Detailed Replication Relation Properties panel, as shown in Figure 7-19.
When the connectivity between System A and System B is restored, an automatic data synchronization between System A and System B occurs first and HyperSwap is then automatically re-enabled.

Figure 7-20 shows the GUI view when the connectivity between System A and System B has been restored and data is being synchronized between both storage systems.
After the volume synchronization has completed, the GUI shows that the high availability is reestablished and an automatic failover is now possible again, as shown in Figure 7-21.

![Figure 7-21  Automatic failover reenabled](image)

### 7.2.2 Losing Tertiary volume or CG

This scenario considers an iSCSI connectivity loss between Systems A and C, which disables the active asynchronous mirroring relationship, as shown in Figure 7-22.

![Figure 7-22  Asynchronous mirroring relationship details after Systems A and C connection loss](image)
When the connectivity between System A and System C is reactivated, System C is automatically synchronized with the data changes that occurred at System A while the connection was down. Data changes are all contained in System A’s most recent snapshots. Until synchronization from the most recent snapshot is completed, the asynchronous mirror remains in RPO Lagging state, as shown in Figure 7-23.

During synchronization, the Multi-site relationship status is Degraded, as shown in Figure 7-24.

After a while, synchronization is complete, and the asynchronous mirroring is reestablished.
7.3 DR test at secondary volume site

The procedure for testing the disaster recovery situation is very similar whether using a single volume or volumes that are part of a consistency group. In this section we focus on the most likely scenario using a consistency group.

Typically the testing is executed at the DR site, which site C in the case of a Multi-site configuration.

Asynchronous replication uses snapshots as a way to determine what data needs to be copied to the secondary system. Once copied, the internal snapshot on the secondary is renamed as the Last Replicated Snapshot (LRS). This is an internal snapshot and can not be modified or deleted by the user. For more information on how snapshots used for asynchronous replication see IBM FlashSystem A9000 and A9000R Business Continuity Solutions, REDP-5401.

To perform the test based on the LRS, the practice is to duplicate either the single volume or the LRS snapshot group for a consistency group. Once the LRS volume or CG has been duplicated, the user can unlock the snapshot or snapshots that are part of a group, map to a host and perform testing as needed.

Let's look at those steps using the Hyper-Scale Manager.

If you know which volumes are part of the consistency group, you can navigate to the Snapshots or Snapshot Groups view and select each snapshot.

Alternatively, you can start from the Consistency Group view found under the Pools and Volumes View. Select the consistency group and then select the arrow under the Snapshot Groups column heading to navigate to the associated snapshot group. This is shown in Figure 7-25.

![Figure 7-25 Navigate to the snapshot group](Image)

The selected snapshot group is the only one displayed in the table. Notice that the name begins with “last-replicated” and has an “internal” bubble next to the name.

That LRS is greyed out, indicating that you are unable to open, delete or modify it in any way while there is an active asynchronous relationship associated with the corresponding consistency group.

However, if you right click on the snapshot group, or open the Actions menu you do have a few options available. The list of actions includes duplicating and restoring, found under Snapshot Group Data. An example is shown in Figure 7-26.

For the purposes of testing at the DR site, you will select Duplicate Snapshot Group.
Chapter 7. Multi-site scenarios

A new window opens, giving you an option to change the default name of the duplicate before creating it, as shown in Figure 7-27. Names can also be changed later if needed.

Click **Apply** to create the duplicate snapshot group.

![Duplicate snapshot group](image)

**Figure 7-26** Duplicate snapshot group

You should now be able to see the snapshot group you created as well as the internal LRS snapshot group in the table view. To use any snapshot within the group, you must first unlock it for writing (although a host can read from a locked snapshot, it is be unable to make any changes if the snapshot remains locked).

To unlock, right click on the duplicate or select the **Actions** menu followed by **Change Lock State** as shown in Figure 7-28.

![Unlock the snapshot group](image)

**Figure 7-28** Unlock the snapshot group
You are prompted to confirm the lock state change as shown in Figure 7-29.

Once unlocked, the snapshots can be mapped to a host for testing. With the volumes selected, select **Mapping > View/Modify Mapping** under the **Actions** menu. Click on **ADD** for host or cluster, depending on the environment. See Figure 7-30.

This selection allows you to map multiple volumes as the same time to a defined host as shown in Figure 7-31.

If the host has not yet been defined, you have to option to create one at this point.

For the purposes of this illustration, we have selected an existing host from the list and then clicked **Apply**.
The testing of the DR copies can now be done to ensure, in the event of an actual failure of the primary site (or the HyperSwap systems in the case of Multi-Site HA/DR) that you are able to recover from the secondary (or tertiary) site.

7.4 Primary and Secondary volume or CG failure

This scenario considers a failure of System A, followed by a failure of System B.

7.4.1 Systems A and B failure scenario

We assume that the Multi-site configuration prior to the failures was fine and fully operational, with a HyperSwap connection between Systems A and B, an active asynchronous mirroring connection between Systems A and C, and a standby asynchronous mirroring connection between Systems B and C as shown in Figure 7-32.

When System A fails, it disappears from the replication relationship details in the Hyper-Scale Manager HGUI (Figure 7-33), and an automatic failover is performed to the Secondary volume on System B.
Now the multi-site configuration status is compromised, because System B is acting not as designated, and System A is temporarily not being monitored (Figure 7-34):

![Figure 7-34 Multi-site relationship status after failover on System B](image)

In addition, upon the failure of System A, the Standby asynchronous mirror between Systems B and C is automatically activated. Its details now appear in the Async tab of the replication relation properties, as shown in Figure 7-35:

![Figure 7-35 Systems B and C asynchronous mirror status after failover on System B](image)

Following the failure of System A, System B fails as well. As shown in Figure 7-36, currently the only remaining storage system in the Multi-site configuration is System C:

![Figure 7-36 Replication relations details after Systems A and B failure](image)

The Multi-site relationship is effectively destroyed (Figure 7-37), and the asynchronous mirror between the Systems B and C is deactivated (Figure 7-38):
7.4.2 Multi-site relationship recovery after Systems A and B failure

The recovery process of a Multi-site relationship after the failure of Systems A and B begins with changing the System C role from Tertiary to Primary (Figure 7-39), via the replication Actions menu (Failover/Recovery → Change Role):

![Figure 7-39 Changing System C role to Primary]
Since none of the Multi-site relationship peers is now performing its designated role, the Multi-site relationship status is inconsistent (Figure 7-40).

![Figure 7-40  Multi-site status with the System C volume defined as Primary](image)

At this point, the volume on System C must be mapped in order to start serving the I/O from the host.

When Systems A and B are back online, a role conflict occurs, since each peer is now defined as Primary: System A because Primary is its designated role, System B because it became Primary after the System A failure, and System C because we changed its role to Primary (Figure 7-41).

![Figure 7-41  Role conflict with three Primary volumes](image)

The volumes on Systems A and B (former Primary and Secondary volumes) are now out-of-date and need to be synchronized with the data currently residing on System C. While the host workload is directed to System C, the volumes on Systems A and B can be safely unmapped and the blocked I/O can be fixed on System A (Figure 7-42).

![Figure 7-42  Fixing blocked I/O](image)
The next step is to clean up the configuration to an original state by deleting the Multi-site relationships from each of the three systems. This procedure must be performed with CLI commands.

First of all, delete the Multi-site relationship from System A with the `multisite_delete` command (Example 7-12).

**Example 7-12 Deleting Multi-site on System A**

```
A9000R-Site_A>>multisite_delete vol=ITSO_FA_Volume_2
Command executed successfully
```

Then, repeat the command on System B (Example 7-13).

**Example 7-13 Deleting Multi-site on System B**

```
A9000-Site_B>>multisite_delete vol=ITSO_FA_Volume_2
Command executed successfully.
```

And finally on System C - note the use of the `force` flag (Example 7-14).

**Example 7-14 Forcing the deletion of Multi-site on System C**

```
A9000R-Site_C>>multisite_delete vol=ITSO_FA_Volume_2 force=yes
Command executed successfully.
```

The next step is the deletion of the HyperSwap relationship left over from the Multi-site on both Systems A and B - this is performed with the `ha_delete` command (Example 7-15).

**Example 7-15 Deleting HyperSwap on System A and System B**

```
A9000R-Site_A>>ha_delete vol=ITSO_FA_Volume_2
Warning: Are you sure you want to delete this HyperSwap relation? y/n: y
Command executed successfully.
```

```
A9000-Site_B>>ha_delete vol=ITSO_FA_Volume_2
Warning: Are you sure you want to delete this HyperSwap relation? y/n: y
Command executed successfully.
```

It is then necessary to delete both the Asynchronous mirroring relationships left over, by issuing a `mirror_delete` command on all three Systems, as shown in Example 7-16.

**Example 7-16 Deleting Asynchronous Mirroring**

```
A9000R-Site_A>>mirror_delete vol=ITSO_FA_Volume_2 target=A9000R-Site_C
Warning: Are you sure you want to delete this mirroring relation? y/n: y
Command executed successfully.
```

```
A9000-Site_B>>mirror_delete vol=ITSO_FA_Volume_2 target=A9000R-Site_C
Warning: Are you sure you want to delete this mirroring relation? y/n: y
Command executed successfully.
```

```
A9000R-Site_C>>mirror_delete vol=ITSO_FA_Volume_2 target=A9000R-Site_A
Warning: Are you sure you want to delete this mirroring relation? y/n: y
Command executed successfully.
```
Warning: Are you sure you want to delete this mirroring relation? y/n: y
Command executed successfully.
A9000R-Site_C>>mirror_delete vol=ITSO_FA_VOLUME_2 target=A9000-Site_B

Warning: Are you sure you want to delete this mirroring relation? y/n: y
Command executed successfully.

Finally, delete the snapshots from each of the three systems, as shown in Figure 7-43.

![Deleting snapshots](image)

Figure 7-43 Deleting snapshots

From this point on, we can restore the Multi-site relationship.
To restore the Multi-site relationship, start by creating an asynchronous mirroring relationship between C and A with offline initialization, as shown in Figure 7-44, to transfer data from System C to System A.

Once the C-A asynchronous mirror is created, activate it (Figure 7-45) and wait for the synchronization to complete (Figure 7-46).
Properly speaking, the rebuilding of the Multi-site relationship begins at this point. Consider shutting down your disaster recovery hosts in order to suspend data changes until they are brought over to System A.

Switch roles between Systems A and C and change their designations (Figure 7-47) to make System A Primary, and System C Secondary (Figure 7-48):

Figure 7-46  C-A synchronization in progress

Figure 7-47  Switching roles and changing designations

Figure 7-48  A-C asynchronous mirror status after swapping roles
Extend the A-C asynchronous mirroring relationship to Multi-site, using offline initialization, as shown in Figure 7-49.

![Figure 7-49](image)

When the Multi-site relationship is created, its status will be Initializing, as shown in Figure 7-50.

![Figure 7-50](image)
Finally, activate a HyperSwap relationship between A and B (Figure 7-51).

Activating the HyperSwap relationship, initializes the synchronization of all Multi-site relationship components, including the Standby asynchronous relationship between Systems B and C. Upon completion of synchronization, the restored Multi-site relationship will be fully operational, as shown in (Figure 7-52):
7.5 Quorum Witness failures

Issues with Quorum Witness connectivity on one or more systems involved in a Multi-site relationship result in compromised high availability, affecting the HyperSwap leg of the relationship as detailed in section 5.3, “Quorum Witness failure scenarios” on page 61.

The next subsections describe two Quorum Witness failure scenarios similar to the ones already seen for a simple HyperSwap relationship; in this case, the HyperSwap is part of a Multi-site relationship.

In the first scenario, only the Primary system loses connectivity to the Quorum Witness; this situation still allows for Automatic Failover in case of system failure, and recovery is straightforward.

In the second scenario, both the Primary and the Secondary system lose connectivity to the Quorum Witness; this is a more complex scenario, which involves manual intervention in case of rolling disaster.

7.5.1 Loss of connection between the Primary system and the Quorum Witness

In this example, System A is the Primary in a Multi-site environment involving System B (Secondary) and System C (Tertiary). The initial configuration is depicted in Figure 7-53 and Figure 7-54.

![Figure 7-53 Multi-site relationship configuration for our example](image)

![Figure 7-54 Multi-site relationship details for our example](image)
Upon loss of connectivity between System A and the Quorum Witness (Figure 7-55), the Multi-site relationship enters the *Compromised* state (Figure 7-56).

<table>
<thead>
<tr>
<th>System</th>
<th>Active Quorum Witness Status</th>
<th>Quorum Witness</th>
<th>Co...</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>A9000R-Site_A</td>
<td>Down (Connection Down)</td>
<td>ITSO_QW</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>A9000R-Site_B</td>
<td>OK</td>
<td>ITSO_QW</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>A9000R-Site_C</td>
<td>OK</td>
<td>ITSO_QW</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

Figure 7-55  System A lost connectivity to the active Quorum Witness

Figure 7-56  The Multi-site relationship enters the Compromised state as a result

The details for the HyperSwap leg of the Multi-site relationship show potential problems (Figure 7-57 on page 146); however at this time, Automatic Failover is still possible.

Figure 7-57  The Hyperswap leg of the Multi-site relationship shows potential problems, because of the loss of connectivity between System A and the Quorum Witness
If System A were to fail now, the Automatic Failover would occur as shown in Figure 7-58 and Figure 7-59.

Figure 7-58  Multi-site relationship configuration after System A failure

<table>
<thead>
<tr>
<th>Local Peer</th>
<th>System</th>
<th>Remote Peer</th>
<th>Remote System</th>
<th>Synchronization status</th>
<th>Local Role</th>
<th>Multi-site Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO, FA, Volume_2</td>
<td>A9000R-Site_B</td>
<td>TS0, FA, Volume_2</td>
<td>A9000R-Site_A</td>
<td>Automatic failover was performed on the secondary Peer (Link Down)</td>
<td>Primary (not in Sync)</td>
<td>Compromised</td>
</tr>
<tr>
<td>ISO, FA, Volume_2</td>
<td>A9000R-Site_C</td>
<td>ITSO, FA, Volume_2</td>
<td>A9000R-Site_A</td>
<td>Inactive (Link Down)</td>
<td>Tertiary</td>
<td>Compromised</td>
</tr>
<tr>
<td>ISO, FA, Volume_2</td>
<td>A9000R-Site_B</td>
<td>ITSO, FA, Volume_2</td>
<td>A9000R-Site_D</td>
<td>PPO-DK</td>
<td>Primary (not in Sync)</td>
<td>Compromised</td>
</tr>
<tr>
<td>ISO, FA, Volume_2</td>
<td>A9000R-Site_C</td>
<td>ITSO, FA, Volume_2</td>
<td>A9000R-Site_B</td>
<td>PPO-DK</td>
<td>Tertiary</td>
<td>Compromised</td>
</tr>
</tbody>
</table>

Figure 7-59  Multi-site relationship details after System A failure

The HyperSwap relationship details panel shows that System B has assumed the Primary role while System A is down (Figure 7-60 on page 147).

Figure 7-60  HyperSwap relationship details after the Automatic Failover occurred
It is interesting to note that when the Automatic Failover is performed, the Active and Standby Asynchronous mirroring legs of the Multi-site relationship are switched - the mirroring between System A and System C is deactivated and becomes Standby (Figure 7-61), while the mirroring between System B (now temporarily Primary) and System C is activated, ensuring replication of data to the DR site (Figure 7-62).

When the failure situation is resolved (System A comes back online, and connectivity with both System B, System C and the Quorum Witness is restored) the Multi-site relationship enters a role conflict scenario (Figure 7-63), which needs to be resolved as indicated in 7.1.2, “Multi-site relationship recovery after System A failure” on page 119.
7.5.2 Loss of connection between both Primary and Secondary systems and the Quorum Witness

The initial configuration for this example is the same as the previous, seen in Figure 7-53 on page 145 and Figure 7-54 on page 145. However, in this case both System A (Primary) and System B (Secondary) lose connectivity with the Quorum Witness. This situation compromises the HyperSwap leg of the Multi-site relationship, preventing Automatic Failure in case of subsequent failures (Figure 7-64 on page 149).

If System A (Primary) then fails, the Multi-site relationship enters a Compromised state as shown in Figure 7-65, in which host IO is temporarily blocked (since A is down, while System B could not assume the Primary role as it did not hold the quorum).
In order to restore operations on the volume, a manual role change is needed first. This change can be performed by selecting **Failover/Recovery → Change Role** in the relationship’s **Actions** menu (Figure 7-66 on page 150). System B can be given the role of Primary while System A is down (Figure 7-67 on page 150).
Figure 7-68 depicts the relationship details after the role change.

![Multi-site relationship details after the manual role change](image)

Note that in this case, since the Failover was manual and not automatic, the Asynchronous relationship between System B and System C was not automatically activated, as shown in Figure 7-69 on page 151. This activation, however, can be done manually as well after the role change, by using the **Configuration → Change Activation** option in the relationship's **Actions** menu (Figure 7-70 and Figure 7-71).

![Details for the Asynchronous relationship between System B and System C](image)
The activation allows for data to be asynchronously replicated from System B to the DR site (System C) while System A is down.

When all failures are resolved, a manual role switch can be performed to return to the original configuration (System A - Primary, System B - Secondary, Active Asynchronous mirroring between System A and System C, Standby Asynchronous mirroring between System B and System C).
Using VMware Site Recovery Manager and HyperSwap

This chapter explains how to use VMware Site Recovery Manager (SRM) in conjunction with the HyperSwap feature for FlashSystem A9000 and A9000R. The solution is enabled by the IBM Spectrum Accelerate Family's Storage Replication Adapter (SRA) Version 3.0.0 or later.

Important: VMware Cross vCenter vMotion is a feature that allows virtual machines (VMs) to switch from one vCenter Server instance to another to change the compute, network, storage and management of the VM concurrently. This is a critical underlying component of SRM. This capability allows SRM to do a "live migration" of VMs from the primary site to a secondary site with no outage or downtime. This was introduced in SRM version 6.1

Several requirements must be in place for a Virtual Machine (VM) to migrate across vCenters.

- Both the original and destination vCenter Server instances and ESXi servers must be running vSphere version 6.0 or later.
- Both vCenter Server instances need to be on the same single sign-on (SSO) domain.

By using stretched storage between the sites, as in IBM HyperSwap technology, the VMs can be live migrated between these sites.

VMware will use IBM Spectrum Accelerate Family's Storage Replication Adapter (SRA) to do a vMotion to migrate running VM's from the production site to the recovery site for planned failover. (Version 3.0.0 added support for IBM HyperSwap, as introduced by the FlashSystem A9000 and A9000R storage systems (software Version 12.1 or later). During an unplanned failover, where the primary site is down, VMware will still use similar technologies as with a planned outage failover.

The benefits of using IBM HyperSwap to VMware mirrors are an increased speed of recovery, including expedited sub-steps, and the SRA will also update site preferences automatically. In practical terms HyperSwap will dramatically reduce failover times and make the entire process far more streamlined.
Please see Figure 8-1 on page 154, below, for the overview of the solution.

**Figure 8-1  SRM Overview with IBM FlashSystem A9000/R HyperSwap Replication**

For HyperSwap, if there is no target connectivity between the individual peers, and if the recovery storage system has the volumes in secondary role, the IBM Spectrum Accelerate Family SRA cannot perform failover from the protected to recovery site. Make sure that none of the names of both storage systems have (been) changed after the mirroring targets have been defined, otherwise both arrays will not pair in SRM.

**Note:** The name of the target system may be changed by using the Hyper-Scale Manager or by using xcli command: `target_rename`.

This chapter covers the following topics:
- Prerequisites
- Installing IBM Spectrum Accelerate Family SRA
- Configuring SRM

**Note:** For the HyperSwap solution, a synchronous remote mirroring pairing must be used. For information about installing and configuring SRM, refer to *Site Recovery Manager Installation and Configuration*, which you can find at the VMware Site Recovery Manager Documentation website.
8.1 Prerequisites

- The source and destination vCenter Server instances and ESXi hosts must be running version 6.0 or later. See the following VMware article:
  https://kb.vmware.com/s/article/2106952

- IBM SRA Version 3.0 or later, installed on each vCenter server.

- VMware SRM Version 6.1 or later, installed. Each vCenter server must have its own SRM server.

- Both vCenter Server instances need to be on the same single sign-on domain. For a list of supported topologies, refer to this VMware article:
  https://kb.vmware.com/s/article/2147672

- VMware’s Cluster Mapping is engaged by SRM to list all the hosts defined inside the A9000 /R cluster. A 1:1 mapping between the A9000 /R Cluster (of volumes) and the Vsphere cluster is needed. This means that any host not part of the Vsphere cluster, but is defined as part of the A9000 / R cluster might interfere with the VMware’s Cluster mapping, which may then at some point, interfere with SRM failover operations.

- Volumes at the protected site should be marked as “Primary” inside the A9000 /R Mirroring relationship. Volumes can also be part of A9000/R Consistency Groups.

- VMware Storage Tags and Policies must be defined before creating SRM definitions. Please see Figure 8-2 on page 155, below for an brief overview.
8.2 Installing IBM Spectrum Accelerate Family SRA

IBM Spectrum Accelerate Family SRA is a software add-on that integrates VMWare SRM to run failovers together with supported storage systems. The IBM Spectrum Accelerate Family SRA extends SRM capabilities and uses HyperSwap as part of the SRM comprehensive Disaster Recovery Planning (DRP) solution. VMWare administrators can automate the failover of HyperSwap volumes at the protected SRM site to a recovery SRM site. Immediately upon a failover, the VMWare ESX/ESXi servers at the recovery SRM site initiate the replicated datastores on the HyperSwap volumes.

The IBM Spectrum Accelerate Family SRA Version 3.0 and later is available for download at the VMware website.

IBM Spectrum Accelerate Family SRA Version 3.0 works with the SRM instance at the protected and recovery site. Therefore, it needs to be installed on the same server where the SRM instance is installed at both the protected and recovery sites.

8.3 Configuring tags

Before Configuring SRM, the VMware administrator must create and assign VMware tags to the resources that will be configured in the SRM relationship. These tags should be created before the Storage Policy is created. This overview is shown in Figure 8-3 on page 156.

**Figure 8-3  Initial Creation of VMware Vcenter Tag**

After the tag has been created, it then needs to be assigned to the vCenter resources involved, as shown in Figure 8-4 on page 157.
VMware SRM & IBM HyperSwap configuration: Define a HyperSwap Tag for the Storage Policy

Figure 8-4  Assigning the created tag, inside Vcenter
8.4 Configuring a VMware Storage Policy

After the appropriate tags have been created, the next step is to create a storage policy for SRM to use for its failover, and fail-back steps. From the VMware vSphere Web Client Home page, select VM Storage Policies, which opens the dialog shown in Figure 8-5.

![Figure 8-5 VMware Vcenter Storage Policy creation for IBM HyperSwap replication and protection](image-url)
8.5 Configuring SRM

The following information assumes that the SRM was configured to the point of assigning tags to datastores and Storage Policies and that the site to site pairing is already configured.

Important: The solution will not work unless both vCenter Servers are configured with a single sign-on domain, i.e. same PSC (platform service controller).

For information about setting up tags, Storage Policies, site pairing, and resource mappings, reference VMware publications Site Recover Manager Installation and Configuration, and Site Recovery Manager Administration, which is available at the VMware website.

After initial configuration is completed, the array manager must be configured so that SRM can discover the HyperSwap devices, compute datastores, and initiate storage operations.

To configure the Array Based Replication use the following steps:

1. In the vSphere web client, click the Site Recovery plug-in. Then click Array Based Replication, then Add Array Manager.
2. Select the Add a pair of array managers option and click Next.
3. Select the paired site relationship to use in the array manager wizard and click Next.
4. Select the IBM Spectrum Accelerate Family SRA displayed in the wizard for both the array managers and click Next.
5. On the Configure array manager page, specify a display name for the site that is displayed, in this example, the recovery site was the first site displayed in the wizard. Enter the management IP addresses for the displayed FlashSystemA9000 or A9000R site. Provide a username and password to access the FlashSystem A9000 or A9000R, and click Next, as shown in Figure 8-6.

Figure 8-6   Defining recovery array manager

6. On the Configure paired array manager, specify a display name for the second displayed site. In this example the second site was the protected site. Enter the management IP addresses for the displayed site FlashSystem A9000 or A9000R. Provide
a username and password to access the FlashSystem A9000 or A9000R, and click Next, as shown in Figure 8-7.

![Figure 8-7 Defining the protected array manager](image)

7. On the **Enable array pairs** page, select the FlashSystem A9000 or A9000R array pair to be enabled and click Next, as shown in Figure 8-8.

![Figure 8-8 Selecting the array pair](image)

8. Review the content displayed in the wizard, and click **Finish**, as shown in Figure 8-9.
The Array Based Replication For both protected and recovery sites is displayed as \textit{OK}, as shown in Figure 8-10.

8.5.1 Create storage policy based protection group

A storage policy-based protection group enables automated protection of virtual machines that are associated with that storage policy. When a virtual machine is associated with or disassociated from a storage policy, SRM automatically protects or unprotects it.

To configure a protection group, use the following steps:

1. In the vSphere web client, click the Site Recovery plug-in, from there click Protection Group, then Create a Protection Group to start a new protection group.
2. In the Create Protection Group wizard enter a name description and location of the protection group then click OK, as shown in Figure 8-11.
3. In the **Protection group type** page, specify the required direction of protection and ensure to only select **Storage policies (array-based replication)**, as shown in Figure 8-12.

![Figure 8-11 Defining the protection group name](image1)

![Figure 8-12 Setting the direction and type of protection group](image2)
4. In the **Storage policies** page, select the storage policies to be used for the protection group and click **Next**, as shown in Figure 8-13.

![Figure 8-13 Selecting the protection groups storage policies](image)

5. Review the protection group settings and click **Finish** to create a storage policy-based protection group, as shown in Figure 8-14.

![Figure 8-14 Verifying the protection group configuration](image)
8.5.2 Create a recovery plan

In VMware’s Site Recovery Manager, a recovery plan is similar to an automated run book. It controls every step of the recovery process, including the order in which the SRM powers on and off the virtual machines, as well as, the network addresses that the recovered virtual machines use, and so on. Recovery plans are flexible and customizable. A recovery plan includes one or more protection groups and a protection group can be included in multiple recovery plans. For example, one recovery plan can be created to handle a planned migration of services from the protected site to the recovery site, and another plan to handle an unplanned event such as a power failure or natural disaster.

After configuring a protection group, a recovery plan must be created and tested. Perform the following steps to create a recovery plan:

1. In the vSphere web client click the Site Recovery plug-in, then click Recovery Plans, and then Create a Recovery Plan to start creating a new recovery plan.

2. In the Create Recovery Plan wizard enter a name, description, and location for a recovery plan then click Next, as shown in Figure 8-15.

![Figure 8-15 Defining the recovery plan name](image)

3. Select the site which the virtual machines in the recovery plans will recover to and click Next, as shown in Figure 8-16.

![Figure 8-16 Selecting the recovery site](image)
4. In the Protection groups page, select Storage policy protection groups as the group type from the drop-down list. Select the protection group created earlier and click Next, as shown in Figure 8-17.

![Figure 8-17 Selecting the protection groups](image1)

5. Select the networks that will be used while running the recovery plan and click Next, as shown in Figure 8-18.

![Figure 8-18 Selecting the test networks](image2)

6. Verify the recovery plan settings and click Finish to create a recovery plan, as shown in Figure 8-19.

![Figure 8-19 Verifying the recovery plan](image3)
7. After it successfully completes, the recovery plan will show a status of **Ready**, as shown in Figure 8-20.

![Figure 8-20 The plan is in a ready state](image)

### 8.5.3 Test a recovery plan

After a recovery plan is created, it must be tested to ensure that the virtual machines in the protection group are correctly recovered across the recovery sites. The tested recovery plans make the environment ready for disaster recovery situations. If the recovery plan is not tested, the actual virtual machines might not be recovered in case of a disaster, which can result in data loss. Testing a recovery plan runs almost every aspect of the recovery scenario.

In the array-based replication feature of SRM, a snapshot of the HyperSwap volume hosting the virtual machines disk file will be created while testing a recovery plan. Snapshots are removed upon cleanup.

**Note:** It is strongly advised to test all recovery plan for a planned migration and disaster recovery situation.

To test a recovery plan perform the following steps:

1. In the vSphere web client, click the **Site Recovery** plug-in. Then select a recovery plan and click the **Test Recovery Plan** icon, as shown in Figure 8-21.

![Figure 8-21 Clicking the Test Recovery Plan icon](image)
2. Confirm the protected and recovery sites and click **Next**, as shown in Figure 8-22.

![Figure 8-22 Confirming the protected and recovery sites](image1)

3. Click **Finish** to start the recovery plan test, as shown in Figure 8-23.

![Figure 8-23 Verifying the recovery plan and finishing](image2)
4. After completing the test, on the Monitor tab, check the status of the operation for any failures and warnings. If the Test Recovery Plan returns **Test Complete**, with no warnings, then the recovery plan is ready for a planned migration or disaster recovery operation, as shown in Figure 8-24.

![Figure 8-24 Successful test completion](image)

5. After successfully running a recover plan test click the **Cleanup Recovery Plan** icon to return the recovery plan to a Ready state. This cleanup operation is necessary before running the recovery plan or running a failover, as shown in Figure 8-25.

![Figure 8-25 Clicking the Cleanup Recovery Plan icon](image)
6. On the **Confirm operations** page click **Next**, as shown in Figure 8-26.

![Figure 8-26 Confirming the cleanup operation](image)

7. On the **Ready to complete** page, click **Finish**, as shown in Figure 8-27.

![Figure 8-27 Confirming and starting the cleanup operation](image)

### 8.5.4 Fail over a recovery plan

VMware Site Recovery Manager provides the following types of recovery options:

- In case of *planned migration*, the recovery plan recovers the virtual machines while both sites are running. If an error occurs at the protected site during recovery, planned migration will fail.

- In case of *disaster recovery*, the recovery plan recovers the virtual machines to the recovery site when the protected site encounters a problem. If an error occurs at the protected site, disaster recovery will continue and will not fail.

After testing the recovery plan successfully, the recovery plan is ready for planned migration or disaster recovery operations. Before running a failover, resolve any errors or warnings in the *Test Recovery Plan* operation.
Before running a recovery plan, the virtual machine, or machines to be failed need to reside at the protected site, as shown in Figure 8-28.

Figure 8-28  VM location prior to recovery operation

At this point, the HyperSwap volume is on FlashSystem A9000 designated as the secondary. When the migration is completed, it will show as the primary, as illustrated in Example 8-1.

Example 8-1  HyperSwap volume information about the FlashSystem A9000 prior to recovery operation

A9000>>ha_list vol=ITSO_HA_VOLUME3 -x

<XCLIRETURN STATUS="SUCCESS" COMMAND_LINE="ha_list vol=ITSO_HA_VOLUME3 -x">
  <OUTPUT>
    <ha id="5dc19b00006">
      <creator value="admin"/>
      <creator_category value="storageadmin"/>
      <id value="5dc19b00006"/>
      <domain_uid value="-1"/>
      <local_peer_id value="5881920005b"/>
      <local_peer_name value="ITSO_HA_VOLUME3"/>
      <designation value="Secondary"/>
      <current_role value="Slave"/>
      <remote_mirror_id value="6e61a800006"/>
      <remote_peer_name value="ITSO_HA_VOLUME3"/>
      <target_id value="59719000000"/>
      <target_name value="A9000R"/>
      <sync_state value="Synchronized"/>
      <active value="yes"/>
      <ha_connected value="yes"/>
      <operational value="yes"/>
      <sync_progress value="100"/>
      <size_to_synchronize value="-1"/>
      <estimated_sync_time value="0"/>
      <mirror_error value="No_Error"/>
      <ha_object value="Volume"/>
      <init_type value="online"/>
    </ha>
  </OUTPUT>
</XCLIRETURN>
Complete the following steps:

1. In the vSphere web client, select the Site Recovery plug-in. Then select the recovery plan and click the Run Recovery Plan icon, as shown in Figure 8-29.
2. Select the option to provide recovery confirmation and select the recovery type. In this example, Planned migration is used. If planned migration is selected, ensure that Enable vMotion of eligible VMs is also selected, as shown in Figure 8-30.

![Figure 8-30  Defining the recover options](image)

3. In the Ready to complete page, verify all the settings and click Finish to run the recovery plan, as shown in Figure 8-31.

![Figure 8-31  Verifying the recovery options](image)
4. When the recovery plan completes successfully, the virtual machine or machines are now shown at the recovery site (Figure 8-32).

![Figure 8-32 VM location after recovery operation](image)

At this point, the HyperSwap volume will show on the A9000 as primary, as illustrated in Example 8-2.

**Example 8-2 HyperSwap volume information about the FlashSystem A9000 after recovery**

```
A9000>> ha_list vol=ITSO_HA_VOLUME3 -x
<XCLIRETURN STATUS="SUCCESS" COMMAND_LINE="ha_list vol=ITSO_HA_VOLUME3 -x">
<OUTPUT>
  <ha id="5dc19b00006">
    <creator value="admin"/>
    <creator_category value="storageadmin"/>
    <id value="5dc19b00006"/>
    <domain_uid value="-1"/>
    <local_peer_id value="5881920005b"/>
    <local_peer_name value="ITSO_HA_VOLUME3"/>
    <designation value="Primary"/>
    <current_role value="Master"/>
    <remote_mirror_id value="6e61a800006"/>
    <remote_peer_name value="ITSO_HA_VOLUME3"/>
    <target_id value="59719000000"/>
    <target_name value="A9000R"/>
    <sync_state value="Synchronized"/>
    <active value="yes"/>
    <ha_connected value="yes"/>
    <operational value="yes"/>
    <sync_progress value="100"/>
    <size_to_synchronize value="-1"/>
    <estimated_sync_time value="0"/>
    <mirror_error value="No_Error"/>
    <ha_object value="Volume"/>
    <init_type value="online"/>
    <crash_consistent value="Consistent"/>
  </ha>
</OUTPUT>
```
8.5.5 Reprotect virtual machines after recovery

After recovery is complete, the recovery site becomes the primary site, but the virtual machines are not protected. Manually re-establishing protection in the opposite direction by re-creating all protection groups and recovery plans is time consuming and prone to errors. SRM provides the reprotect function, which is an automated way to revert protection and automating all of the above.

By running `reprotect`, when the protected site comes back online, the direction of replication can be reversed to protect the recovered virtual machines on the recovery site back to the original protected site. The reprotect operation can be initiated only after successful recovery without any errors.

The reprotect operation reverses the direction of protection, and then forces the synchronization of virtual machines from the new protected site to the new recovery site.

To perform a reprotect operation, use the following steps:

1. In the vSphere web client, click the Site Recovery plug-in. Then select a recovery plan and click the Reprotect Recovery Plan icon, as shown in Figure 8-33.

![Figure 8-33 Clicking the Reprotect Recovery Plan icon](image_url)
2. Provide the reprotect confirmation and click **Next**, as shown in Figure 8-34.

![Figure 8-34  Confirming the reprotect operation options](image)

3. Verify the information in the **Ready to complete** page and click **Finish** to reprotect the recovery plan, as shown in Figure 8-35.

![Figure 8-35  Verifying the reprotect operation options](image)

4. After the recovery operation is completed, and the virtual machine or machines are reprotected, run the test recovery to ensure that the new configuration of protected and recovery sites completes successfully.
Chapter 9. Microsoft Multi-Site Failover Clustering and HyperSwap

This chapter shows how to use Microsoft Failover Clustering between two sites using HyperSwap for FlashSystem A9000 and A9000R.

This chapter covers the following topics:
- Microsoft Multi-Site Failover Clustering configuration
- Planned and unplanned failovers
9.1 Microsoft Multi-Site Failover Clustering configuration

IBM FlashSystem A9000 and A9000R supports Microsoft Failover Clustering, starting with Windows Server version 2008. For more details about the supported versions, see the System Storage Interoperability Center (SSIC).

For information about installing and configuring Microsoft Failover Clustering using Windows 2008 R2, see the Microsoft publication Failover Clustering Overview.

Note: The following illustrations are based on Microsoft SQL Server 2012 Failover Cluster running on Windows 2008 R2, Hyper-Scale Manager 5.4, and FlashSystem A9000 and A9000R Version 12.2.1.

If the hosts on FlashSystem A9000 and A9000R are defined by Host Attachment Kit, the host type Windows 2008 is automatically chosen. Make sure that for host and cluster type, Windows 2008 on FlashSystem A9000 and A9000R is selected. In the Hyper-Scale Manager UI, select HOST & CLUSTERS VIEWS → Clusters to show the type, as shown in Figure 9-1. If another host and cluster type is chosen, it might result in a stop error on one Microsoft cluster node after a failover.

![Figure 9-1 Cluster on FlashSystem A9000 and A9000R](image)

The HyperSwap volumes are configured and mapped to the cluster, and an SQL Server 2012 Failover Cluster is running on one of these volumes.

Volume ITSO_Cluster_Data holds the SQL server data and ITSO_Cluster_Quorum is the Microsoft Cluster Quorum Disk Witness. Select POOLS & VOLUMES VIEWS → Volumes and filter for the volumes as depicted in Figure 9-2.

![Figure 9-2 Failover Cluster HyperSwap volumes](image)
To list the HyperSwap volumes on one of the cluster nodes, open a command prompt or a PowerShell window and run the `xiv_devlist.exe` command, as shown in Example 9-1.

**Example 9-1  FlashSystem A9000 and A9000R volumes on Cluster node ITSO-WIN2**

```
C:\Users\Administrator.ITSO> xiv_devlist
IBM storage devices

<table>
<thead>
<tr>
<th>Device</th>
<th>Size (GB)</th>
<th>Paths</th>
<th>Vol Name</th>
<th>Vol ID</th>
<th>Storage ID</th>
<th>Storage Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>\PHYSICALDRIVE1</td>
<td>500.0</td>
<td>8/8</td>
<td>ITSO_Cluster_Data</td>
<td>27130</td>
<td>1322131</td>
<td>FlashSystem A9000</td>
</tr>
<tr>
<td>\PHYSICALDRIVE2</td>
<td>10.0</td>
<td>8/8</td>
<td>ITSO_Cluster_Quorum</td>
<td>27131</td>
<td>1322131</td>
<td>FlashSystem A9000</td>
</tr>
</tbody>
</table>

Non-IBM storage devices

<table>
<thead>
<tr>
<th>Device</th>
<th>Size (GB)</th>
<th>Paths</th>
</tr>
</thead>
<tbody>
<tr>
<td>\PHYSICALDRIVE0</td>
<td>199.0</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Both disks are shown with their assignment in the Failover Cluster Manager, as illustrated in Figure 9-3.

![Failover Cluster Manager Disks](image)
The SQL Server is running on Node WINDOWS-OPZX5FL, as depicted in Figure 9-4.

**Tip:** To speed up path failover, change two registry values on all Microsoft cluster nodes. In HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\services\mpio\Parameters, set the parameter PathVerifyEnabled to 1 and PathVerificationPeriod to 5, as illustrated in Figure 6-5. These two changes allow faster path failure detection, but are not mandatory. More details about these parameters can be found at the following web page:

9.1.1 File Share Witness or Disk Witness

Microsoft’s recommendation for Multi-Site failover clustering is to use Node and File Share Majority (file share witness) as quorum option, which requires a file share on a third site. But with FlashSystem A9000 and A9000R HyperSwap volumes, the use of Node and Disk Majority (disk witness) as quorum option is even better because the HyperSwap volumes are available to both sites. In split-brain situations with file share witness, that Microsoft cluster might decide that one site is the surviving site while FlashSystem A9000 and A9000R decide the other site is the surviving site. This situation can lead to an access loss. More details about the requirements and recommendations from Microsoft for a Multi-Site Failover Cluster can be found at Requirements and Recommendations for a Multi-Site Failover Cluster.

The scenarios in 9.2, “Planned and unplanned failovers” use a disk witness for the Microsoft failover cluster.

9.2 Planned and unplanned failovers

This section describes the following failover scenarios for IBM HyperSwap on FlashSystem A9000 and FlashSystem A9000R with Microsoft Multi-Site failover clustering:

- Planned failover
- All Fibre Channel connections lost on the owner node
- Site failure

9.2.1 Planned failover

You can use both IBM Hyper-Scale Manager and Microsoft Failover Cluster Manager to make it simple to move a clustered application to another site at metro distance. The following steps must be performed:

1. To fail over the HyperSwap volume from primary to secondary site, right-click the HyperSwap volumes and select **HyperSwap → Switch Roles**, as depicted in Figure 9-6.

![Figure 9-6 Switch Roles](image)
2. Select **Switch the Volume designation as well** and click **Apply** to confirm the action, as illustrated in Figure 9-7.

![Figure 9-7  Prompt to confirm role switch for HyperSwap (Primary to Secondary)](image1.png)

3. The HyperSwap volumes are switched, as shown in Figure 9-8.

![Figure 9-8  Roles switched](image2.png)

4. In the Failover Cluster Manager, right click the application and select **Move this service or application to another node** → **Move to node windows-oub6s4p**, as shown in Figure 9-9.

![Figure 9-9  Failover Cluster Manager: Failover to other node](image3.png)
5. Click **Move SQL Server (MSSQLSERVER) to windows-oub6s4p** as depicted in Figure 9-10.

![Figure 9-10 Failover Cluster Manager: Confirm Move of application](image)

6. Application has moved to node windows-oub6s4p, as illustrated in Figure 9-11.

![Figure 9-11 Failover Cluster Manager: Application moved](image)

To return to the initial configuration, just execute the same sequence of steps, using the other storage system and other cluster node.
9.2.2 All Fibre Channel connections lost on the owner node

After the planned failover, the application resides on Site B. Assume that all Fibre Channel connections to the owner node (windows-oub6s4p) of the SQL Server Failover Cluster are lost, as depicted in Figure 9-12.

Using `xiv_devlist.exe` does not list the volumes anymore, as shown in Example 9-2.

**Example 9-2 FlashSystem A9000 and A9000R volumes on Cluster node windows-oub6s4p after failure**

```
PS C:\Users\Administrator.ITSO> xiv_devlist.exe
IBM storage devices
Device  Size (GB)  Paths  Vol Name  Vol ID  Storage ID  Storage Type  Hyper-Scale Mobility
Non-IBM storage devices
Device  Size (GB)  Paths
\\ PHYSICALDRIVE0  200.0  N/A
```

Figure 9-12  FC connection loss scenario
The SQL Server starts to fail over to the other cluster node WINDOWS-0PZX5FL, as shown in Figure 9-13.

After the failover finishes, the SQL Server is running on WINDOWS-0PZX5FL, as can be seen in Figure 9-14.
9.2.3 Site failure

Assume a site failure of the owner node of the SQL server (WINDOWS-OPZX5FL) and FlashSystem A9000 where the primary volume resides, as depicted in Figure 9-15.

Only two paths connected to FlashSystem A9000R on the surviving node (windows-oub6s4p) are left, as shown in Example 9-3.

Example 9-3  FlashSystem A9000 volumes on Cluster node windows-oub6s4p after site failure

```
PS C:\Users\Administrator.ITSO> xiv_devlist.exe
IBM storage devices
-------------------------------------------------------------------------------------------------
Device              Size (GB)  Paths  Vol Name             Vol ID  Storage ID  Storage Type
-------------------------------------------------------------------------------------------------
\\.\PHYSICALDRIVE1  500.0      2/2    ITSO_Cluster_Data    27130   1320902     FlashSystem A9000R
-------------------------------------------------------------------------------------------------
\\.\PHYSICALDRIVE2  10.0       2/2    ITSO_Cluster_Quorum  27131   1320902     FlashSystem A9000R
-------------------------------------------------------------------------------------------------
Non-IBM storage devices
-----------------------------------------------------------------------------------------------
Device              Size (GB)  Paths
-----------------------------------------------------------------------------------------------
\\.\PHYSICALDRIVE0  200.0      N/A
-----------------------------------------------------------------------------------------------
```
After the Microsoft cluster failover completes, the SQL Server is running on `windows-oub6s4p` as illustrated in Figure 9-16.

![Figure 9-16  Failover Cluster Manager Roles: Failover finished after site failure](image)

A9000 is down and the automatic failover for the HyperSwap volumes has happened as shown in Figure 9-17.

![Figure 9-17  Automatic HyperSwap volumes failover](image)
After System A comes up again, all paths and all connections are back, but the HyperSwap volumes are not synchronized. Select REMOTE VIEWS → Mirrored/HyperSwap Volumes (Availability), as shown in Figure 9-18.

![HyperSwap volume after A9000 up again](image)

The necessary steps for the recovery after a failure are described in 5.4, “Failback scenarios” on page 76.
Volume Shadow Copy Service and HyperSwap Snapshots

This chapter discusses how to combine the use of Windows Volume Shadow Copy Service (VSS) with FlashSystem A9000 and A9000R snapshots in a HyperSwap configuration.

Microsoft first introduced VSS in Windows 2003 Server and has included it in all subsequent releases. VSS provides a framework and the mechanisms to create consistent point-in-time copies (known as shadow copies) of databases and application data. It consists of a set of Microsoft COM APIs that enable volume-level snapshots to be performed while the applications that contain data on those volumes remain online and continue to write. This approach enables third-party software like IBM Spectrum Protect™ Snapshot or IBM Spectrum Copy Data Management to centrally manage the backup and restore operations.

You can find more details about VSS at the Microsoft TechNet website.

You can find details about Spectrum Protect Snapshot at the IBM Spectrum Project website.

For more details about Spectrum Copy Data Manager refer to Using IBM Spectrum Copy Data Management with IBM FlashSystem A9000 or A9000R and SAP HANA, REDP-5439.

This chapter covers the following topics:
- VSS product and components
- VSS function
- IBM Spectrum Accelerate family VSS Provider (xProv)
10.1 VSS product and components

Without VSS, if you do not have an online backup solution implemented, you either must stop or quiesce applications during the backup process or live with the side effects of an online backup with only crash-consistent data and open files that could not be backed up.

With VSS, you can produce application-consistent shadow copies by coordinating tasks with business applications, file system services, backup applications, fast recovery solutions, and storage hardware such as the FlashSystem A9000 or A9000R.

VSS enables you to perform online backup of applications, which otherwise is not possible, and is supported on the FlashSystem A9000 and A9000R. VSS accomplishes this backup by facilitating communications between the following entities:

► Requestor
   An application that requests that a volume shadow copy be taken. These are applications, such as backup (like Spectrum Protect Snapshot or Spectrum Copy Data Management) or storage management, that request a point-in-time copy of data or a shadow copy.

► Writer
   A component of an application that stores persistent information about one or more volumes that participate in shadow copy synchronization. Writers are software that is included in applications and services to help provide consistent shadow copies. Writers serve the following main purposes:
   – Responding to signals provided by VSS to interface with applications to prepare for shadow copy.
   – Providing information about the application name, icons, files, and a strategy to restore the files.

Writers prevent data inconsistencies. A database application (such as SQL Server or Exchange Server) or a system service (such as Active Directory) can be a writer.

► Provider
   A component that creates and maintains the shadow copies. A provider can be implemented in the software or in the hardware. For FlashSystem A9000 and A9000R, you must install and configure the IBM Spectrum Accelerate Family Provider for Microsoft Windows Volume Shadow Copy Service.
Figure 10-1 shows the Microsoft VSS architecture and how the software provider and hardware provider interact through Volume Shadow Copy Services.

VSS uses the following terminology to characterize the nature of volumes participating in a shadow copy operation:

- **Persistent**
  A shadow copy that remains after the backup application completes its operations. This type of shadow copy also survives system reboots.

- **Non-persistent**
  A temporary shadow copy that remains only while the backup application needs it to copy the data to its backup repository.

- **Transportable**
  A shadow copy volume that is accessible from a secondary host so that the backup can be off-loaded. Transportable is a feature of hardware snapshot providers. You can mount a snapshot volume to another host.

- **Source volume**
  The volume that contains the data to be shadow copied. These volumes contain the application data.

- **Target or snapshot volume**
  The volume that retains the shadow-copied storage files. It is an exact copy of the source volume at the time of backup.
VSS supports the following shadow copy methods:

- **Clone (full copy/split mirror)**
  A clone is a shadow copy volume that is a full copy of the original data as it resides on a volume. The source volume continues to take application changes while the shadow copy volume remains an exact read-only copy of the original data at the point-in-time that it was created.

- **Copy-on-write (differential copy)**
  A copy-on-write shadow copy volume is a differential copy (rather than a full copy) of the original data as it resides on a volume. This method makes a copy of the original data before it is overwritten with new changes. Using the modified blocks and the unchanged blocks in the original volume, a shadow copy can be logically constructed that represents the point-in-time at which it was created.

- **Redirect-on-write (differential copy)**
  A redirect-on-write shadow copy volume is a differential copy (rather than a full copy) of the original data as it resides on a volume. This method is similar to copy-on-write, without the double-write penalty, and it offers storage-space- and performance-efficient snapshots. New writes to the original volume are redirected to another location set aside for the snapshot. The advantage of redirecting the write is that only one write takes place, whereas with copy-on-write, two writes occur (one to copy original data onto the storage space, the other to copy changed data). The XIV storage system supports redirect-on-write.

### 10.2 VSS function

VSS accomplishes the fast backup process when a backup application initiates a shadow copy backup. VSS coordinates with the VSS-aware writers to briefly hold writes on the databases, applications, or both. VSS flushes the file system buffers and requests a provider to initiate an IBM FlashCopy® of the data. When the FlashCopy is logically completed, VSS allows writes to resume and notifies the requestor that the backup has completed successfully. The volumes are mounted, hidden, and for read-only purposes, to be used when rapid restore is necessary. Alternatively, the volumes can be mounted on a different host and used for application testing or backup to tape.

The steps in the VSS FlashCopy process are as follows:

1. The requestor notifies VSS to prepare for shadow copy creation.
2. VSS notifies the application-specific writer to prepare its data for making a shadow copy.
3. The writer prepares the data for that application by completing all open transactions, flushing the cache, and writing in-memory data to disk.
4. When the data is prepared for shadow copy, the writer notifies the VSS, and it relays the message to the requestor to initiate the commit copy phase.
5. VSS temporarily quiesces application I/O write requests for a few seconds and the hardware provider performs the FlashCopy on the Storage Unit.
6. After the completion of FlashCopy, VSS releases the quiesce, and database writes resume.
7. VSS queries the writers to confirm that write I/Os were successfully held during Volume Shadow Copy.
10.3 IBM Spectrum Accelerate family VSS Provider (xProv)

A VSS hardware provider, such as the xProv, is used by third-party software to act as an interface between the hardware (storage system) and the operating system. The third-party application uses Spectrum Accelerate family VSS Provider to instruct FlashSystem A9000 or A9000R to perform a snapshot of a volume attached to the host system.

Note: The following illustrations are based on Windows 2012 R2, Hyper-Scale Manager 5.2, and FlashSystem A9000 and A9000R Version 12.1.

10.3.1 Install the VSS Provider

This section illustrates how to install the VSS Provider. At the time of writing, the IBM Spectrum Accelerate family VSS Provider Version 2.9.0 was available. To obtain the system requirements, see the IBM Spectrum Accelerate Family Provider for Microsoft Windows Volume Shadow Copy Service Release Notes, which includes a chapter about the system requirements. Download the Spectrum Accelerate family VSS Provider user guide and release notes from IBM Fix Central.

Verify that the prerequisites are met and the following packages are installed:

- Microsoft Visual C++ 2012 Redistributable-64
- Microsoft Visual C++ 2012 Redistributable-68

Installing the VSS Provider is a straightforward Windows application installation:

1. Locate the Spectrum Accelerate family VSS Provider installation file, also known as the xProv installation file. If the VSS Provider is downloaded from the Internet, the file name is xProvSetup_2.9.0.0-b64_for_Windows-x64.exe. Execute the file to start the installation.

2. Select the language for the installation, the Welcome window opens, as shown in Figure 10-2. Click Next.

![Figure 10-2 VSS provider installation: Welcome window](image)

3. The License Agreement window also displays. To continue the installation you must accept the license agreement.
4. Specify the VSS Provider configuration file directory and the installation directory. Keep the default directory folder and installation folder or change it to meet your needs.

5. A dialog window for post-installation operations is opened, as shown in Figure 10-3. You can perform a post-installation configuration during the installation process or at a later time. When done, click **Next**.

![Figure 10-3](Image)

**Figure 10-3** Installation: Post-installation operation

6. A Confirm Installation window is displayed. You can go back to make changes if required, or confirm the installation by clicking **Install**.

7. Click **Close** to exit after the installation is complete.

### 10.3.2 Configure VSS Provider

Configure the VSS Provider using the following steps:

1. If the post installation option was selected during installation (Figure 10-3), the MachinePoolEditor window opens automatically.

2. Right-click the Machine Pool Editor window.

3. In the dialog shown in Figure 10-4, click **New System** to open the New System window.

![Figure 10-4](Image)

**Figure 10-4** Configuration: Machine Pool Editor
4. The Add System Management window shown in Figure 10-5 displays. Enter the user
name and password of a FlashSystem A9000 or A9000R user with administrator
privileges (storageadmin role) and the primary IP address of the FlashSystem A9000 or
A9000R. If the snapshot is taken of a volume that is in a mirror relationship and you want
to have the snapshot on source and target system, then select Enable Replicated
Snapshots and click Add.

![Figure 10-5 Configuration: Add Storage System](image)

5. Repeat steps 3 and 4 to add additional FlashSystem A9000 or A9000R.

6. You are returned to the VSS Machine Pool Editor window. The VSS Provider collected
additional information about the storage systems, as illustrated in Figure 10-6.

![Figure 10-6 Configuration: Machine Pool Editor](image)

7. If the snapshot needs to be taken of a volume that is in a HyperSwap relationship and the
snapshots are needed on both sites, open `C:\Program Files\IBM\IBM XIV Provider for
Microsoft Windows Volume Shadow Copy Service\etc\MachinePool.xml` and set
create_ha_snapshots="true", as shown in Example 10-1.

**Example 10-1 MachinePool.xml file**

```xml
<?xml version="1.0" encoding="us-ascii"?>
<MachinePool>
<machine serial="1322131" username="admin"
create_mirror_snapshots="false" create_ha_snapshots="true"
password="LZnLBheVIybKTxpbCkMeDQ==">
    <aserver hostname="9.155.120.218" port="7778" />
    <aserver hostname="10.0.20.108" port="7778" />
    <aserver hostname="10.0.20.109" port="7778" />
</machine>
<machine serial="1320902" username="admin"
create_mirror_snapshots="false" create_ha_snapshots="true"
password="LZnLBheVIybKTxpbCkMeDQ==">
    <aserver hostname="9.155.116.200" port="7778" />
    <aserver hostname="9.155.116.201" port="7778" />
    <aserver hostname="9.155.116.202" port="7778" />
</machine>
</MachinePool>
```
8. At this point, the VSS Provider configuration is complete and you can close the Machine Pool Editor window.

After the VSS Provider has been configured, ensure that the operating system can recognize it. To do so, launch the `vssadmin list providers` command from the operating system command line.

Make sure that IBM VSS HW Provider is in the list of installed VSS providers returned by the `vssadmin` command, as shown in Example 10-2.

**Example 10-2 Output of vssadmin command**

```plaintext
PS C:\Users\Administrator.ITSO> vssadmin list providers
vssadmin 1.1 - Volume Shadow Copy Service administrative command-line tool
(C) Copyright 2001-2013 Microsoft Corp.

Provider name: 'Microsoft CSV Shadow Copy Helper Provider'
  Provider type: Software
  Provider Id: {26d02d81-6aac-4275-8504-b9c6edc5261d}
  Version: 1.0.0.1

Provider name: 'Microsoft CSV Shadow Copy Provider'
  Provider type: Software
  Provider Id: {400a2ff4-5eb1-44b0-8a05-1fcac0bdf9ff}
  Version: 1.0.0.1

Provider name: 'Microsoft File Share Shadow Copy provider'
  Provider type: Fileshare
  Provider Id: {89300202-3cec-4981-9171-19f59559e0f2}
  Version: 1.0.0.1

Provider name: 'Microsoft Software Shadow Copy provider 1.0'
  Provider type: System
  Provider Id: {b5946137-7b9f-4925-af80-51abd60b20d6}
  Version: 1.0.0.7

Provider name: 'IBM XIV VSS HW Provider'
  Provider type: Hardware
  Provider Id: {d51fe294-36c3-4ead-b837-1a67b3844b1d}
  Version: 2.9.0.0
```

**10.3.3 Diskshadow command line utility**

All editions of Windows Server 2012 R2 contain a command line utility (`DiskShadow.exe`) for the creation, deletion, and restoration of shadow copies (snapshots). It is the first in-box VSS requestor that can create hardware shadow copies and one of many utilities for validating VSS operations. The tool is similar to vshadow (a tool included with the Volume Shadow Copy/VSS SDK), but has an interface similar to diskpart utility.

For more information about diskshadow, see the Microsoft TechNet website.
The steps to test the creation of a persistent snapshot of a basic disk on FlashSystem A9000 or A9000R via VSS are shown in Example 10-3. The snapshot will be automatically unlocked and mapped to the server. Assign a drive letter to the volume and access the data on the file system.

*Example 10-3  Diskshadow snapshot creation*

```
PS C:\Users\Administrator.ITSO> diskshadow
Microsoft DiskShadow version 1.0
Copyright (C) 2013 Microsoft Corporation
On computer:  ITSO-WIN1,  6/1/2017 10:49:45 AM
DISKSHADOW> set context persistent
DISKSHADOW> add volume y:
DISKSHADOW> create
COM call "lvssObject4->GetRootAndLogicalPrefixPaths" failed.
Alias VSS_SHADOW_1 for shadow ID {023be8e3-cb55-4e07-a318-d1f81d97d433} set as environment variable.
Alias VSS_SHADOW_SET for shadow set ID {008cca0e-922a-46dd-8902-e764e3ca3d17} set as environment variable.
Querying all shadow copies with the shadow copy set ID {008cca0e-922a-46dd-8902-e764e3ca3d17}

* Shadow copy ID = {023be8e3-cb55-4e07-a318-d1f81d97d433}
%VSS_SHADOW_1% - Shadow copy set: {008cca0e-922a-46dd-8902-e764e3ca3d17}
%VSS_SHADOW_SET% - Original count of shadow copies = 1
  - Original volume name: \?\Volume{1dc9da0e-45dd-11e7-80cb-6eae8b4b61ab}\[Y:\]
    - Creation time: 6/1/2017 10:50:48 AM
    - Shadow copy device name:
\?\Volume{1dc9da1c-45dd-11e7-80cb-6eae8b4b61ab}
  - Originating machine: ITSO-WIN1.ITSO.local
  - Service machine: ITSO-WIN1.ITSO.local
  - Not exposed
  - Provider ID: {d51fe294-36c3-4ead-b837-1a6783844b1d}
  - Attributes: No_Auto_Release Persistent Hardware
Number of shadow copies listed: 1
```
Select **POOLS & VOLUMES VIEWS → Snapshots** and filter for the snapshot, the snapshot with his Shadow Copy ID is visible as depicted in Figure 10-7.

![Figure 10-7 VSS snapshot](image)

### 10.3.4 Create HyperSwap VSS snapshot

Starting with xProv Version 2.9, you can create snapshots through VSS on a HyperSwap volume. Before you begin, the HyperSwap relationship must exist, be active and must be enabled in the `MachinePool.xml` file.

The VSS process will create a snapshot of the source volume and also the target volume. Use the following steps to create the snapshot on both mirror sites:

1. Add both FlashSystem A9000 and A9000R to xProv and set `create_ha_snaps`="true".
2. Set up HyperSwap for the volumes you want to use for VSS snapshot replication.
3. Run a VSS create operation on the HyperSwap volume.

To be able to import the shadow copy to the same or a different computer the transportable option must be used and the `example1.cab` is needed, as shown in Example 10-4. The import works on basic disks only.

**Example 10-4 VSS snapshot creation of a HyperSwap volume**

```bash
PS C:\Users\Administrator.ITSO> diskshadow
Microsoft DiskShadow version 1.0
Copyright (C) 2013 Microsoft Corporation
On computer:  ITSO-WIN1,  6/1/2017 2:48:31 PM

DISKSHADOW> set context persistent
DISKSHADOW> set option transportable
DISKSHADOW> set metadata c:\Users\Administrator\example1.cab
DISKSHADOW> add volume z:
DISKSHADOW> create
```
COM call "lvssObject4->GetRootAndLogicalPrefixPaths" failed.
Alias VSS_SHADOW_1 for shadow ID {ca2992c2-3f30-4025-a6d6-82d186563901} set as environment variable.
Alias VSS_SHADOW_SET for shadow set ID {e7870a5b-9b2f-4c67-a1ed-7f7e589f7c83} set as environment variable.

The snapshots created by VSS on the source and target XIV storage system are depicted in Figure 10-8.

![Figure 10-8 VSS mirrored snapshot on source and target](image)

To test the import of the data afterwards to another server, copy the example1.cab file to that server. The host and its ports must be defined on the storage systems.

The commands to load the metadata and import the VSS snapshot to the server are shown in Example 10-5. Afterwards assign a drive letter to the volume and access the data on the file system.

**Example 10-5  VSS import to another server**

```powershell
PS C:\Users\Administrator.ITSO> diskshadow
Microsoft DiskShadow version 1.0
Copyright (C) 2013 Microsoft Corporation
On computer: ITSO-WIN2, 6/1/2017 3:31:02 PM

DISKSHADOW> load metadata example1.cab
Could not find metadata file.

DISKSHADOW> load metadata C:\Users\Administrator\example1.cab
Alias VSS_SHADOW_1 for value {ca2992c2-3f30-4025-a6d6-82d186563901} set as an environment variable.
Alias VSS_SHADOW_SET for value {e7870a5b-9b2f-4c67-a1ed-7f7e589f7c83} set as an environment variable.

DISKSHADOW> import

DISKSHADOW>
```
Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this paper.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *IBM FlashSystem A9000 and IBM FlashSystem A9000R Architecture and Implementation*, SG24-8345
- *IBM Hyper-Scale Manager for IBM Spectrum Accelerate Family: IBM XIV, IBM FlashSystem A9000 and A9000R, and IBM Spectrum Accelerate*, SG24-8376
- *IBM FlashSystem A9000 and A9000R Business Continuity Solutions*, REDP-5401
- *IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System: Host Attachment and Interoperability*, SG24-8368
- *Using the IBM Spectrum Accelerate Family in VMware Environments: IBM XIV, IBM FlashSystem A9000 and IBM FlashSystem A9000R, and IBM Spectrum Accelerate*, REDP-5425

You can search for, view, download or order these documents and other Redbooks, Redpapers, web docs, draft and additional materials, at the following website:

[ibm.com/redbooks](https://ibm.com/redbooks)

Other publications

These publications are also relevant as further information sources:

- IBM FlashSystem A9000 and A9000R product overview
- IBM FlashSystem A9000 and A9000R CLI reference guide
- IBM Hyper-Scale Manager user guide

Online resources

These websites are also relevant as further information sources:

- IBM Fix Central
  
  [https://www.ibm.com/support/fixcentral](https://www.ibm.com/support/fixcentral)
- IBM Knowledge Center
  
  [https://www.ibm.com/support/knowledgecenter/](https://www.ibm.com/support/knowledgecenter/)
Help from IBM

IBM Support and downloads
ibm.com/support

IBM Global Services
ibm.com/services