ibm.com/redbooks

IBM Storwize HyperSwap with IBM i

Jana Jamsek

Falk Schneider

Storage

Redpaper

International Technical Support Organization

IBM Storwize HyperSwap with IBM i

May 2018

REDP-5490-00

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

First Edition (May 2018)

This edition applies to the software and hardware versions documented in this paper.

Contents

Notices

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

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

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

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

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

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

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

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

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

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

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

COPYRIGHT LICENSE:

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

Trademarks

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

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

DS8000®
FlashCopy®
HyperSwap®
IBM®
IBM FlashSystem®
IBM Spectrum™
IBM Spectrum Control™
POWER®
POWER7®
PowerHA®
Redbooks®
Redpaper™
Redbooks (logo) ®
Storwize®
System i®
System Storage®

The following terms are trademarks of other companies:

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

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

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

Preface

In this paper, we describe the implementation steps and business continuity solutions for IBM® Storwize® HyperSwap® and IBM i.

Authors

This paper was produced by a team of specialists from around the world.

Jana Jamsek, at the time of writing, was an IT specialist for IBM Slovenia. She worked in Storage Advanced Technical Skills for Europe as a specialist for IBM Storage Systems and IBM i systems. Jana has 8 years of experience in the IBM System i® and IBM AS/400 areas, and 13 years of experience in Storage. She has a Master's degree in computer science and a degree in mathematics from the University of Ljubljana, Slovenia. Jana worked on complex customer cases that involved IBM i and Storage systems in various European and Middle East countries. She has presented at IBM Storage and Power universities, and she ran workshops for IBM employees and customers. She is the author or co-author of many IBM publications in this area.

Falk Schneider is an IT Specialist working for the IBM Systems Lab Services team at the European Storage Competence Center (ESCC), Germany. Falk has over 15 years of experience in IT. He has worked in different roles, such as IBM Systems Lab services consultant, Advanced Technical Support (ATS), and Proof of Concepts (PoC) at the IBM lab in Mainz. He also has many years of field experience supporting clients with storage and SAN products. His areas of expertise include disaster recovery (DR), IBM Storage virtualization, and IBM FlashSystem®.

This project was managed by:

Jon Tate is a Project Manager for IBM System Storage® Virtualization and SAN Solutions at the ITSO, San Jose Center. Before joining the ITSO in 1999, he worked in the IBM Technical Support Center, providing Level 2/3 support for IBM mainframe storage products. Jon has 32 years of experience in storage software and management, services, and support. He is an IBM Certified IT Specialist, an IBM SAN Certified Specialist, and is Project Management Professional (PMP) certified. He is also the UK Chairman of the Storage Networking Industry Association (SNIA).

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:

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:

  ibm.com/redbooks

- Send your comments in an email:

  redbooks@us.ibm.com

- Mail your comments:

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

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

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

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

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

- Stay current on recent Redbooks publications with RSS Feeds:

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

1

Introduction

IBM Storwize HyperSwap is a response to increasing demand for continuous application availability, minimizing downtime in the event of an outage, and non-disruptive migrations.

IT centers with IBM i can take full advantage of the HyperSwap solution.

In this IBM Redpaper™ publication, we provide instructions to implement Storwize HyperSwap with IBM i. We also describe some business continuity scenarios in this area, including solutions with HyperSwap and IBM i Live Partition Mobility, and a solution with HyperSwap and IBM PowerHA® for i.

1.1 IBM Storwize HyperSwap overview

The IBM HyperSwap function is a high availability feature that provides dual-site, active-active access to a volume. It provides continuous data availability in case of hardware failure, power failure, connectivity failure, or disasters. HyperSwap capabilities are also available on other IBM storage technologies that can support more than one I/O group (for example, Storwize V5030 systems) and also IBM FlashSystem V9000 and A9000.

This feature was introduced with IBM Spectrum™ Virtualize V7.5 on Storwize and SVC devices.

The HyperSwap function is a new level of security since it can also handle an outage of a Storwize V7000 control enclosure or a cluster in a single site. Before HyperSwap was introduced the Storwize V7000 was able to handle an outage of external (virtualized) or internal storage via the Volume mirroring feature. This did not cover an outage of one site or controller; this could only work at a host level in combination with a second storage system.

Starting with V7.5 a technology is available that provides a high availability (HA) solution transparent to a host over two locations which can have a distance of up to 300 km. This is the same as the Metro Mirror limitation for the distance.

The HA solution is based on HyperSwap volumes which have a copy at two different and independent sites. Data that is written to the volume is automatically sent to both sites, even if one site is no longer available then the other remaining site allows access to the volume.

A new Metro Mirror capability, *active-active Metro Mirror*, is used to maintain a fully independent copy of the data at each site. When data is written by hosts at either site, both copies are synchronously updated before the write operation is completed. The HyperSwap function automatically optimizes itself to minimize the data that is transmitted between sites and to minimize host read and write latency.

To define HyperSwap volumes, active-active relationships are made between the copies at each site. This is normally done using the GUI, but can also be created using the CLI. The relationships provide access to whichever copy is up to date through a single volume, which has a unique ID. As with normal remote mirror relationships, the HyperSwap relationships can be grouped into consistency groups. The consistency groups fail over consistently as a group based on the state of all of the copies in the group. An image that can be used in the case of a disaster recovery is maintained at each site.

As redundancy over the locations is needed, a Storwize V7000 HyperSwap configuration requires one system (one control enclosure with two nodes) in both locations. This results in a minimum configuration of two control enclosures, one in each site. Based on the used hardware additional devices can be used, as long as is supported by the used system (Storwize V7000 allow more control enclosures than Storwize V5000).

In addition to the two sites that are holding the data a third, independent site is required for any HyperSwap solution. This is mandatory to avoid so-called “split-brain” solutions, which can happen if both sites are not able to communicate anymore, but both sites are still up and running. This third site is the location of the quorum disk, which acts as a tiebreaker. For this, the third site needs an independent connection to both data sites.

Previously, this had to be a Fibre Channel-attached storage device. However, since V7.6 there is also the possibility to use an IP Quorum, which is a piece of software installed on a supported server (for example, Linux). Check the product documentation for all prerequisites that come along with this setup. By using this IP Quorum, this is selected automatically as the active quorum.

Several requirements must be validated for the Storwize V7000 HyperSwap implementations, specifically for the SAN extension. For more information about HyperSwap prerequisites, see the IBM Storwize V7000 Knowledge Center.

How a HyperSwap setup works is shown in Figure 1-1.

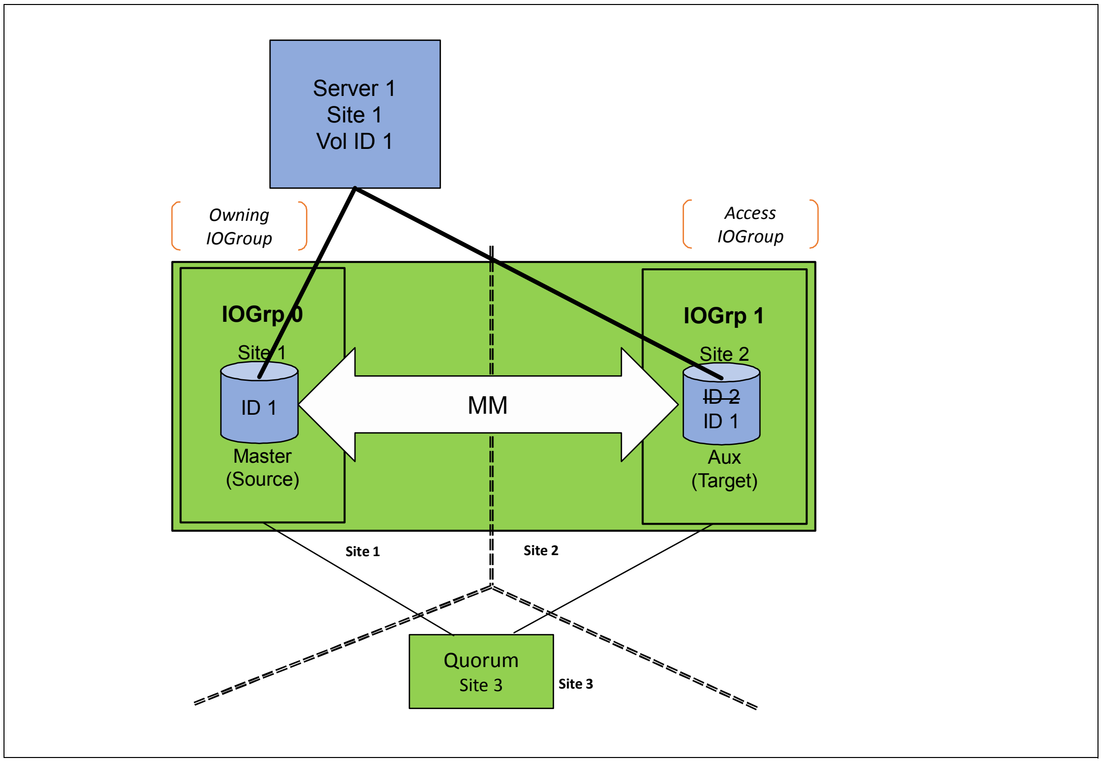


Figure 1-1   How HyperSwap works

Every volume has a copy that is mirrored via a special Metro Mirror (MM). With this mirror, the target volumes get the same ID as the source and is seen as the same volume. Because every volume is accessed at the same time via two I/O groups, the maximum number of paths is doubled here. So the server sees four nodes for every volume (two out of I/O group 0 and two from I/O group 1). The multipath driver selects the preferred path (via ALUA) and uses this path, as long as this path is online and reachable. For a case that the traffic is directed to the access I/O group, the data is forwarded to the owning I/O group.

1.2 IBM i overview

IBM i servers are the first-choice of systems for companies that want the benefits of business solutions without the complexity. IBM i product line offers the most integrated and flexible set of servers in the industry, designed for small and medium businesses, and scalability for large business solutions. IBM i servers run in partitions of IBM POWER® systems.

Next we present some features of IBM i servers that are important for working with external storage systems.

1.2.1 Single-level storage and object-orientated architecture

When you create a new file in a UNIX system, you must tell the system where to put the file and how big to make it. You must balance files across different disk units to provide good system performance. If you discover later that a file needs to be larger, you need to copy it to a location on disk that has enough space for the new, larger file. You may need to move files between disk units to maintain system performance.

IBM i server is different in that it takes responsibility for managing the information in auxiliary storage pools (also called disk pools or ASPs).

When you create a file, you estimate how many records it should have. You do not assign it to a storage location; instead, the system places the file in the best location that ensures the best performance. In fact, it normally spreads the data in the file across multiple disk units. When you add more records to the file, the system automatically assigns additional space on one or more disk units.

Therefore, it makes sense to use disk copy functions to operate on either the entire disk space or the IASP. Power HA supports only an asp-based copy.

IBM i uses a single-level storage, object-orientated architecture. It sees all disk space and the main memory as one storage area and uses the same set of virtual addresses to cover both main memory and disk space. Paging of the objects in this virtual address space is performed in 4 KB pages. However, data is usually blocked and transferred to storage devices in bigger than 4 KB blocks. Blocking of transferred data is based on many factors, for example, expert cache usage.

1.2.2 Translation from 520 byte blocks to 512 byte blocks

IBM i disks have a block size of 520 bytes. Most fixed block (FB) storage devices are formatted with a block size of 512 bytes so a translation or mapping is required to attach these to IBM i (IBM DS8000® supports IBM i with a native disk format of 520 bytes).

IBM i performs the following change of the data layout to support 512 byte blocks in external storage: for every page (8 * 520 byte sectors) it uses an additional 9th sector; it stores the 8-byte headers of the 520 byte sectors in the 9th sector, and therefore changes the previous 8* 520-byte blocks to 9* 512-byte blocks. The data that was previously stored in 8 * sectors is now spread across 9 * sectors, so the required disk capacity on V7000 is 9/8 of the IBM i usable capacity, and vice-versa, the usable capacity in IBM i is 8/9 of the allocated capacity in Storwize.

Therefore, when attaching a Storwize to IBM i, whether through vSCSI, NPIV or native attachment, this mapping of 520:512 byte blocks means that you will have a capacity constraint of being able to use only 8/9ths of the effective capacity.

The impact of this translation to IBM i disk performance is negligible.

2

Testing environment and configuration

This details the testing environment and the configuration that was employed.

2.1 Testing environment

We used the following environment for testing Storwize HyperSwap with IBM i:

- **POWER servers**
  - Production server - POWER P730
  - Disaster recovery (DR) server - POWER P770
- **IBM i level**

  We used IBM i level 7.3 Technology Refresh 3
- **Server Connection**

  For the tests we used NPIV connection of Storwize to IBM i
- **VIOS levels**

  We used two VIOS, both of them level 2.2.6.10
- **SAN switches**

  Four Brocade Fibre Channel switches are used. They are divided into a public and private switch (via virtual switch/fabric) to provide a private fabric for internal communication between the nodes and a public SAN for all other I/O traffic.
- **SAN connection and zoning**

  We used two fabrics (A and B) for public and private SAN.

  For the tests with PowerHA for i and IASP we used one fabric to make setup for these tests faster and simpler. However we certainly recommend to use two fabrics also for solutions with PowerHA for i.
- **Storwize V7000**
  - Storage server on Production site - Storwize V7000 Gen1
  - Storage server on DR site - Storwize V7000 Gen1

  The two Storwize V7000 storage servers are set up in a HyperSwap topology.

Figure 2-1 on page 7 shows the Server, Storage and SAN connection scheme.

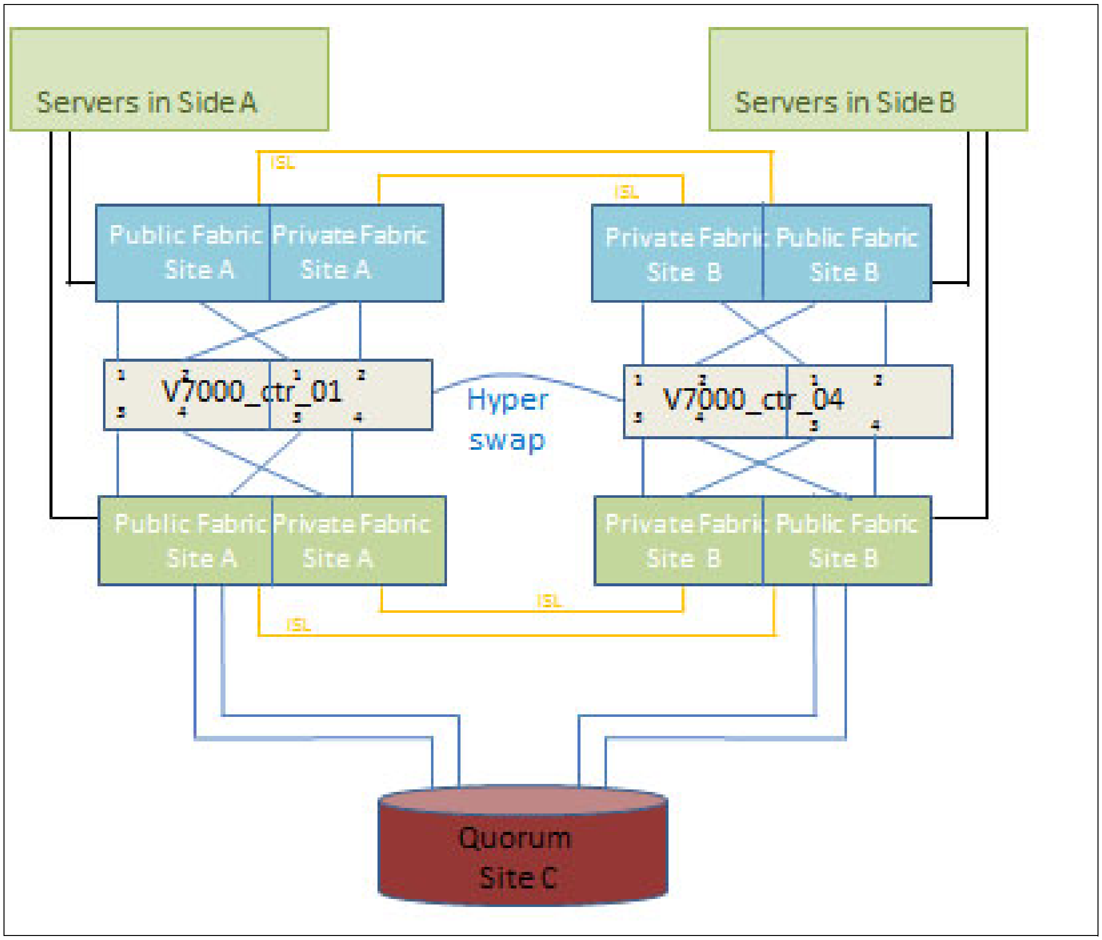


Figure 2-1 SAN connection scheme

2.2 Configuration of Storwize HyperSwap

The IBM Storwize V7000 HyperSwap setup consists of two controller enclosures with one expansion enclosure each. The controller enclosures are connected only to the switches in their location or site, there is no cross cabling to the other site. Due to the fact that only four ports are available per single controller each port is connected to a different switch:

- Port 1: local private switch Fabric A
- Port 2: local public switch Fabric A
- Port 3: local private switch Fabric B
- Port 4: local public switch Fabric B

This can be extended if more ports are available.

The Storwize V7000 control enclosures build a single cluster, so they have to be clustered in advance. This results in a two I/O group setup, one I/O group per site, which is the minimum setup. At the third site, an IBM DS8000 Storage system is used as the quorum, which has to be attached to both sites, independent of each other.

3

Implementation of Storwize HyperSwap

In this chapter, we detail the Storwize HyperSwap steps.

3.1 Storwize HyperSwap implementation

For this setup, we used two Storwize V7000 that were already installed so the initial setup was not done, and is beyond the intended scope of this paper. To be able to cluster all the nodes in a single cluster together, at least one cluster has to be deleted. After the delete, the nodes are brought into a candidate state, which enables them to join an existing cluster. By doing this, the nodes are visible as candidates to the existing cluster and they are added to the cluster, which ends in a four node or two I/O group setup, as shown in Figure 3-1.

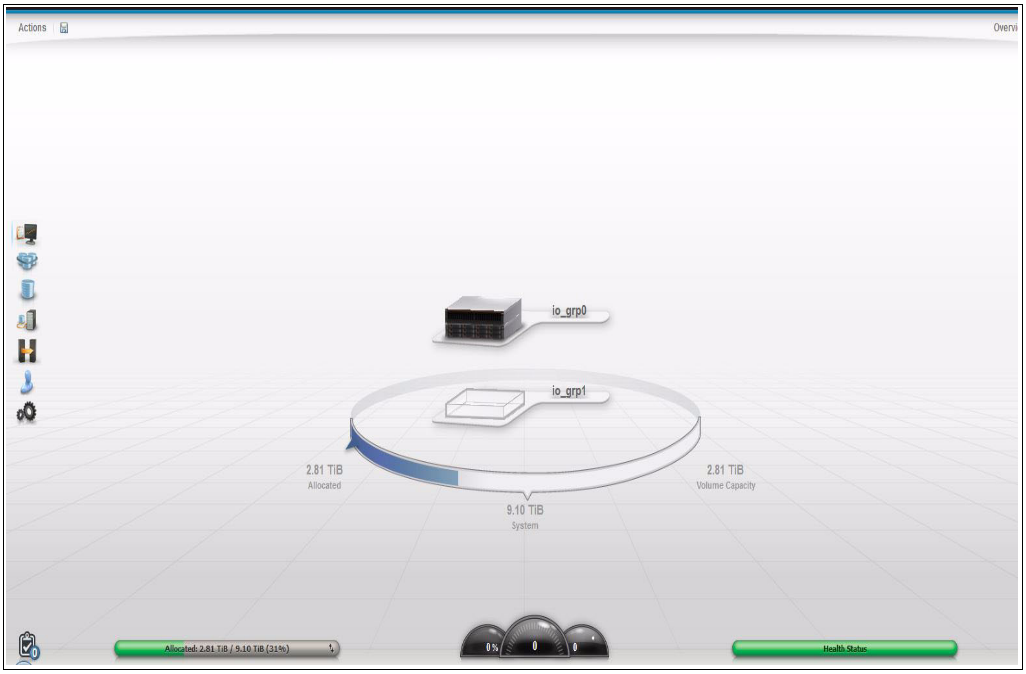


Figure 3-1 Adding nodes to a cluster

At this time, it is a normal Storwize V7000 cluster and there is no HyperSwap Cluster. To change it to a HyperSwap Cluster, make sure that you have fulfilled all the requirements, such as a dedicated SAN for internal communication and a quorum in a third site (FC storage or IP Quorum).

If all prerequisites are fulfilled the topology can be changed to HyperSwap, this can be done via the GUI as described here, or via the **chsystem** command. Make sure that by doing this via the CLI that the Site IDs are set in advance.

To change the topology via the GUI, complete the following steps:

1. Go to the **Action** menu in the start screen and select *Modify System Topology,* as shown in Figure 3-2.

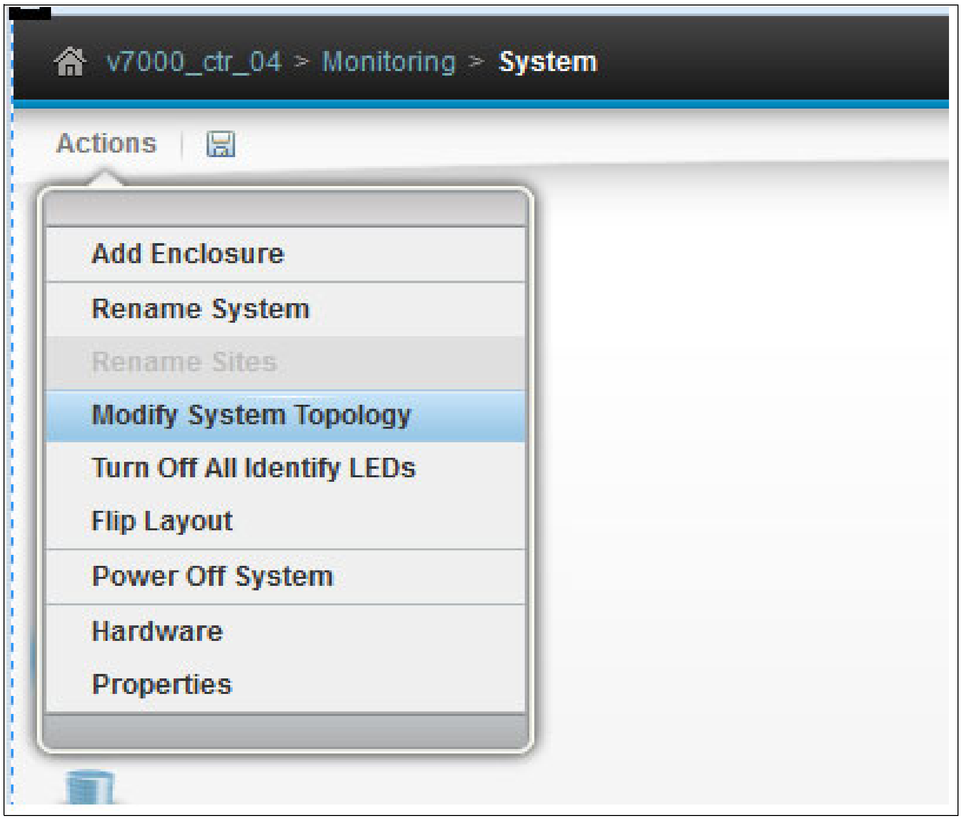


Figure 3-2 Modify System Topology

2. A wizard opens that guides you through the steps, as shown in Figure 3-3.

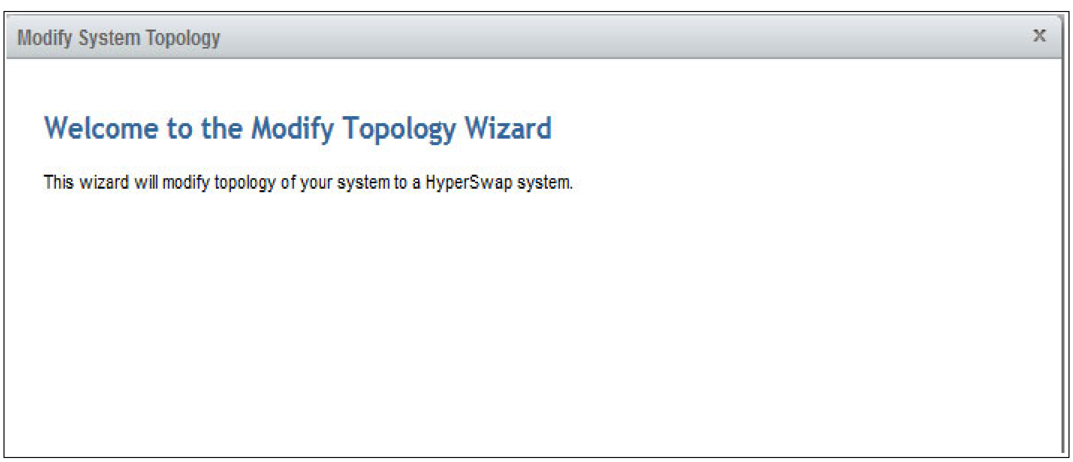


Figure 3-3 Modify topology wizard

3. The first screen of the wizard asks you to enter the site names. The default ones are used, but they can be changed to more descriptive names, such as *Berlin* or *Backup-DC*. You can leave them as they are here and change them later if required. This can be seen in Figure 3-4.

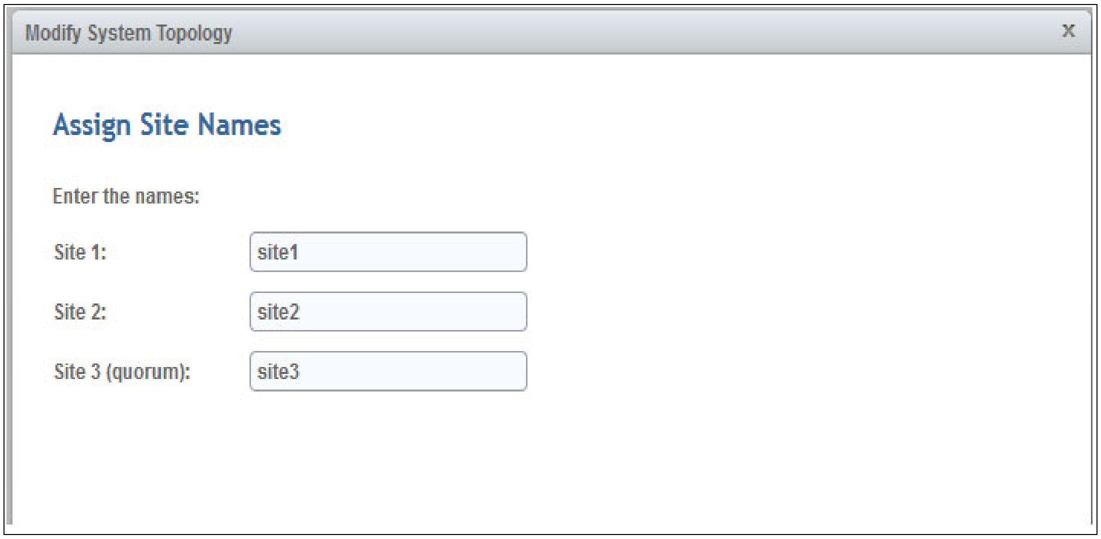


Figure 3-4 Assign site names

4. In the next wizard screen, the enclosures have to be assigned to a site. This is mandatory and has to be done correctly, because the code has to rely on this when the preferred nodes and paths are negotiated, so be careful and double check your actions. The location can be changed by clicking on the small blue button with the arrows. This is shown in Figure 3-5.

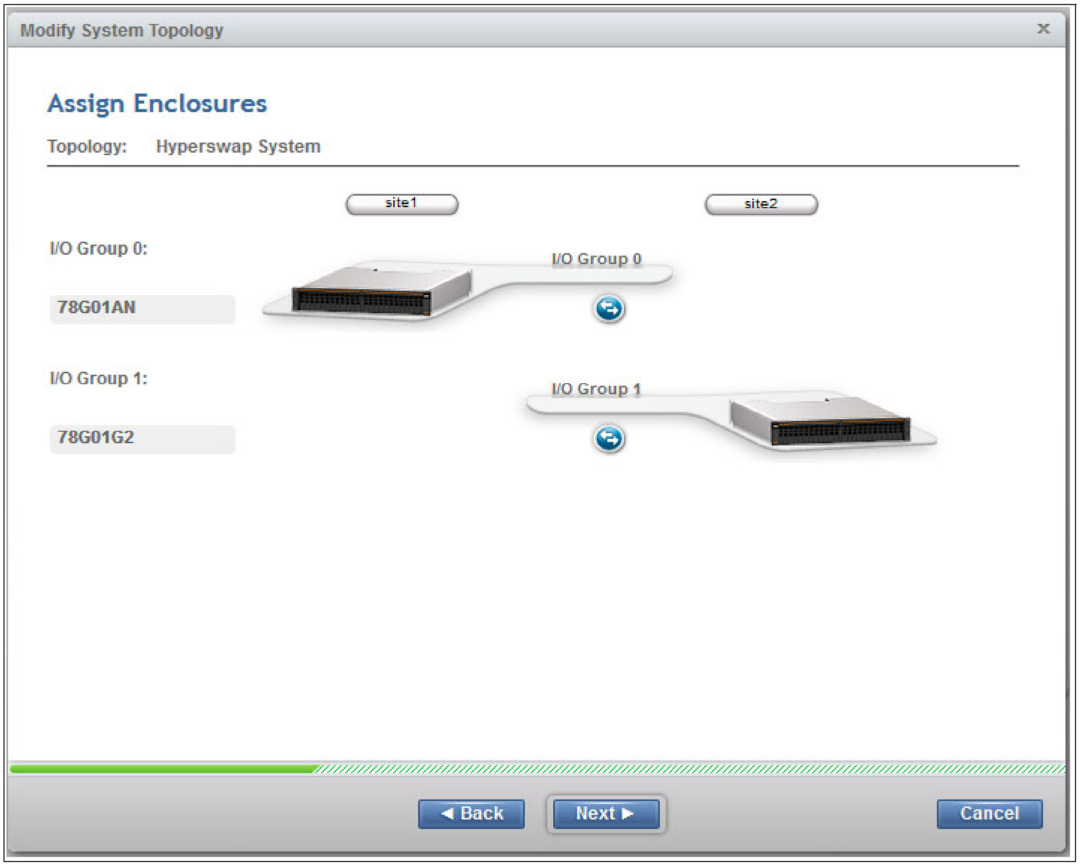


Figure 3-5 Assign enclosures

5. After the nodes get a site assigned, the host needs a site ID. Even if it is not mandatory as it is for the nodes, it is highly recommended to give a site ID (which reflects the location of the host) to every host. This enables the system to select the correct preferred path with the best latencies. This can be done for each host separately. See Figure 3-6.

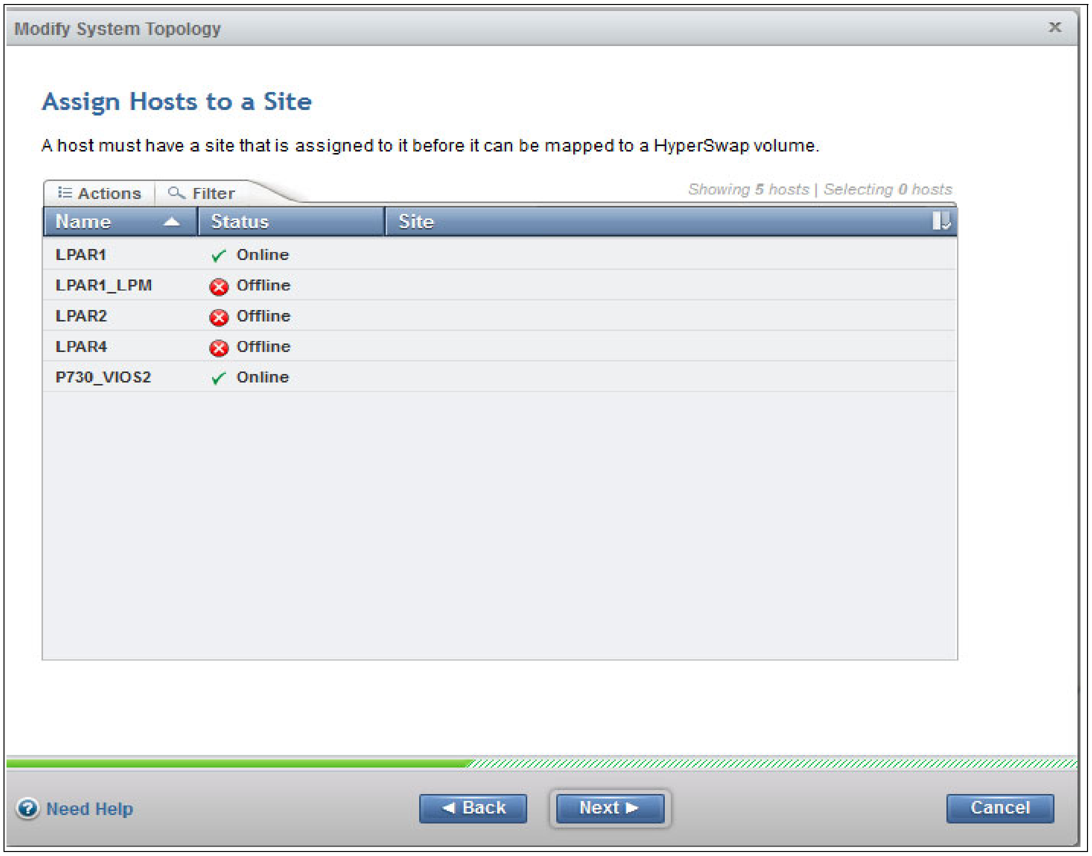


Figure 3-6 Assign hosts to a site

6. After the hosts are assigned, now the external storage system has to be assigned to a site. At a minimum one system has to get a site, the quorum site (an exception is the usage of an IP Quorum). This is the third site (site 3) and is set by default as the active quorum. If there are other storage systems attached and virtualized, they should also get a valid site ID that reflects their location, because this gives greater efficiency. Assigning the quorum site is shown in Figure 3-7 and Figure 3-8 on page 16.

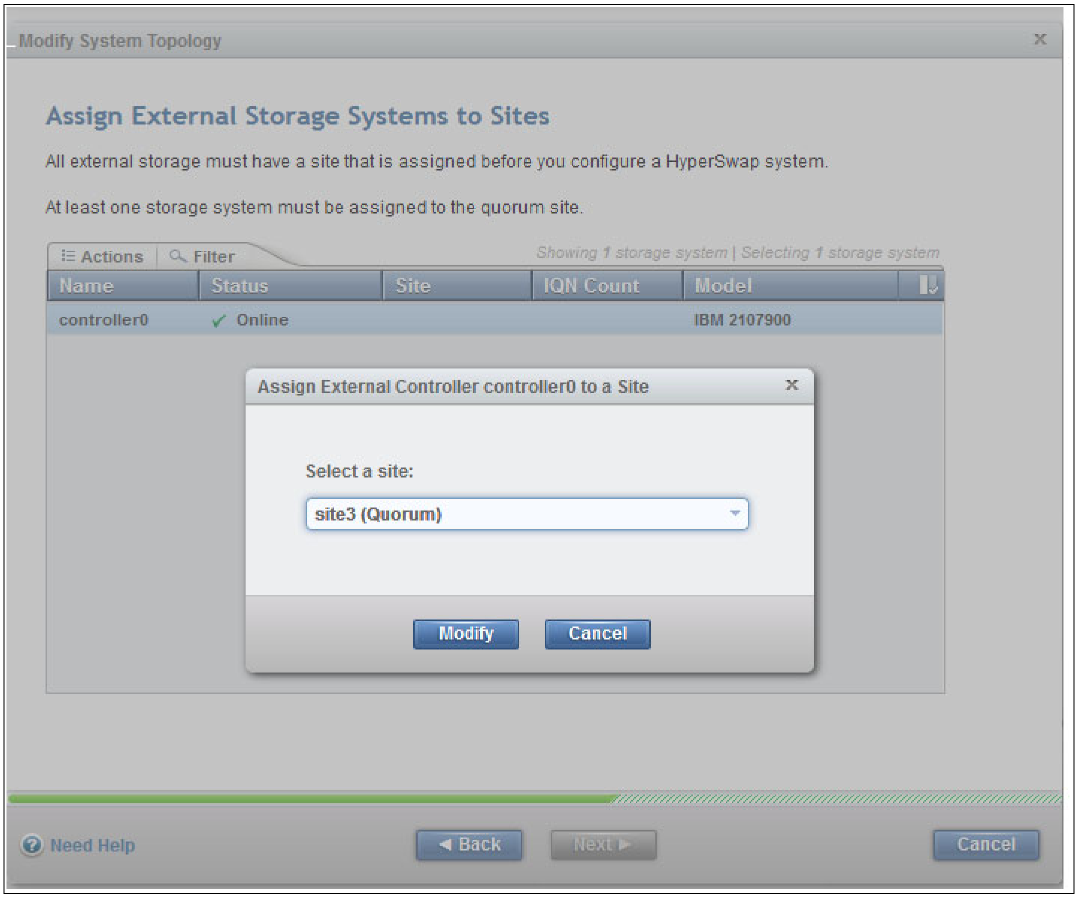


Figure 3-7 Assigning storage to Quorum site - 1

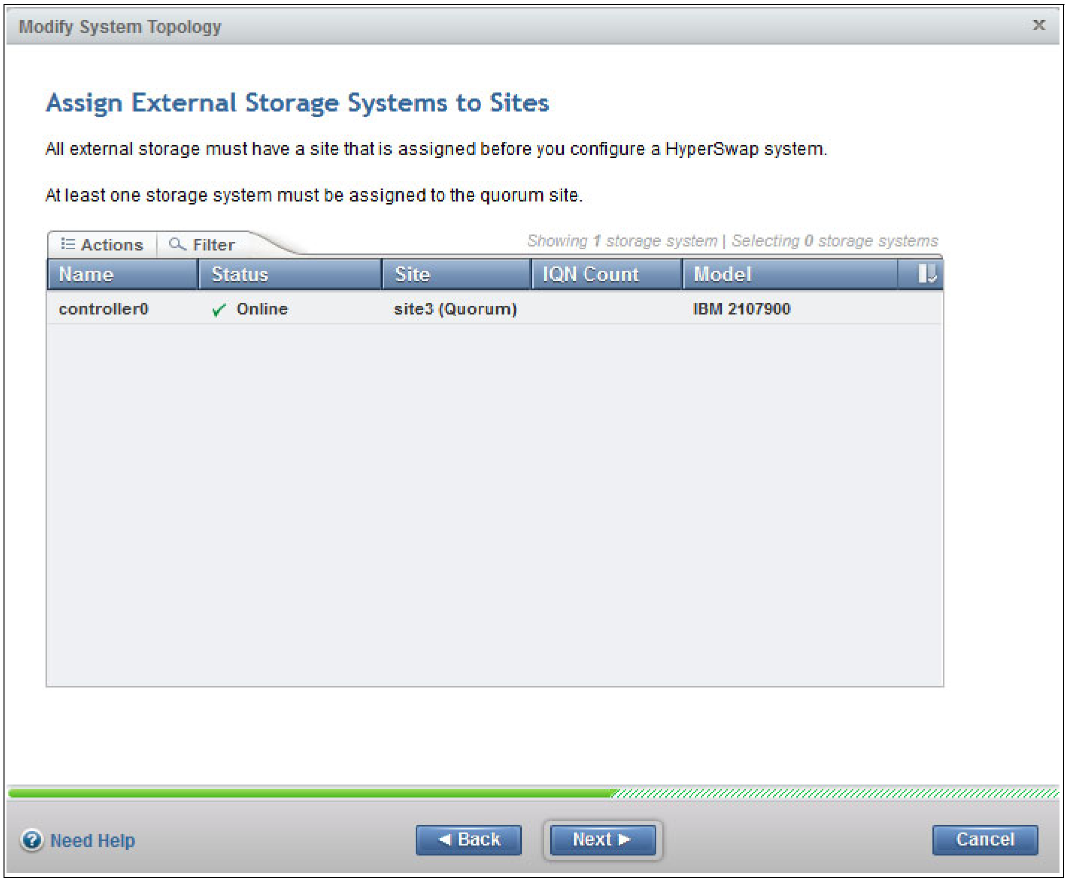


Figure 3-8 Assigning storage to Quorum site - 3

7. In the next screen you have to set the bandwidth which is available between both sites and can be used by the system. This is to limit the system and to not allow it to fill up a shared bandwidth. The bandwidth has to be sized carefully and checked from time to time if there are major changes. Also, the percentage of the bandwidth which can be used for background copies can be set. At the end you see the resulting Total background Copy calculated.

Setting the bandwidth is shown in Figure 3-9 (the numbers in Figure 3-9 are example numbers).

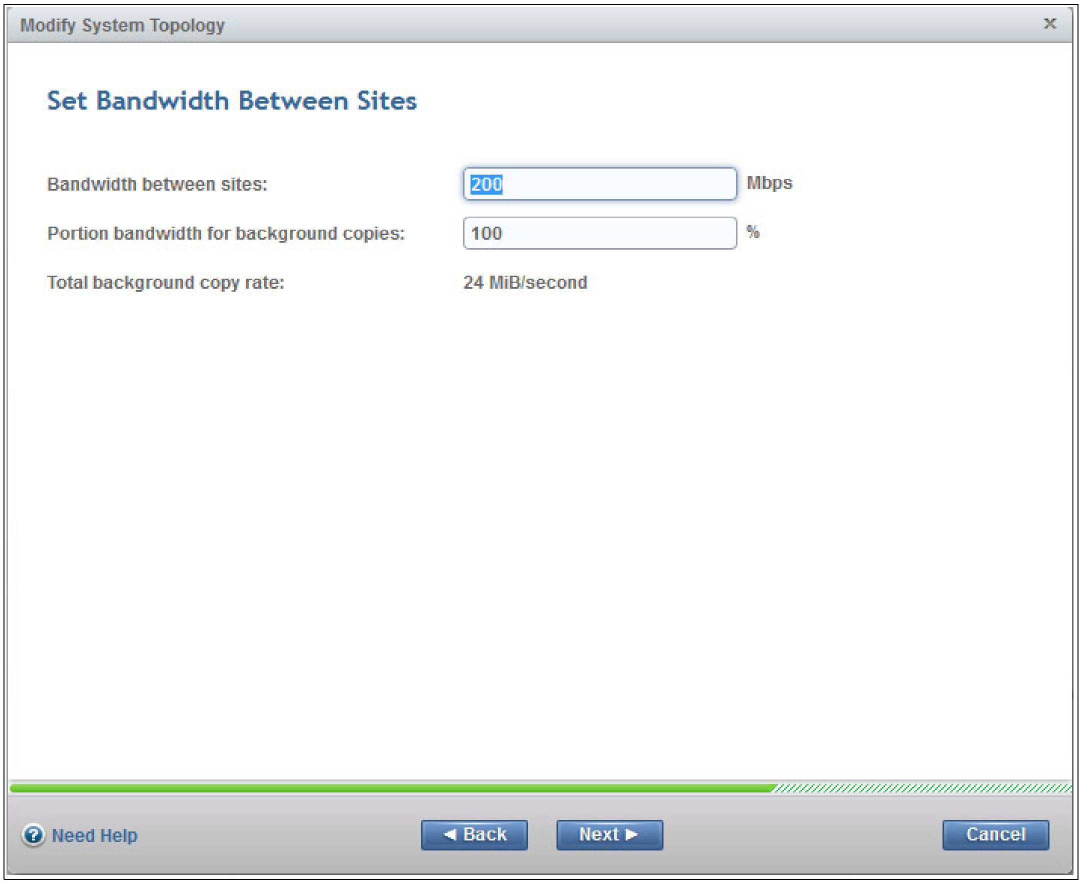


Figure 3-9 Setting the bandwidth between sites

8. Before the wizard is finished you will see a short summary of your setting and everything should be checked. If everything is as expected, click **Finish** to change the topology to HyperSwap. This is shown in Figure 3-10.

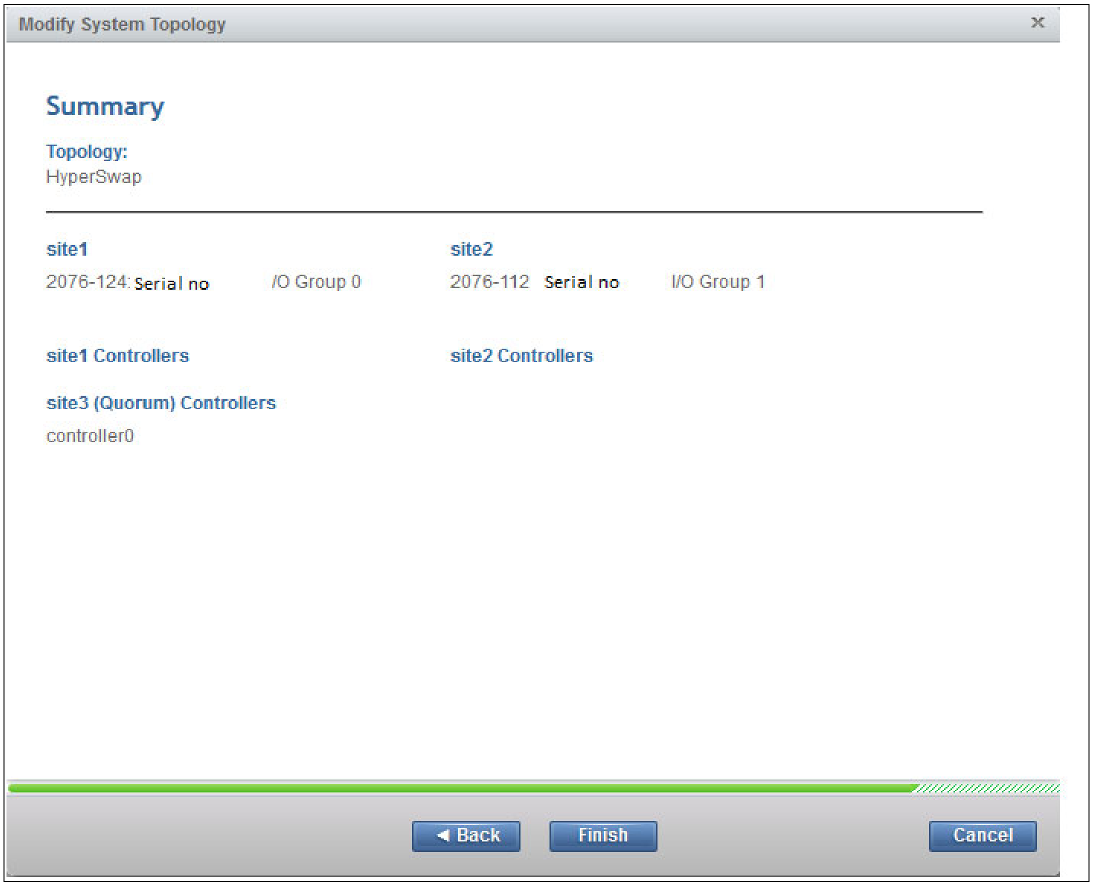


Figure 3-10 Summary

9. In the System view, you see the result (this is also the entry screen), with the sites and the hardware located in the sites. This is a sign that HyperSwap is already active, as is shown in Figure 3-11.

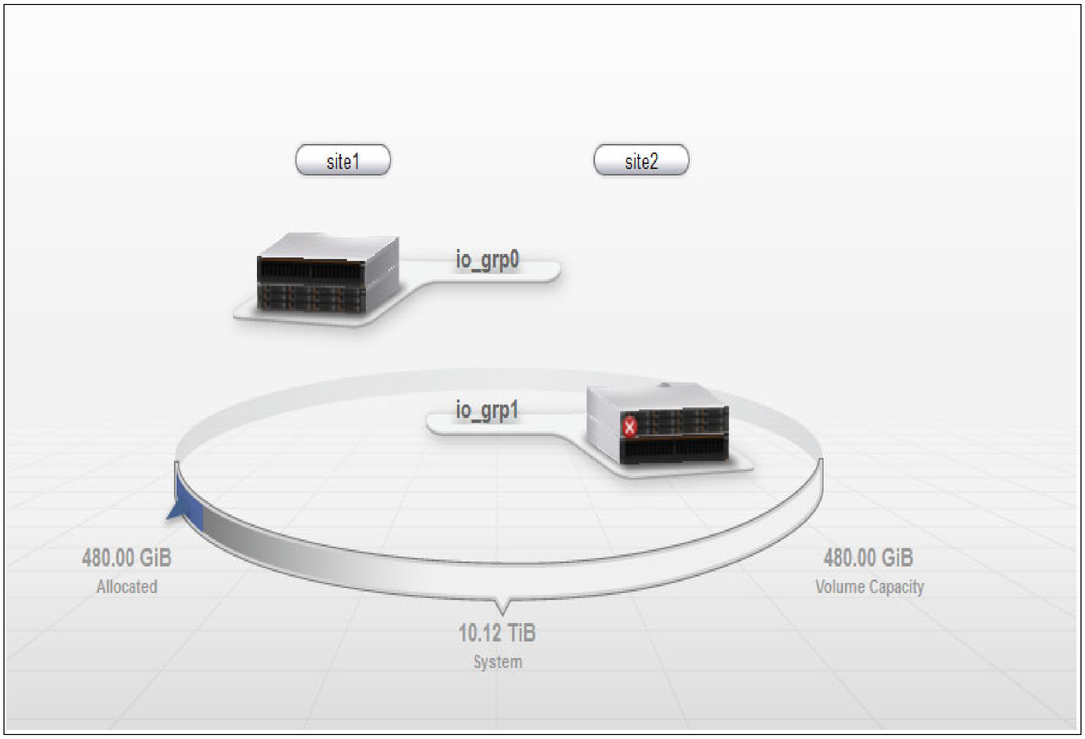


Figure 3-11 Active HyperSwap

10. In Create Volumes, the HyperSwap volume is offered, and here you can create a new HyperSwap where all the settings (creation of IBM FlashCopy® mappings, creation of relationship, and so on) are done using the GUI.

As with any other creation you can select the size, number of volumes, and the pool where the volumes should reside. The site ID is realized by the GUI, so pools from both sites are involved. Creating a HyperSwap volume is shown in Figure 3-12.

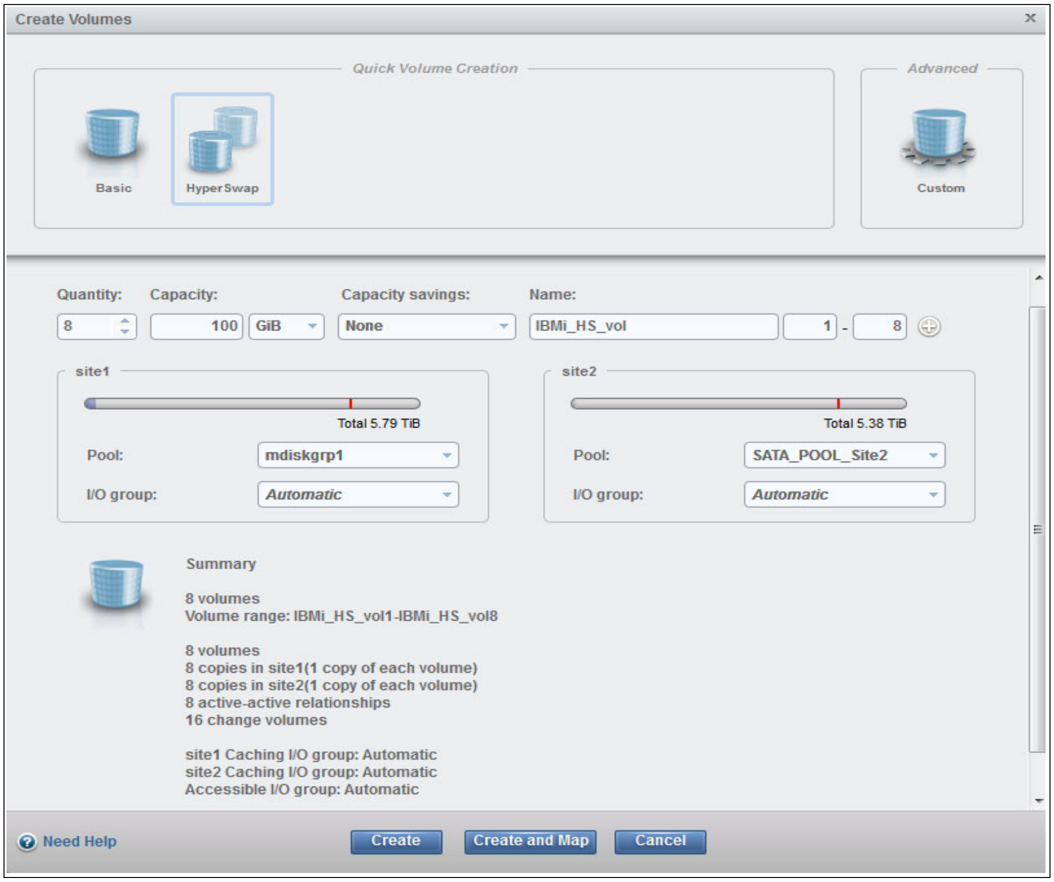


Figure 3-12 Creating a HyperSwap volume

11. The results are the HyperSwap volumes with all the volumes for Metro Mirror, FlashCopy, and so on, as can be seen in Figure 3-13.

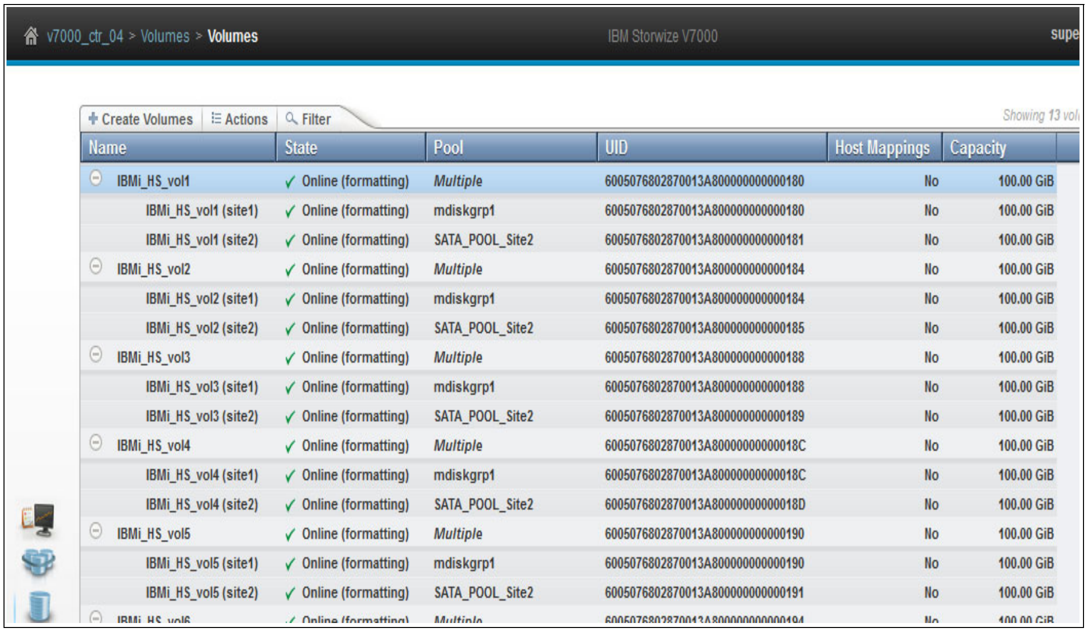


Figure 3-13 HyperSwap volumes

3.2 Configuration of IBM i

On each site, we use a POWER server with an IBM i logical partition connected with two Virtual I/O servers (VIOS) in N-Port ID Virtualization (NPIV) mode.

Two SAN fabrics are implemented, one switch in each fabric on each site. The switches of the same fabric are connected by an Inter Switch Link (ISL). Each switch is divided into one part for the private fabric and one part for the public fabric.

Each VIOS is connected to the SAN with two FC ports, one port to each public fabric.

We use a Storwize V7000 at each site. Each node of the Storwize is connected to the SAN with four ports: a port to Private Fabric A, a port to Public Fabric A, a port to Private Fabric B, a port to Public Fabric B.

The quorum disk resides on the DS8000 which is connected to Public fabric B at either site.

The connection scheme is shown in Figure 3-14.

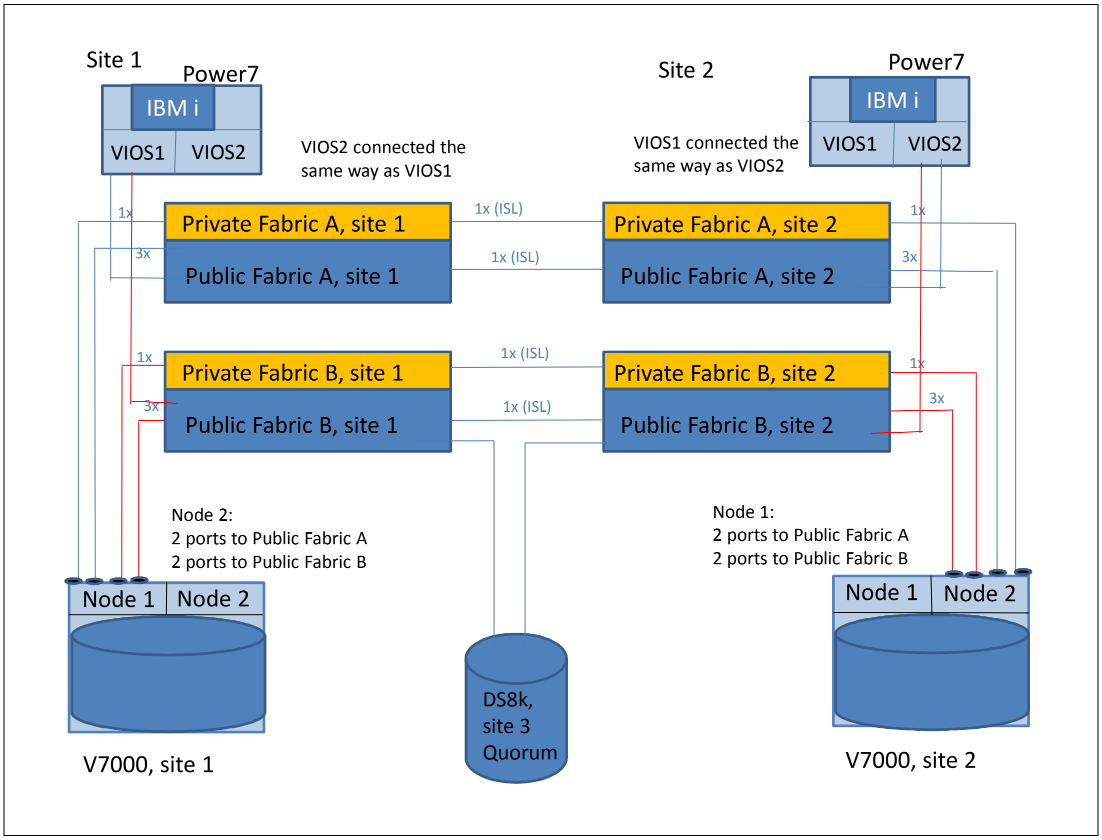


Figure 3-14 Connection scheme

4

IBM i implementation

In this chapter we describe IBM i implementation.

4.1 Creating the LUNs for IBM i

We create 8 HyperSwap LUNs to attach to IBM i. Created LUNs are shown on Figure 4-1.

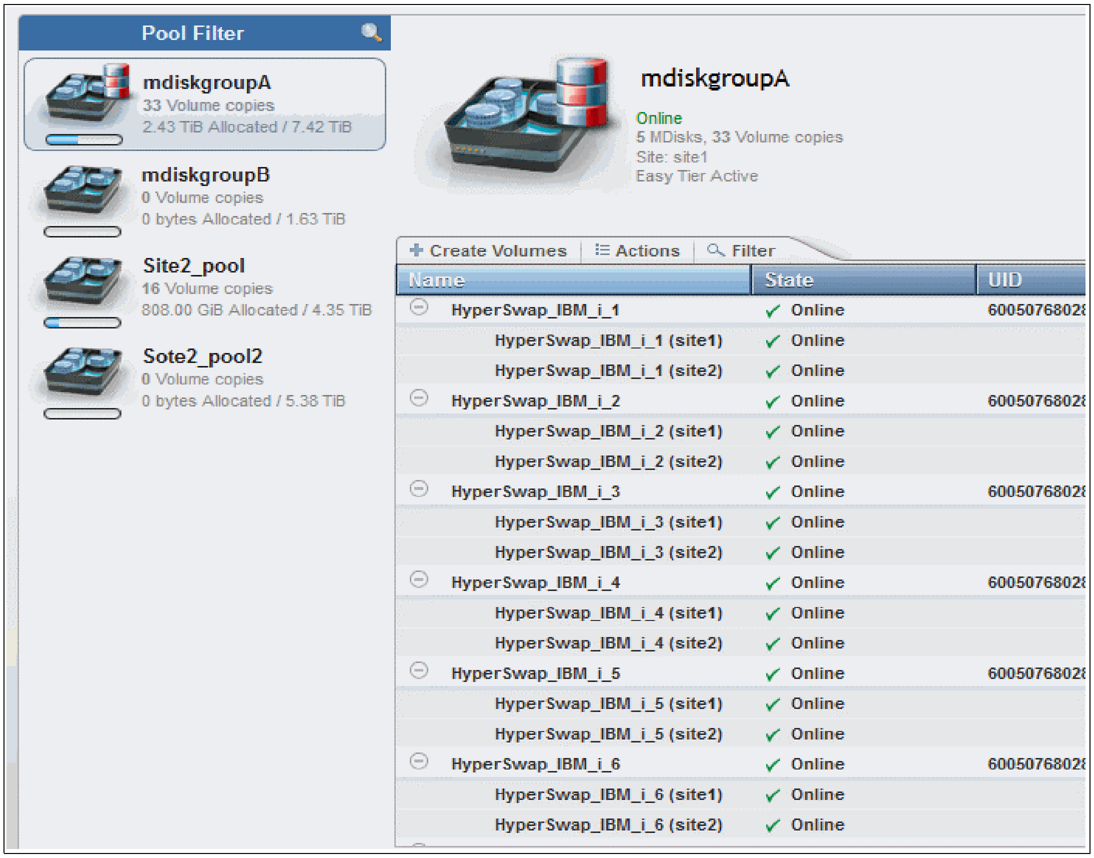


Figure 4-1 HyperSwap LUNs for IBM i

HyperSwap LUNs are in the Metro Mirror Consistency Group, which is shown in Figure 4-2.

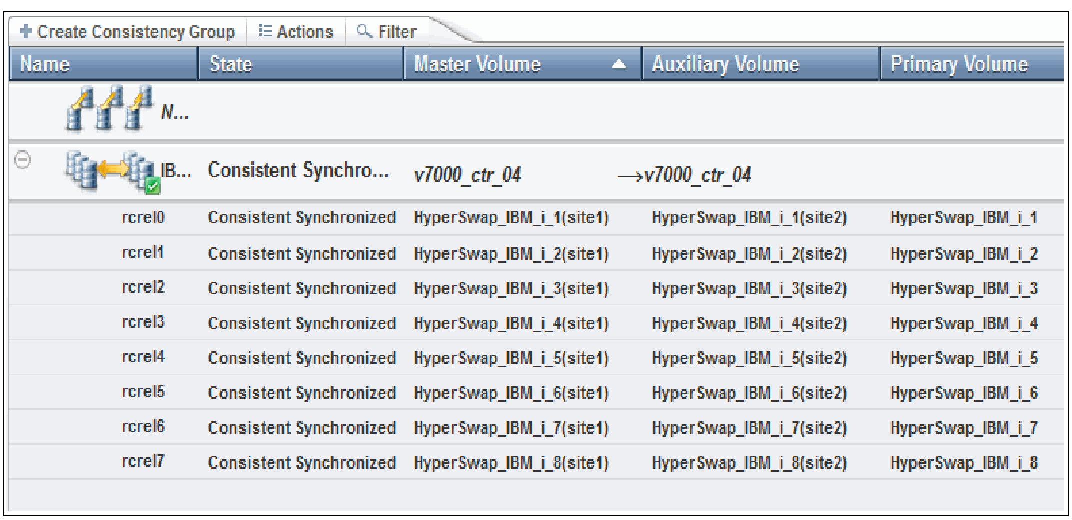


Figure 4-2 Metro Mirror consistency group of HyperSwap volumes

The automatically assigned preferred nodes at each site of the IBM i LUNs are shown in Table 4-1.

Table 4-1 Preferred nodes of IBM i LUNs

LUN	Site 1	Site 2
	Preferred node	Preferred node
HyperSwap_IBM_i_1	1	4
HyperSwap_IBM_i_2	2	3
HyperSwap_IBM_i_3	1	4
HyperSwap_IBM_i_4	2	3
HyperSwap_IBM_i_5	1	4
HyperSwap_IBM_i_6	2	3
HyperSwap_IBM_i_7	1	4
HyperSwap_IBM_i_8	2	3

The preferred nodes of a HyperSwap LUN can be seen in LUN properties, as shown in Figure 4-3.

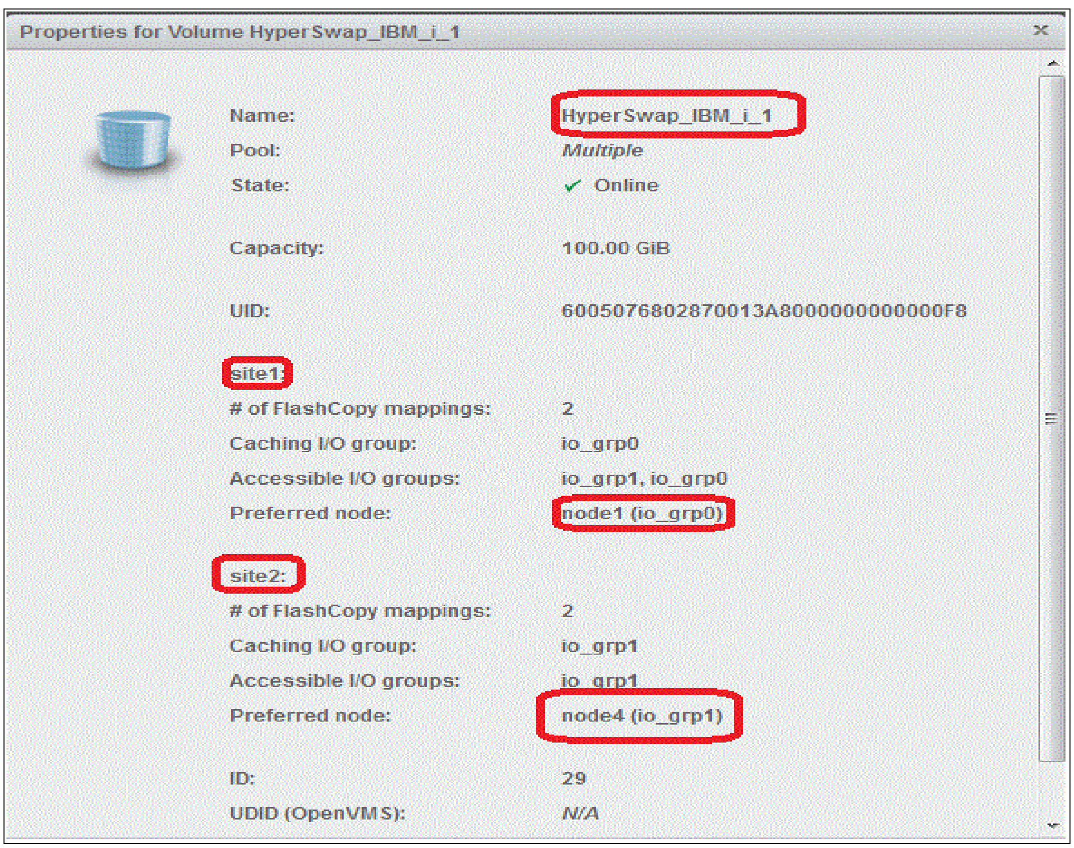


Figure 4-3 Preferred nodes in LUN properties

4.2 Connecting the IBM i LUNs

The Storwize V7000 host for IBM i has four virtual FC adapter with numbers 40, 41, 42, and 43. Two of them are assigned to VIOS1, and two are assigned to VIOS2. Each VIOS is connected via Fabric-A and Fabric-B. Virtual adapter 40 uses VIOS1 Fabric-A, VFC 41 uses VIOS1 Fabric-B, VFC 42 uses VIOS2 Fabric-A, and VFC 43 uses VIOS2 Fabric-B.

Figure 4-4 and Figure 4-5 show zoning of the switches.

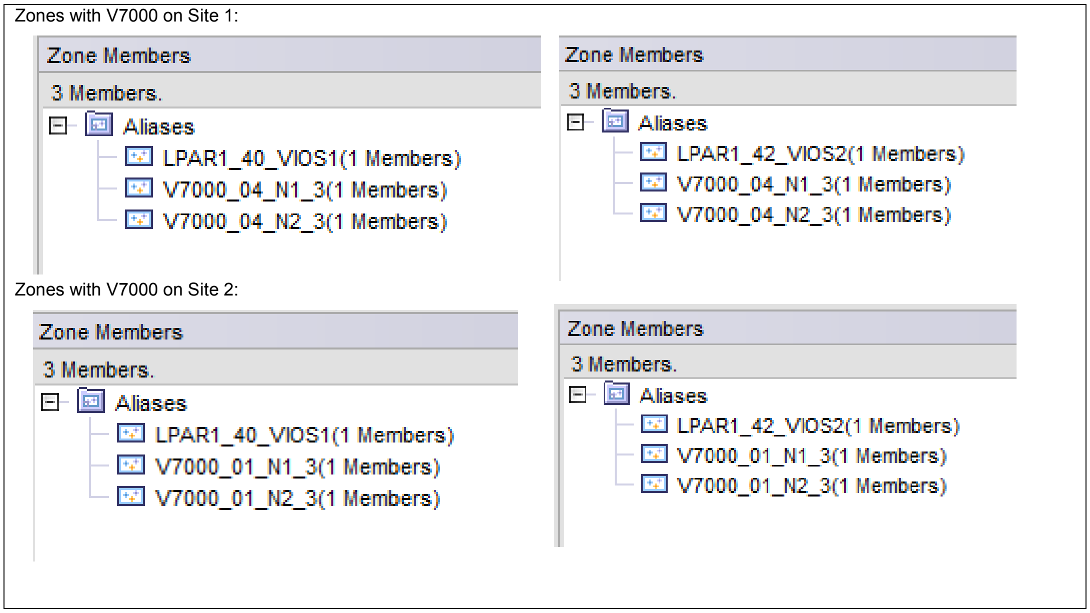


Figure 4-4 Zoning of Fabric A

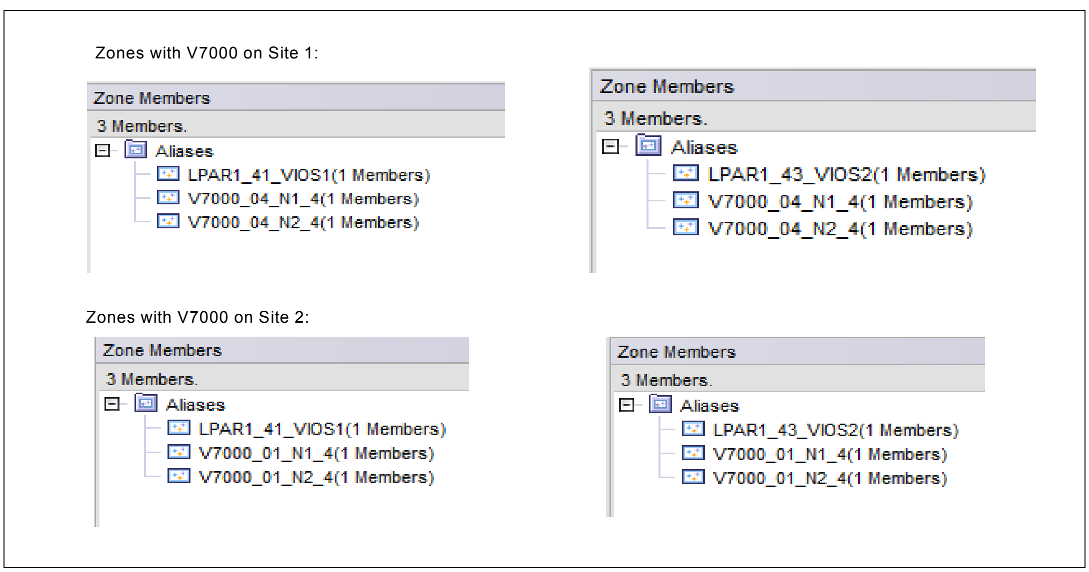


Figure 4-5 Zoning of Fabric B

The attachment and zoning scheme is shown in Figure 4-6.

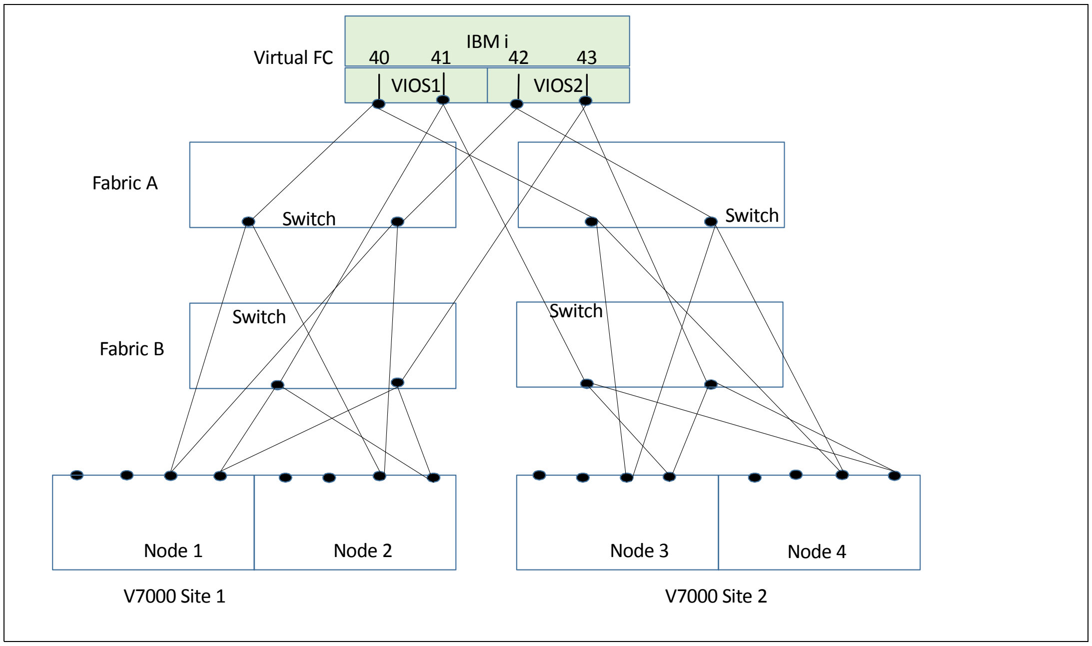


Figure 4-6   Attachment scheme

In the Storwize V7000 cluster we create a Host containing the four WWPNs of IBM i virtual FC adapters, as is shown in Figure 4-7.

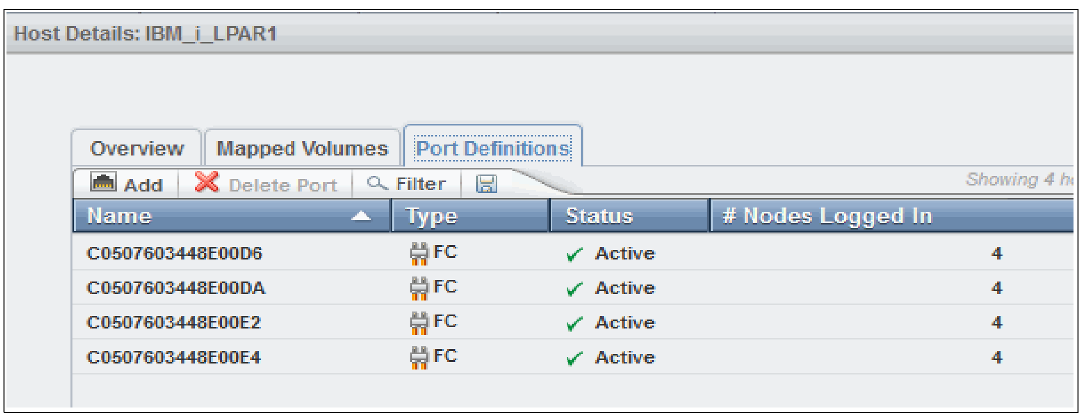


Figure 4-7   IBM i Host on Site 1

We mapped the HyperSwap LUNs to the IBM i Host, as is shown in Figure 4-8.

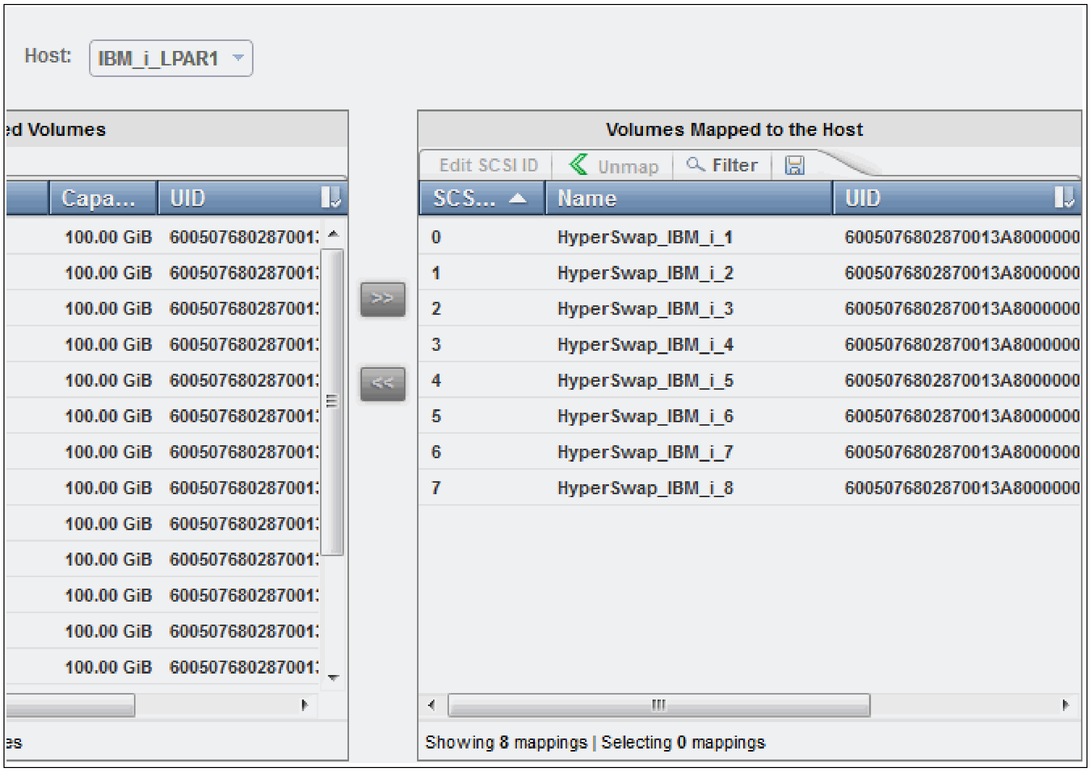


Figure 4-8 Mapping IBM i LUNs to the Host

4.3 Prepared partition on Site 2

Prepared partition IBM_i_Site2 in another Power server is connected to the Storwize cluster with two VIOS in NPIV. Four virtual FC are used to connect the Storwize cluster. SAN cabling and zoning is the same as for the production partition, see Figure 4-6 on page 27.

Host in the Storwize cluster has the WWPNs of the four virtual FC adapters of IBM_i_Site2, as shown in Figure 4-9. Normally, no LUNs are mapped to the host.

Figure 4-9 IBM i Host on Site 2

4.4 Installing IBM i on HyperSwap LUNs

We install IBM i from the installation images in the VIOS repository. Figure 4-10 shows tagged virtual adapters in HMC. Figure 4-11 shows settings of the IBM i LPAR on Site 1. On Figure 4-12 on page 30 see the activating of the partition to install IBM i.

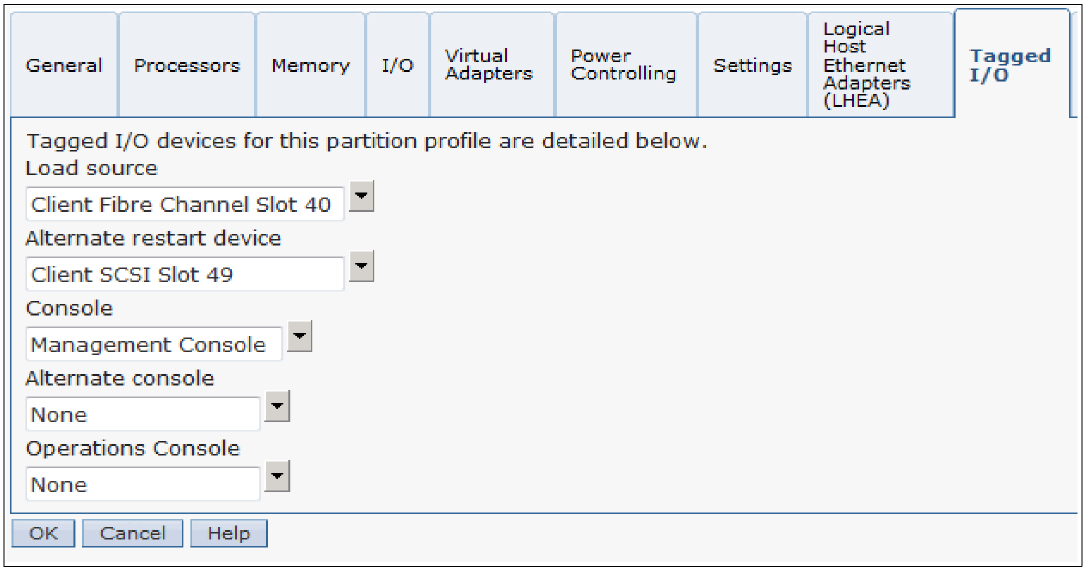


Figure 4-10 Tagged virtual adapters on Site 1

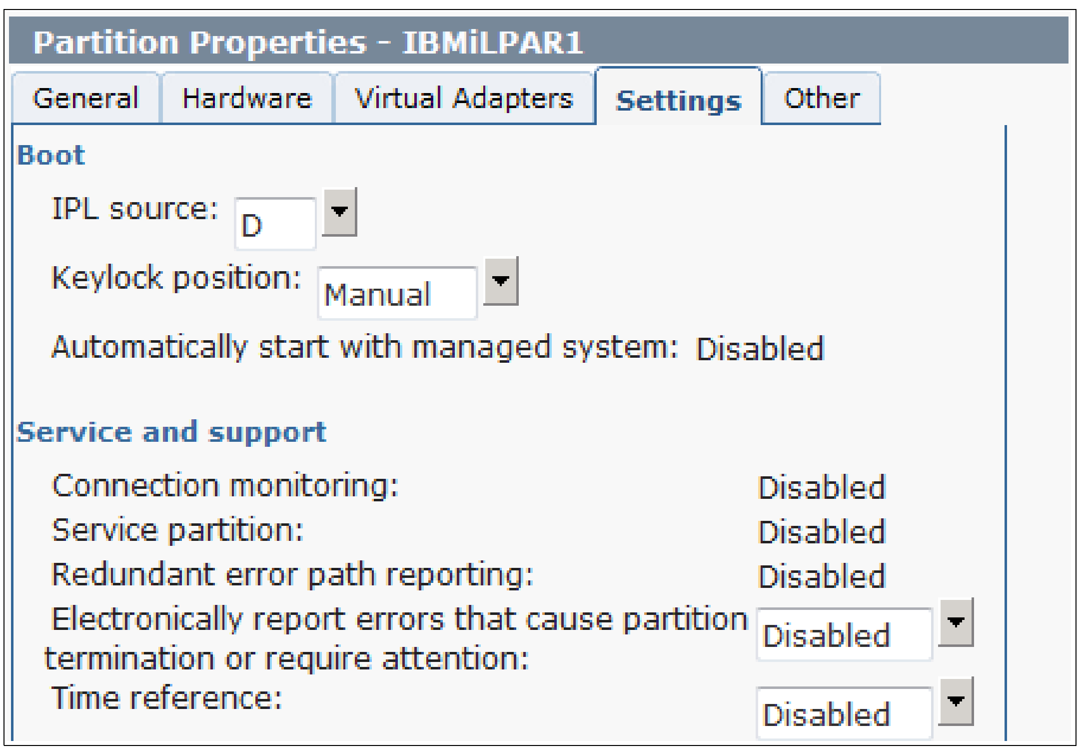


Figure 4-11 Installation settings

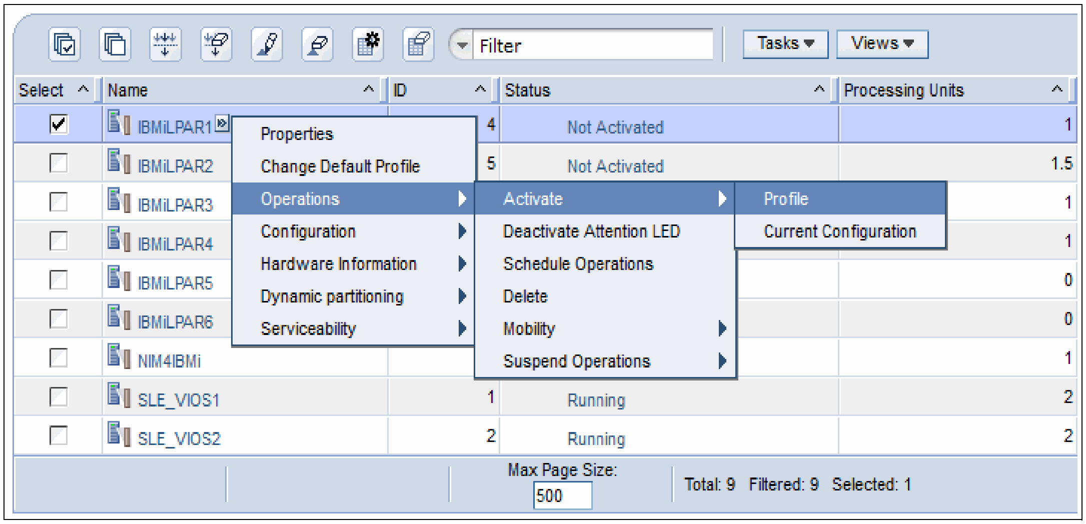


Figure 4-12 Activation of IBM i LPAR

After installation starts, we are prompted to select a LoadSource device, as can be seen in Figure 4-13.

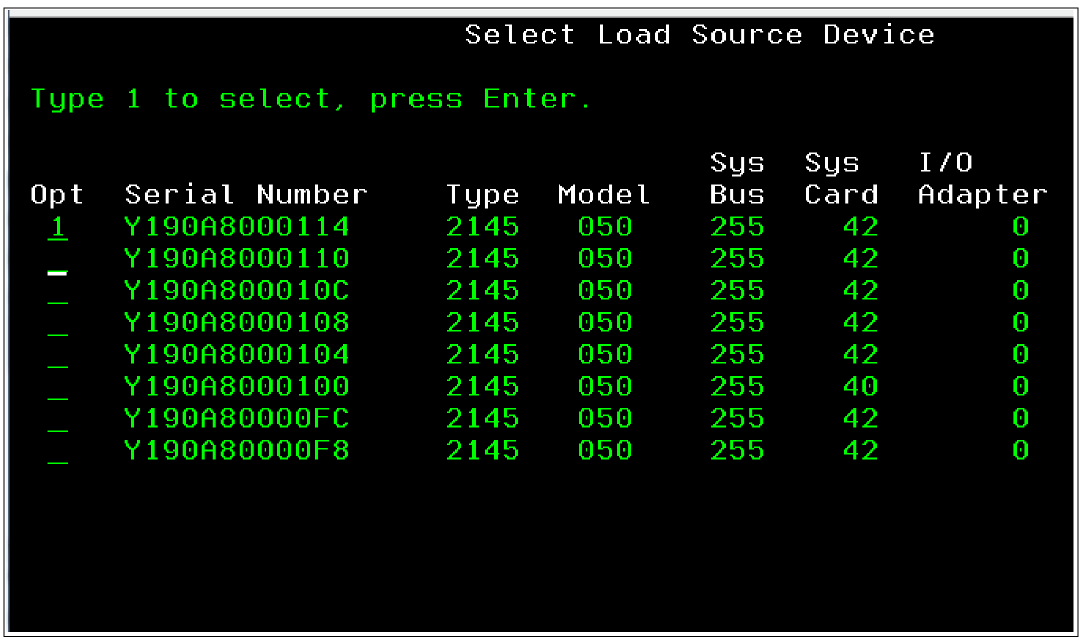


Figure 4-13 Select LoadSource

After IBM i is installed we look for the paths to LUNs, by using command **STRSST** (Start Service Tools) then select 3 Work with disk units, select 1 Display disk configuration, select 9 9 Display Disk Path Status.

Each resource DMPxxx represents one path to the LUN. Each LUN has four active and 12 passive paths, as can be seen in Figure 4-14. Active paths are the paths to the preferred node on Site 1, each path from one VIOS and one fabric. Passive paths are the paths to the non-preferred node on Site 1 and to both nodes at Site 2.

Display Disk Path Status

ASP	Unit	Serial Number	Type	Model	Resource Name	Path Status
1	1	Y190A8000114	2145	050	DMP003	Active
		Y190A8000114	2145	050	DMP062	Passive
		Y190A8000114	2145	050	DMP063	Passive
		Y190A8000114	2145	050	DMP002	Passive
		Y190A8000114	2145	050	DMP061	Passive
		Y190A8000114	2145	050	DMP064	Passive
		Y190A8000114	2145	050	DMP004	Passive
		Y190A8000114	2145	050	DMP001	Active
		Y190A8000114	2145	050	DMP171	Passive
		Y190A8000114	2145	050	DMP172	Passive
		Y190A8000114	2145	050	DMP173	Passive
		Y190A8000114	2145	050	DMP174	Passive
		Y190A8000114	2145	050	DMP175	Active
		Y190A8000114	2145	050	DMP176	Active
		Y190A8000114	2145	050	DMP191	Passive
		Y190A8000114	2145	050	DMP192	Passive

Figure 4-14 Paths to an IBM i LUN

4.5 Using Independent Auxiliary Storage Pool (IASP)

When using solutions with IASP, make sure that both Sysbas and IASP of the production LPAR reside on HyperSwap LUNs, and that Sysbas of the disaster recovery (DR) partition resides on the HyperSwap LUNs. The reason for this is that having IASP alone on the HyperSwap LUNs doesn't bring any benefit of HyperSwap compared to Metro Mirror.

4.6 Migrating IBM i to HyperSwap LUNs

We suggest the following options to migrate an IBM i from existing LUNs or disk units to HyperSwap LUNs:

- Save / Restore
- ASP balancing and copy LoadSource
- Using FlashCopy

4.6.1 Save / Restore

With this migration you save the IBM i system to tape and then restore it from tape to an LPAR with disk capacity on HyperSwap LUNs. Migration is straightforward and doesn't require any additional resources. However, it requires relatively long downtime.

4.6.2 ASP balancing and copy LoadSource

This migration method requires relatively short downtime, but it might require you to temporarily connect additional FC adapters to IBM i. Use the following steps to perform it:

1. Connect HyperSwap LUNs and existing internal disks or LUNs to IBM i LPAR.
2. By using the ASP balancing function, migrate data from the currently used disks or LUNs except LoadSource, to the HyperSwap LUNs. ASP balancing doesn't require any downtime. It is done while the IBM i partition is running. Depending on the installation needs, you may perform load balancing relatively quickly with some impact on performance, or slowly with minimal performance influence.
3. After data (except LoadSource) is migrated to HyperSwap LUNs, copy LoadSource to a HyperSwap LUN, which must be at least as big as the present LoadSource. This action is disruptive, and it requires careful planning.
4. After LoadSource is copied to a HyperSwap LUN, IBM i starts working with the entire disk capacity on HyperSwap LUNs.
5. Disconnect the previous storage system from IBM i, or remove the internal disk.

4.6.3 Using FlashCopy

When the existing IBM i LUN resides on the same Storwize V7000 cluster as the HyperSwap LUNs that you plan to use, you might be able to use FlashCopy of the existing LUNs to the HyperSwap LUNs. You should perform the following steps to migrate:

1. Power Down IBM i.
2. FlashCopy existing IBM i LUNs to the HyperSwap LUNs. Use FlashCopy with background copying.
3. Start IBM i from the HyperSwap LUNs and continue workload.

> **Important:** Note that we didn't test this migration method. Therefore, we strongly recommend that you perform a Proof of Concept (PoC) before using it.

5

Business continuity scenarios

In this chapter we describe some business continuity scenarios.

5.1 Business continuity scenarios for IBM i full system with HyperSwap

In this chapter, we describe business continuity solutions with Storwize HyperSwap and full system IBM i.

5.1.1 Outage of Storwize IO group on Site 1

To trigger the outage of the I/O group on Site 1, we enter a Service state for both Storwize Node 1 and Node 2.

At the outage, the I/O rate automatically transfers to Node 3 and Node 4 at Site 2. The IBM i workload keeps running, and there are no relevant messages in IBM i message queues. IBM i LUNs now have the path status as follows:

- 8 failed paths: The paths from Node 1 and Node 2.
- 4 active paths and 4 passive paths: The paths from Node 3 and Node 4. Active paths are the paths through the preferred node of a LUN, and the passive paths are through the non-preferred node of the LUN.

To end the outage of the I/O group at Site 1, we exit the service state of both nodes at Site 1. The IBM i I/O rate automatically transfers to Node 1 and Node 2, the IBM i workload keeps running, and there are no relevant messages in IBM i message queues. After failback, the status of IBM i paths is the same as initially: 4 active and 16 passive paths.

I/O rate on nodes during the outage is shown on the IBM Spectrum Control™ graph in Figure 5-1.

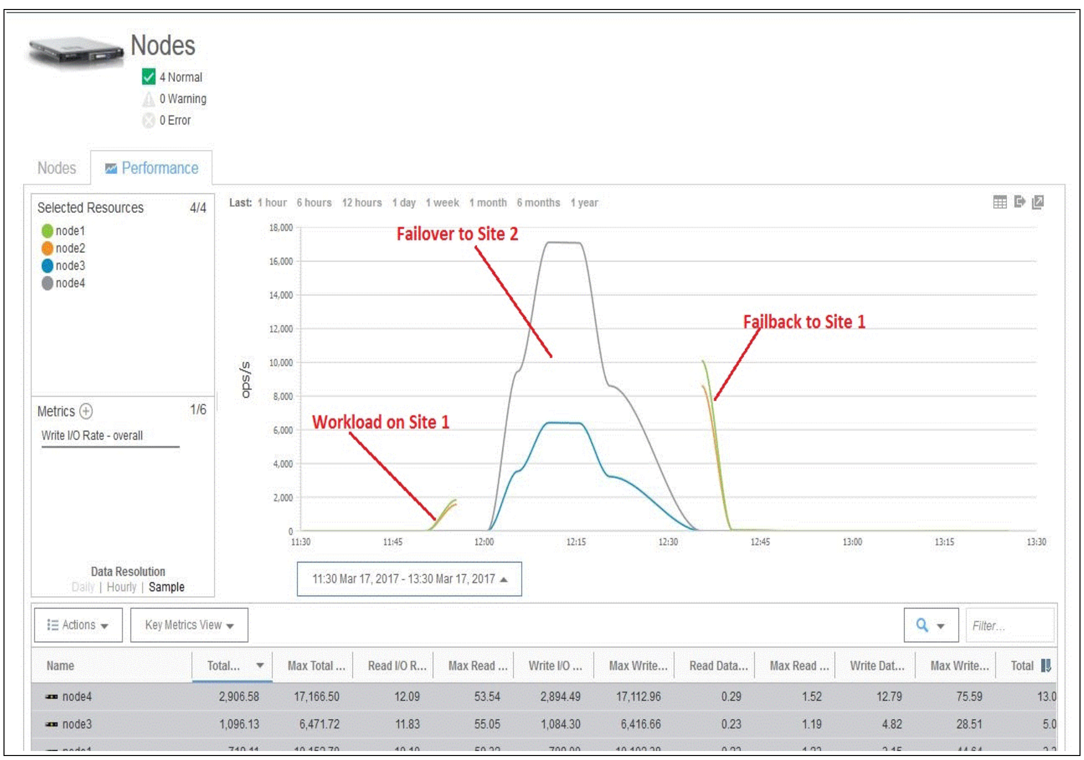
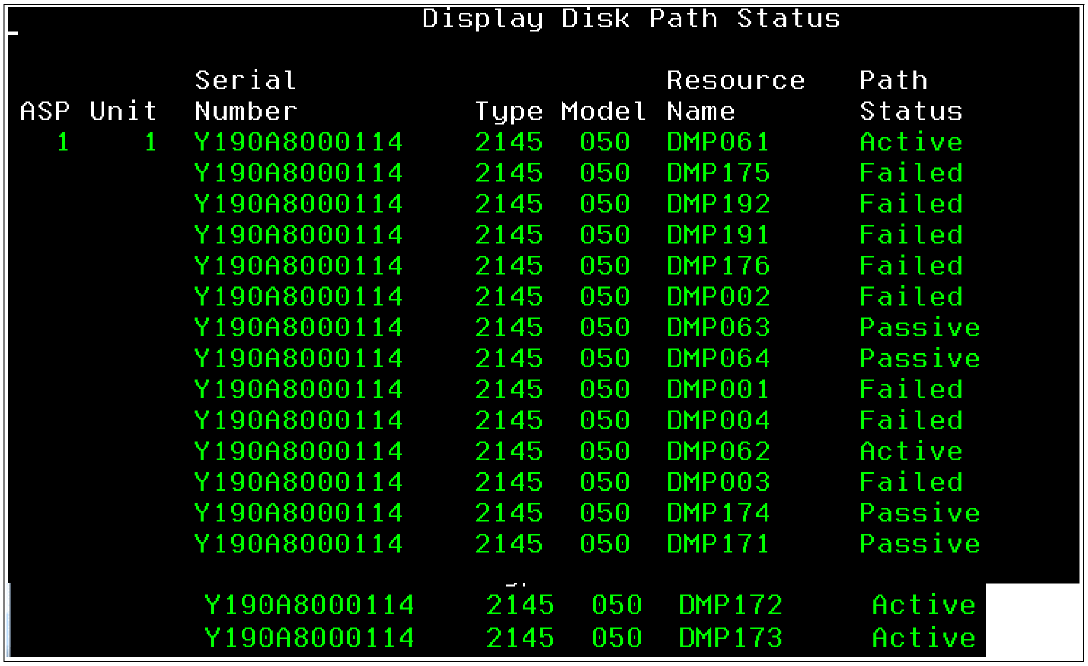


Figure 5-1 Test: I/O group on Site 1 fails

The IBM i paths during the outage are shown in Figure 5-2. Each DMPxxx resource represents one path to the LUN. Note that now different DMPxxx resources are active than before the outage. This means that different paths are used for the I/O rate.



Display Disk Path Status

ASP	Unit	Serial Number	Type	Model	Resource Name	Path Status
1	1	Y190A8000114	2145	050	DMP061	Active
		Y190A8000114	2145	050	DMP175	Failed
		Y190A8000114	2145	050	DMP192	Failed
		Y190A8000114	2145	050	DMP191	Failed
		Y190A8000114	2145	050	DMP176	Failed
		Y190A8000114	2145	050	DMP002	Failed
		Y190A8000114	2145	050	DMP063	Passive
		Y190A8000114	2145	050	DMP064	Passive
		Y190A8000114	2145	050	DMP001	Failed
		Y190A8000114	2145	050	DMP004	Failed
		Y190A8000114	2145	050	DMP062	Active
		Y190A8000114	2145	050	DMP003	Failed
		Y190A8000114	2145	050	DMP174	Passive
		Y190A8000114	2145	050	DMP171	Passive
		Y190A8000114	2145	050	DMP172	Active
		Y190A8000114	2145	050	DMP173	Active

Figure 5-2 IBM i paths during the outage of IO group on Site 1

5.1.2 Outage of Power hardware on Site 1

In Power HMC, we power down the IBM i partition on site 1 to simulate the failure of Power hardware on site1.

Fail-over to site 2

Perform the following steps to fail-over IBM i workload to site 2:

1. In the HyperSwap cluster, unmap HyperSwap LUNs from the Host of IBM i on Site 1, and map the LUNs to the host of IBM i on Site 2, as can be seen in Figure 5-3 and Figure 5-4 on page 37.

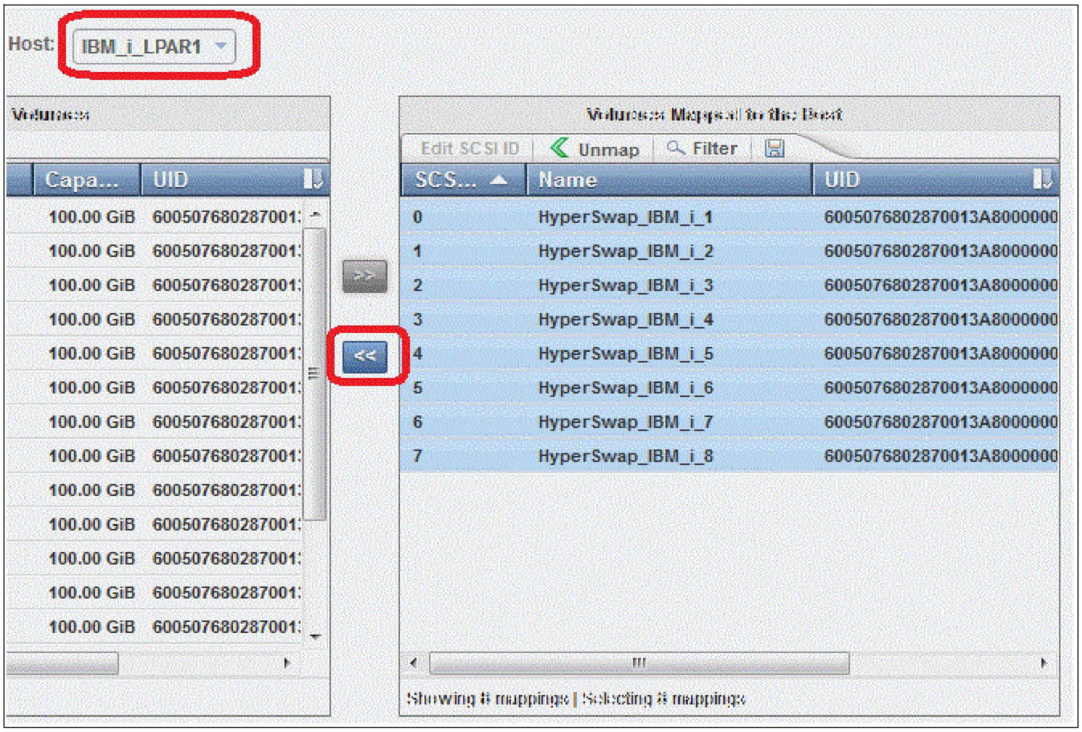


Figure 5-3 Power HW on site 1 falls: unmap the LUNs

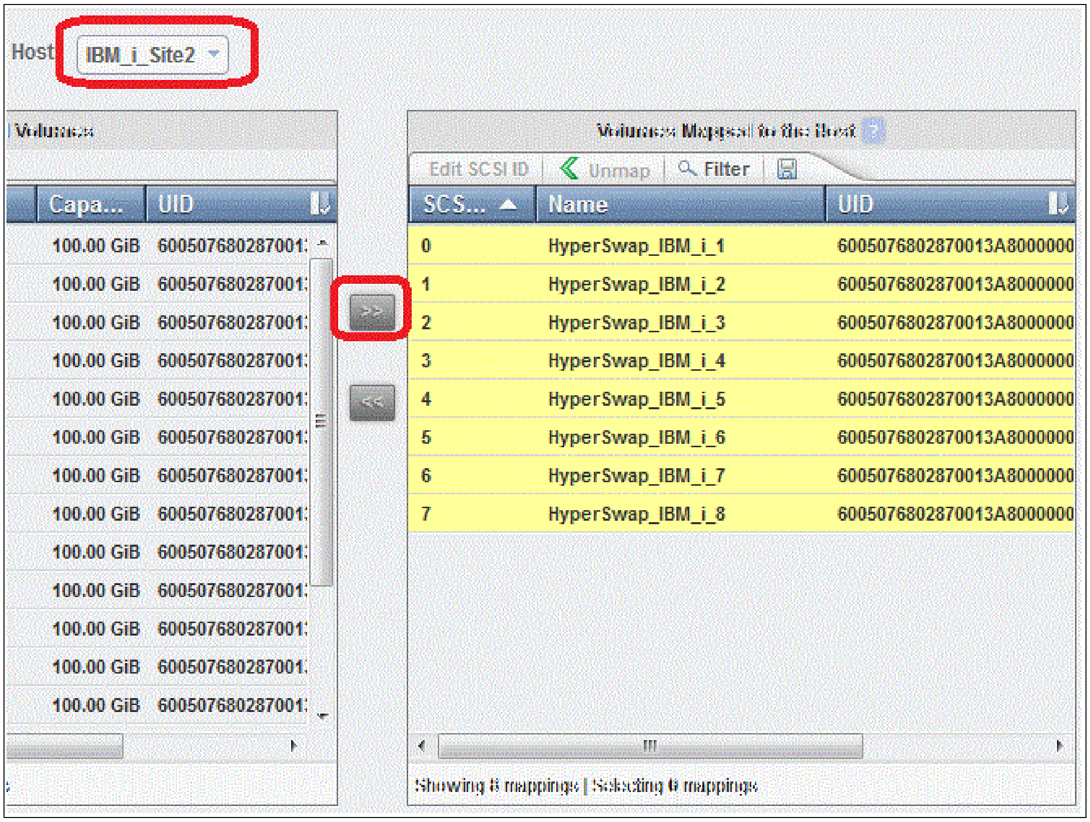
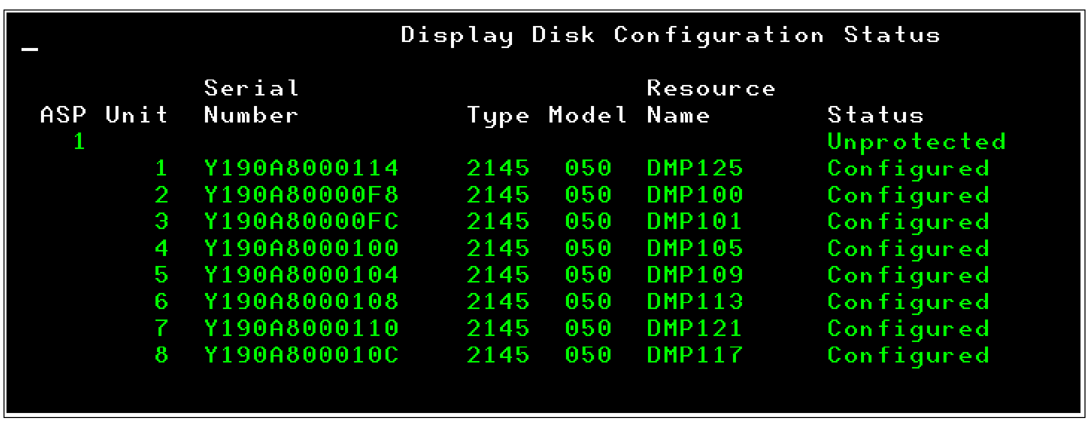


Figure 5-4 Power HW on site 1 falls: remap the LUNs

2. In POWER HMC, activate the IBM i partition at Site 2 and perform an IPL.

   The LUNs in IBM i on site 2 are shown in Figure 5-5. By LUN serial number you can see that they are the same LUNs as in IBM i on site 1.



Display Disk Configuration Status

ASP	Unit	Serial Number	Type	Model	Resource Name	Status
1						Unprotected
	1	Y190A8000114	2145	050	DMP125	Configured
	2	Y190A80000F8	2145	050	DMP100	Configured
	3	Y190A80000FC	2145	050	DMP101	Configured
	4	Y190A8000100	2145	050	DMP105	Configured
	5	Y190A8000104	2145	050	DMP109	Configured
	6	Y190A8000108	2145	050	DMP113	Configured
	7	Y190A8000110	2145	050	DMP121	Configured
	8	Y190A800010C	2145	050	DMP117	Configured

Figure 5-5 IBM i LUNs on site 2

The LUNs in IBM i on site 2 have 4 active paths and 12 passive paths, as can be seen in Figure 5-6.

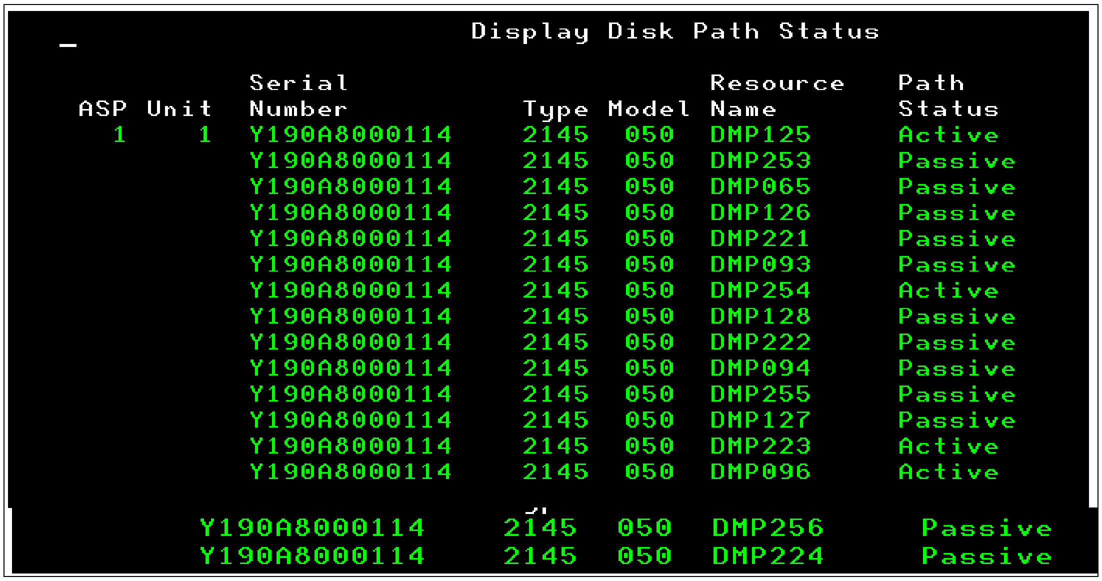


Figure 5-6 Paths to IBM i LUNs on site 2

Failback to Site 1 after outage

Perform the following steps to failback to site 1 after the outage is ended:

1. End the jobs in IBM i on site 2 and power down IBM i using the CLI.
2. After IBM i is powered-down, unmap the IBM i LUNs from the host of site 2 in Storwize V7000 and map them to the Host of site 1. This can be seen in Figure 5-7 and Figure 5-8 on page 39.

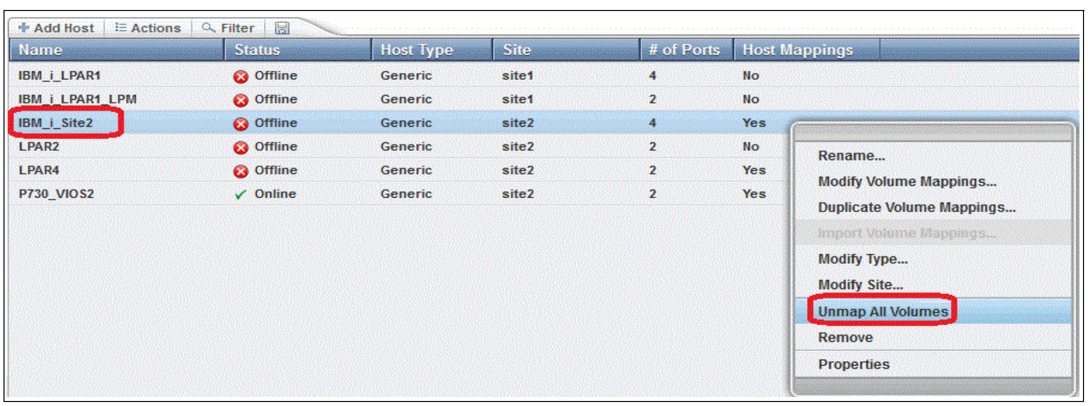


Figure 5-7 Fail-back: unmap LUNs in Storwize

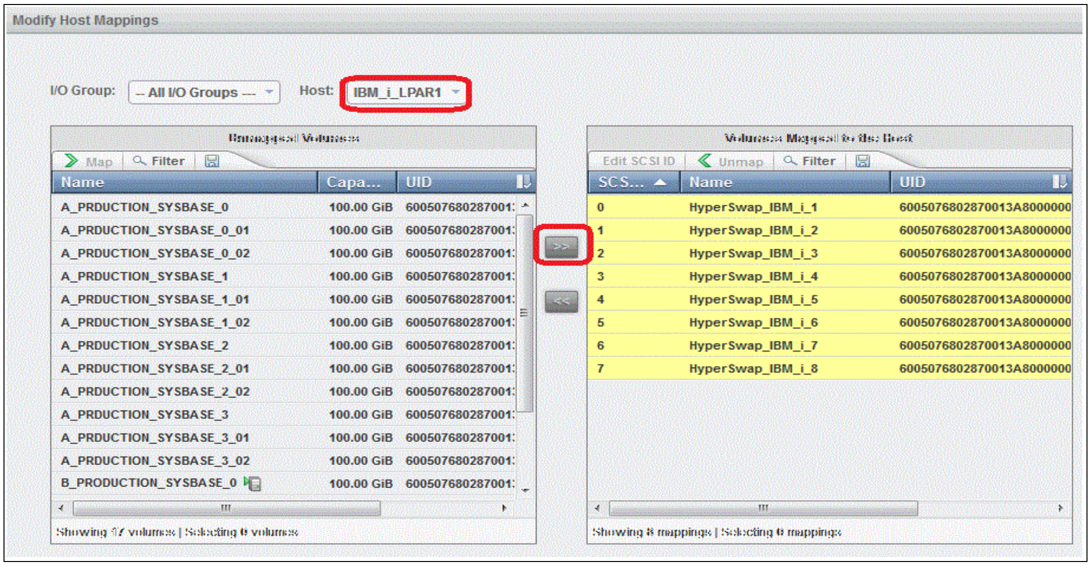


Figure 5-8 Failback: remap LUNs in Storwize

3. IPL IBM i on site 1.

   The IO rate on the nodes is captured by IBM Spectrum Control, and it is shown on the IBM Spectrum Control graph in Figure 5-9.

   The IO rate is initially running on both nodes on site 1: node 1 and node 2. After starting the workload in IBM i on site 2 HyperSwap transfers the workload on node 3 and node 4 on site 2.

   After we failback to site 1, HyperSwap transfers the IO rate to nodes on site 1.

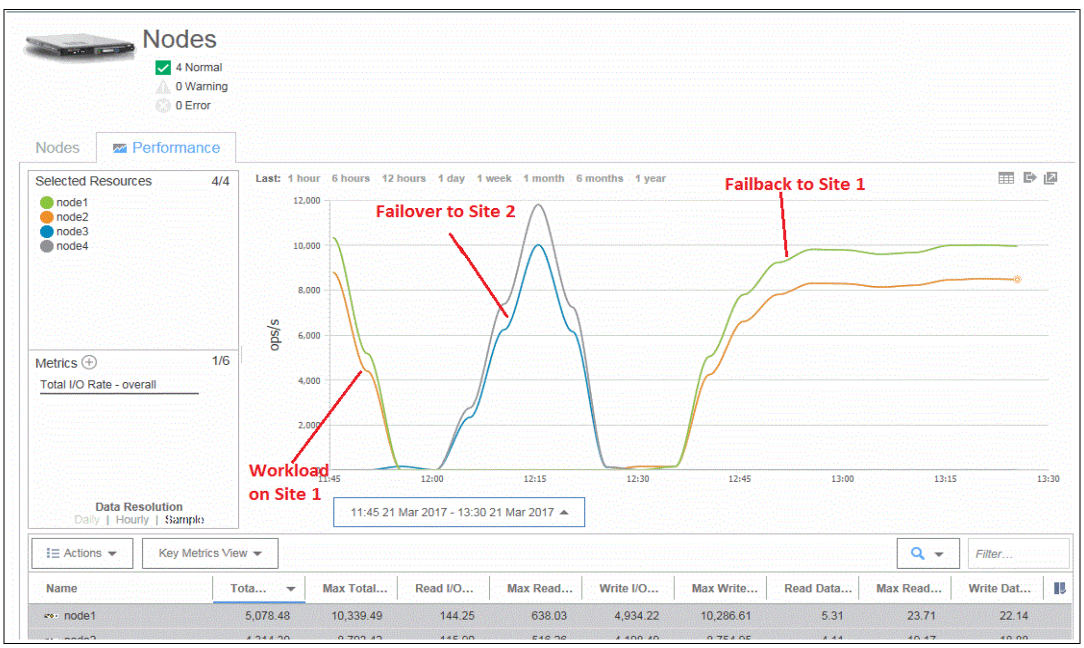


Figure 5-9 IBM Spectrum Control graph of I/O rate

5.1.3 Disaster at site 1

To simulate a disaster at Site 1, we trigger the outage of both Power hardware and Storwize node 1 and node 2. For this, we use Power HMC to power down the IBM i LPAR. At the same time, we use the Storwize GUI to Enter Service State for nodes 1 and 2.

Failover to site 2

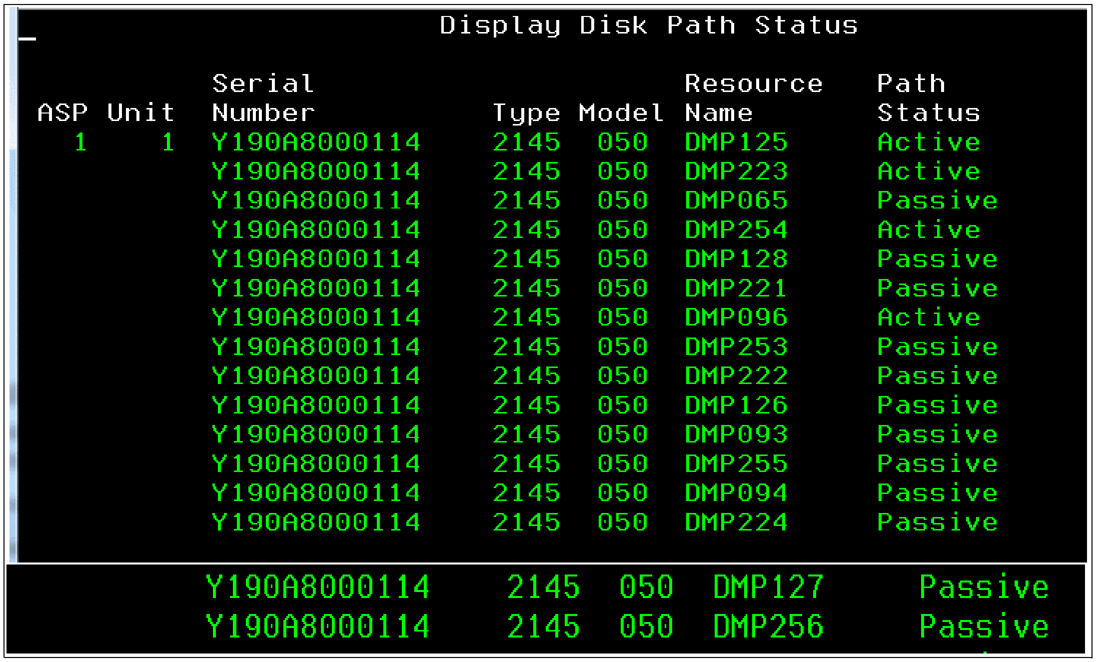
After the simulated disaster, we perform the following steps to fail over to site 2. In Storwize, we remap the IBM i LUNs to the host of IBM i on site 2:

1. In HMC on site 2, we IPL IBM i on site 2, and restart workload on site 2.
2. After failover, there are 8 paths in IBM_i_site2: 4 active and 4 passive. Failed paths are not indicated because we started the partition after failure of paths to node 1 and node 2. The paths in IBM i on site2 are shown in Figure 5-10.

ASP	Unit	Serial Number	Type	Model	Resource Name	Path Status
1	1	Y190A8000114	2145	050	DMP125	Active
		Y190A8000114	2145	050	DMP223	Active
		Y190A8000114	2145	050	DMP065	Passive
		Y190A8000114	2145	050	DMP254	Active
		Y190A8000114	2145	050	DMP128	Passive
		Y190A8000114	2145	050	DMP221	Passive
		Y190A8000114	2145	050	DMP096	Active
		Y190A8000114	2145	050	DMP253	Passive
1	2	Y190A80000F8	2145	050	DMP226	Active
		Y190A80000F8	2145	050	DMP195	Passive
		Y190A80000F8	2145	050	DMP068	Passive
		Y190A80000F8	2145	050	DMP097	Passive
		Y190A80000F8	2145	050	DMP193	Active
		Y190A80000F8	2145	050	DMP227	Passive
		Y190A80000F8	2145	050	DMP066	Active
		Y190A80000F8	2145	050	DMP099	Active

Figure 5-10 Paths after failover

3. The I/O rate is transferred to Storwize node 3 and node 4 on Site 2.
4. For the "end of the disaster" we exit Service state on node 1 and node 2 for the Storwize system using the GUI, and then power on IBM i using the PowerHMC. After enabling the nodes at Site 1, IBM i Site 2 can now see 16 paths: 8 active and 8 passive. This is shown in Figure 5-11 on page 41.



Display Disk Path Status

ASP	Unit	Serial Number	Type	Model	Resource Name	Path Status
1	1	Y190A8000114	2145	050	DMP125	Active
		Y190A8000114	2145	050	DMP223	Active
		Y190A8000114	2145	050	DMP065	Passive
		Y190A8000114	2145	050	DMP254	Active
		Y190A8000114	2145	050	DMP128	Passive
		Y190A8000114	2145	050	DMP221	Passive
		Y190A8000114	2145	050	DMP096	Active
		Y190A8000114	2145	050	DMP253	Passive
		Y190A8000114	2145	050	DMP222	Passive
		Y190A8000114	2145	050	DMP126	Passive
		Y190A8000114	2145	050	DMP093	Passive
		Y190A8000114	2145	050	DMP255	Passive
		Y190A8000114	2145	050	DMP094	Passive
		Y190A8000114	2145	050	DMP224	Passive
		Y190A8000114	2145	050	DMP127	Passive
		Y190A8000114	2145	050	DMP256	Passive

Figure 5-11 Paths after enabling nodes at Site 1

Failback to site 1

To fail back, complete the following steps:

1. Power down IBM_i_site2 from the operating system.
2. Change the LUN mapping to the host of IBM i on site 1, and start IBM i on site 1.
3. Restart workload on site 1.

The IBM Spectrum Control graph shows the I/O rate during failover to Site 2 and failback to Site 1, and is shown in Figure 5-12 on page 41.

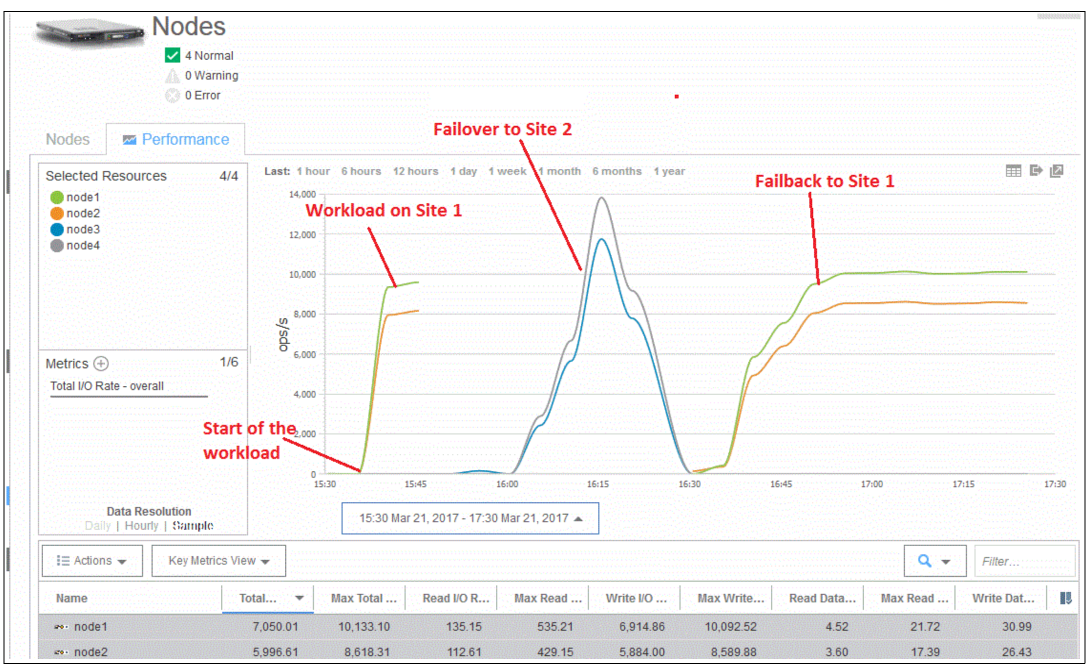


Figure 5-12 IBM Spectrum Control graph of I/O rate

5.2 Business continuity scenarios with Live Partition Mobility and Storwize HyperSwap

Our setup scenario fulfils the following requirements for Live Partition Mobility (LPM):

- IBM POWER7® firmware level.
- SSH Communication between the two consoles enabled.
- IBM i partition has only virtual resources: both disk capacity and Ethernet connected are virtualized.
- The second WWPN of virtual FC adapters in IBM i is zoned in the SAN switches the same way as the first WWPN.
- There is a Host in Storwize for the second WWPN, the HyperSwap LUNs are connected to both hosts: one with the first WWPN and one with the second WWPN.

For the complete checklist of prerequisites for LPM, see *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590.

Note: The Storwize Host with the second WWPN is assigned to site 2, as can be seen in Figure 5-20 on page 46.

5.2.1 Planned outage with Live Partition Mobility and Storwize HyperSwap

To validate that the IBM I partition is feasible for LPM, complete the following steps:

1. Select the partition in HMC. From the pull-down, select **Operations** → **Mobile** → **Validate**, as shown in Figure 5-13.

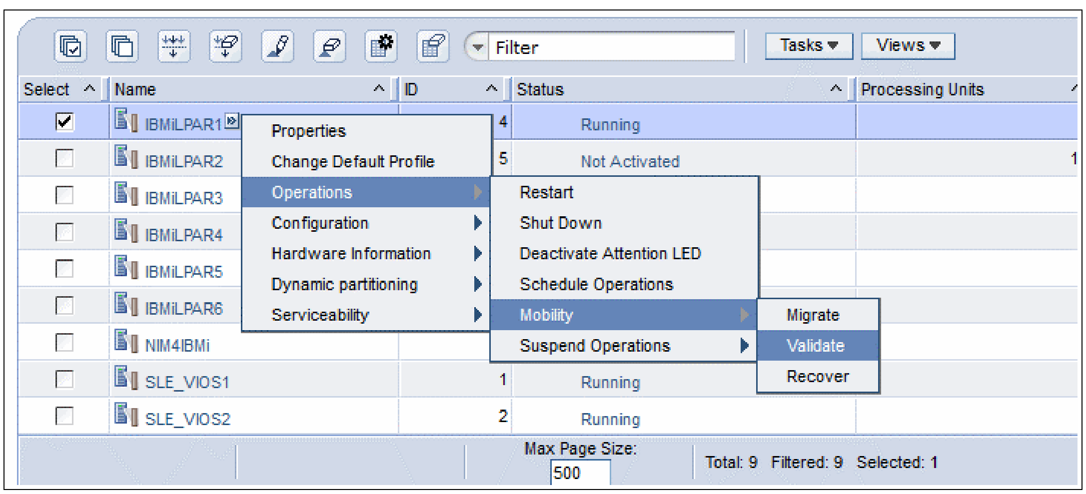


Figure 5-13 Validate LPM

2. Insert the IP address and userid of the server at Site 2 to migrate to, as shown in Figure 5-14.

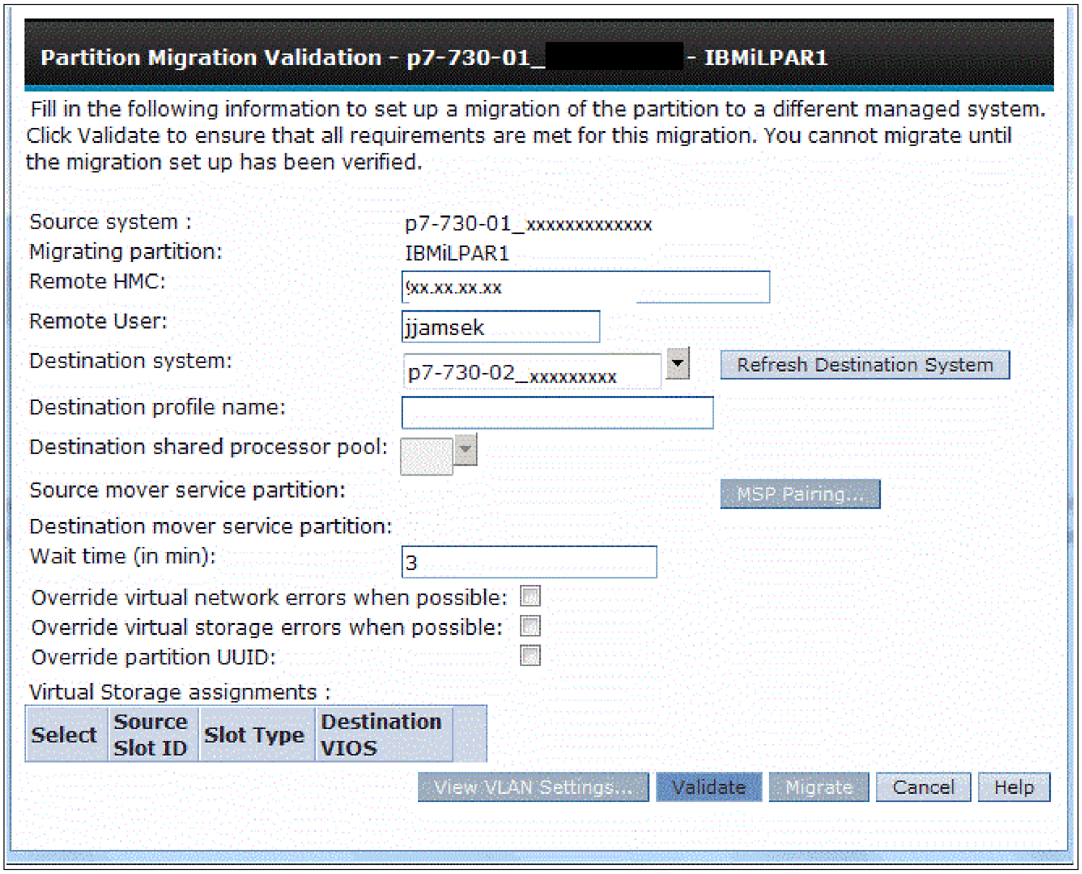


Figure 5-14 Insert credentials of site 2

3. After successful validation, click Migrate as shown in Figure 5-15.

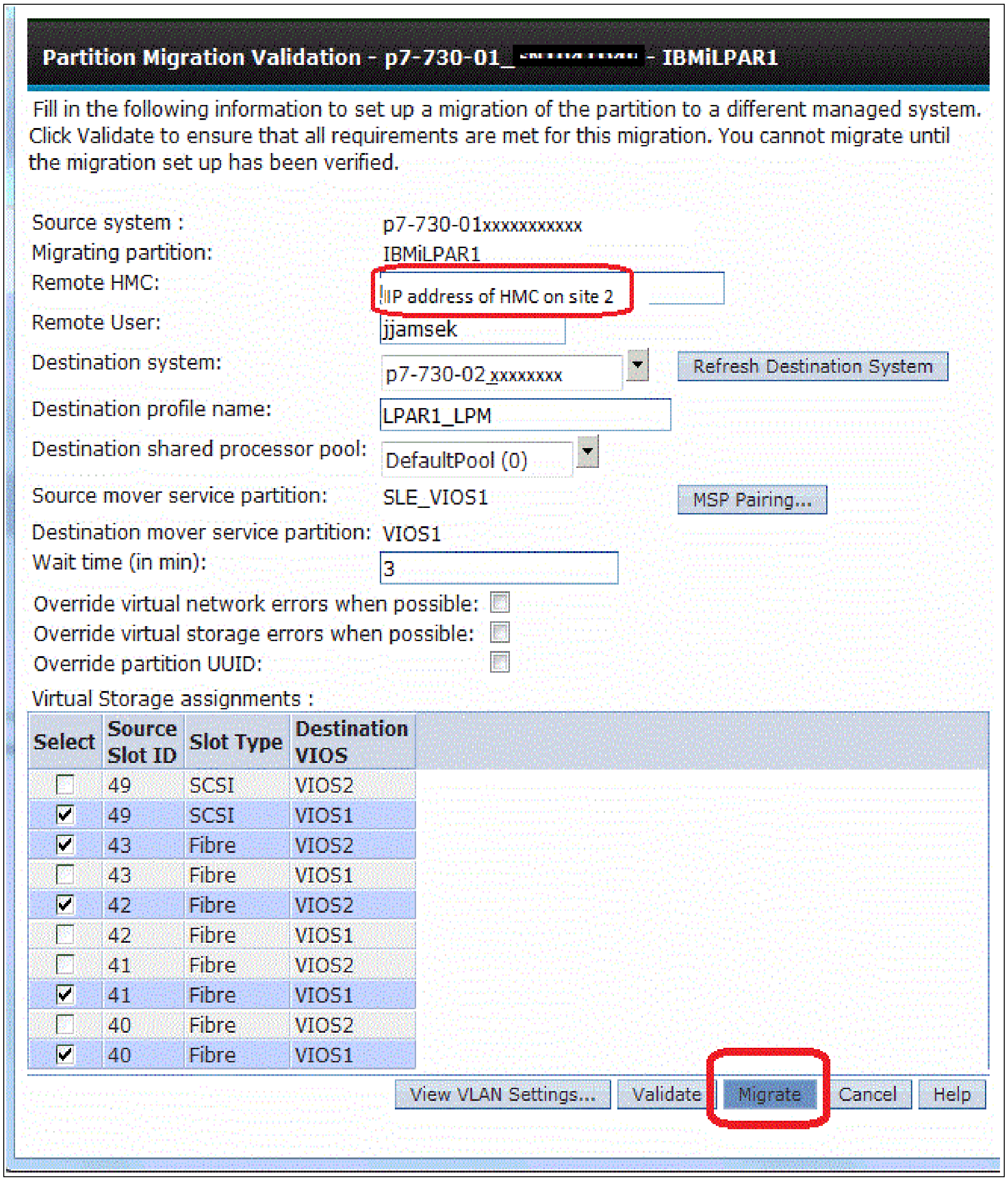


Figure 5-15 Migrate to site 2

During migration, the progress status is shown in an HMC window, as can be seen in Figure 5-16. When migration is finished, the window shows success (Figure 5-17).

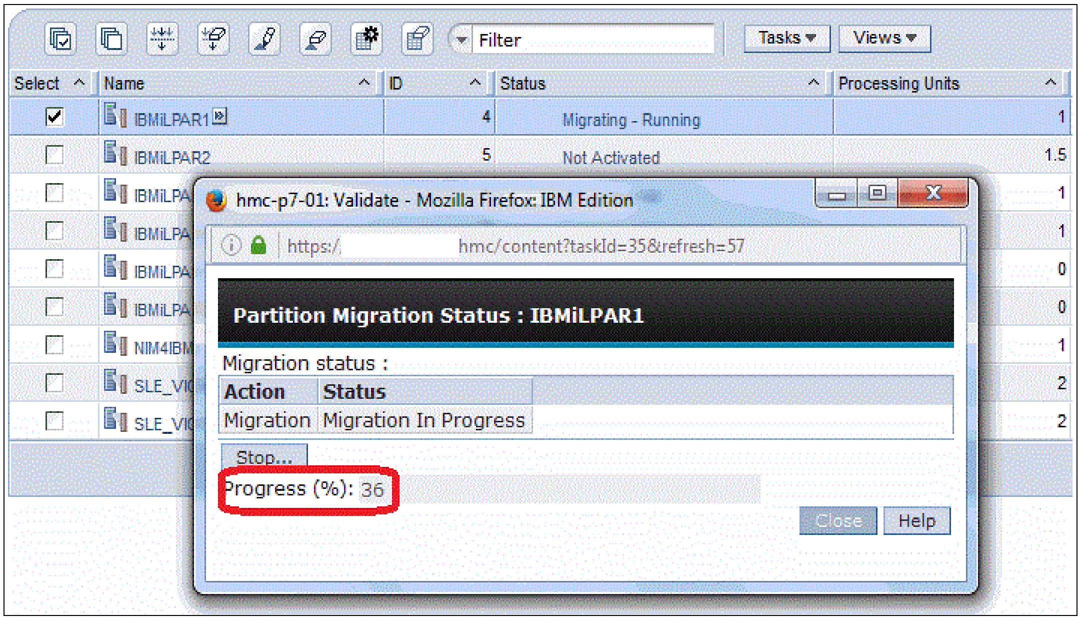


Figure 5-16 Migration progress

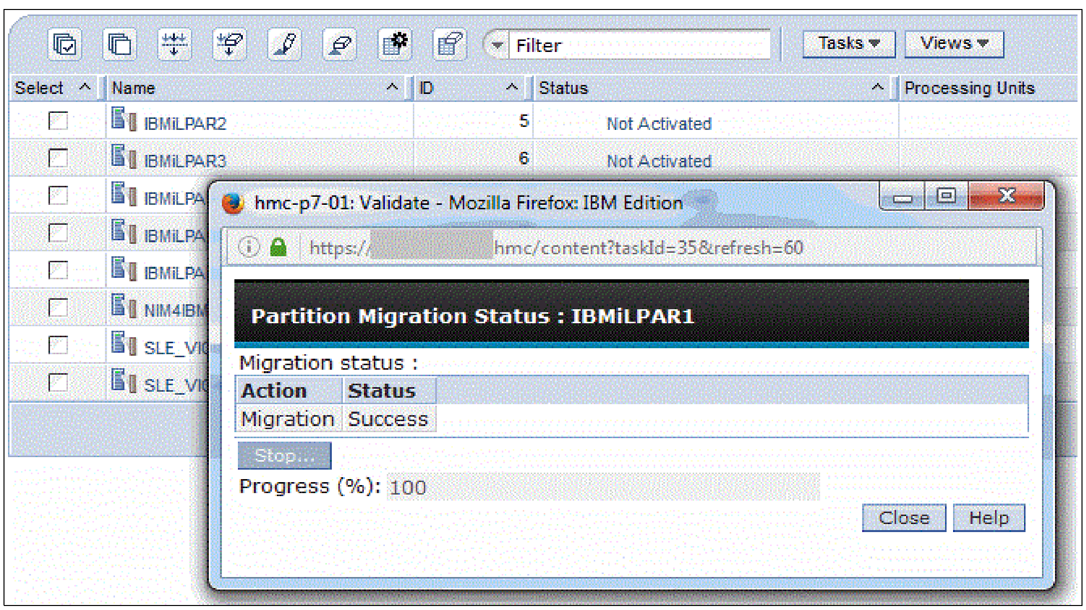
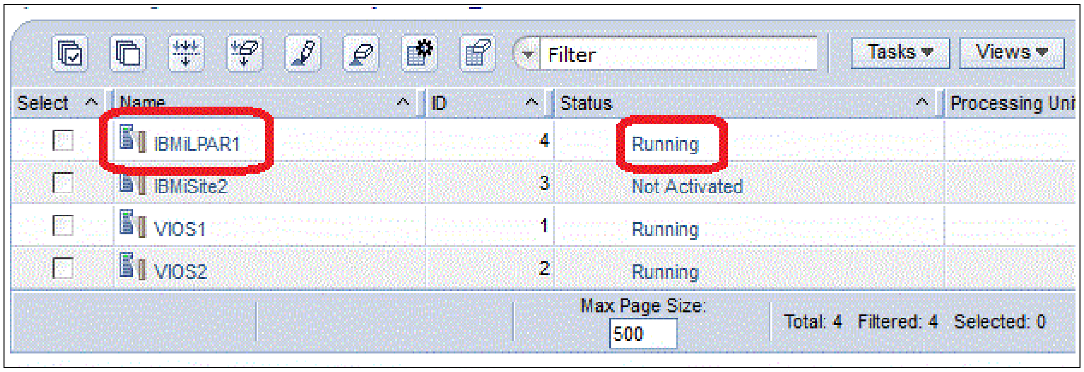


Figure 5-17 Migration success

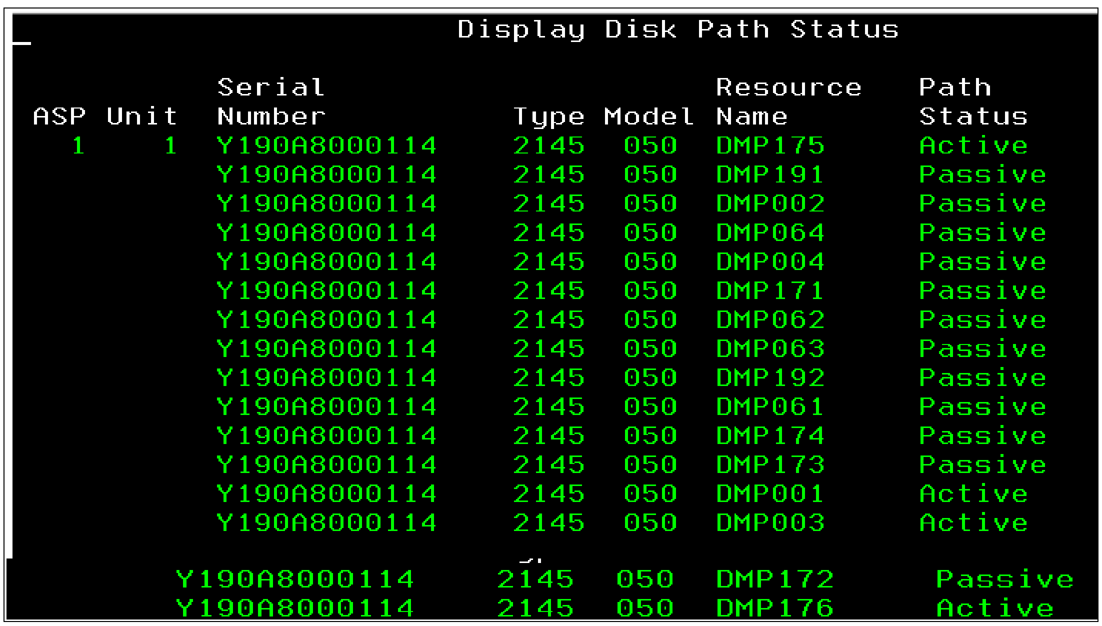
During migration, the IBM i partition with the workload is running, and after migration the IBM i LPAR runs in the Power server at site 2, as can be seen in Figure 5-18.



Select	Name	ID	Status	Processing Uni
☐	IBMiLPAR1	4	Running	
☐	IBMiSite2	3	Not Activated	
☐	VIOS1	1	Running	
☐	VIOS2	2	Running	

Figure 5-18 Migrated LPAR

IBM i workload is running all the time, and there aren't any messages in QSYSOPR. During migration, the I/O rate transfers from the Storwize V7000 nodes at site 1 to the nodes at site 2. The paths to IBM i LUNs in the migrated partition are shown in Figure 5-19.

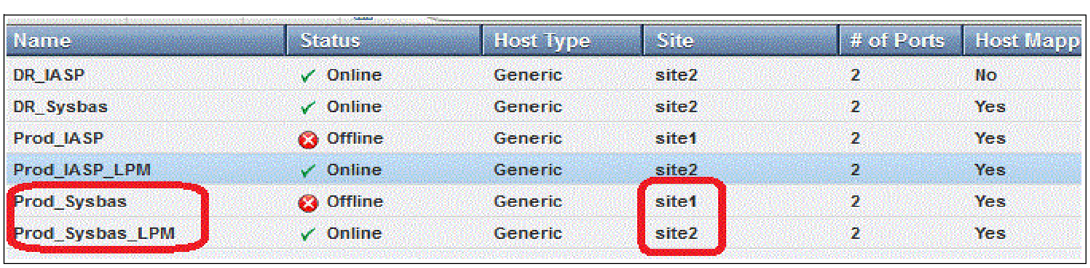


Display Disk Path Status

ASP	Unit	Serial Number	Type	Model	Resource Name	Path Status
1	1	Y190A8000114	2145	050	DMP175	Active
		Y190A8000114	2145	050	DMP191	Passive
		Y190A8000114	2145	050	DMP002	Passive
		Y190A8000114	2145	050	DMP064	Passive
		Y190A8000114	2145	050	DMP004	Passive
		Y190A8000114	2145	050	DMP171	Passive
		Y190A8000114	2145	050	DMP062	Passive
		Y190A8000114	2145	050	DMP063	Passive
		Y190A8000114	2145	050	DMP192	Passive
		Y190A8000114	2145	050	DMP061	Passive
		Y190A8000114	2145	050	DMP174	Passive
		Y190A8000114	2145	050	DMP173	Passive
		Y190A8000114	2145	050	DMP001	Active
		Y190A8000114	2145	050	DMP003	Active
		Y190A8000114	2145	050	DMP172	Passive
		Y190A8000114	2145	050	DMP176	Active

Figure 5-19 Paths to IBM i LUNs in migrated LPAR

The Storwize V7000 hosts after partition is migrated are shown in Figure 5-20.



Name	Status	Host Type	Site	# of Ports	Host Mapp
DR_IASP	Online	Generic	site2	2	No
DR_Sysbas	Online	Generic	site2	2	Yes
Prod_IASP	Offline	Generic	site1	2	Yes
Prod_IASP_LPM	Online	Generic	site2	2	Yes
Prod_Sysbas	Offline	Generic	site1	2	Yes
Prod_Sysbas_LPM	Online	Generic	site2	2	Yes

Figure 5-20 Storwize hosts after partition is migrated

5.2.2 Migration back to site 1

To migrate back to site 1, complete the following steps:

1. When the planned outage is finished, validate the migration back to site 1.
2. After successful validation, start migration by clicking the **Migration** button, as shown in Figure 5-21.

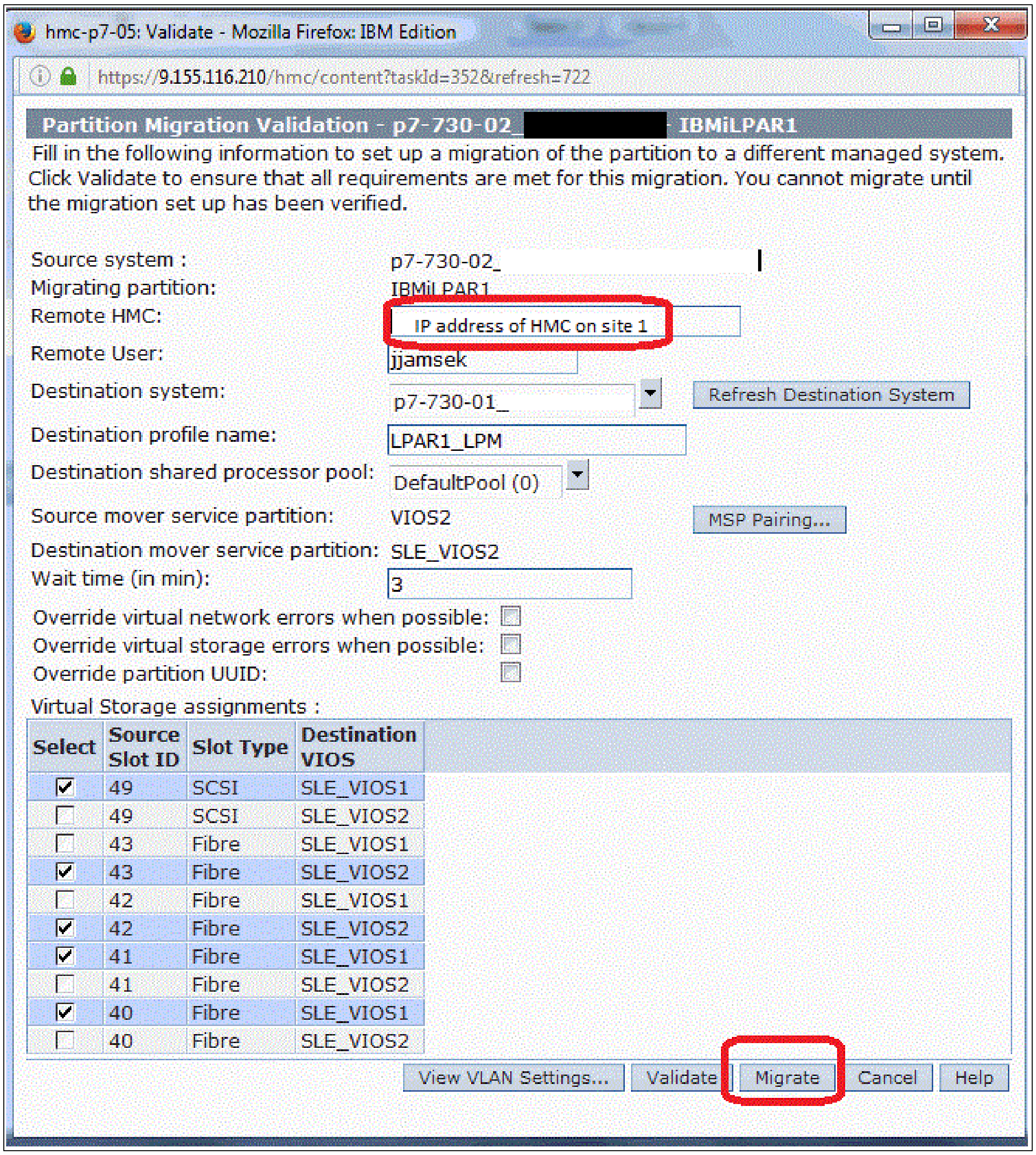


Figure 5-21 Migrate back to site 1

3. After showing the progress (Figure 5-22), the migration back successfully finishes.

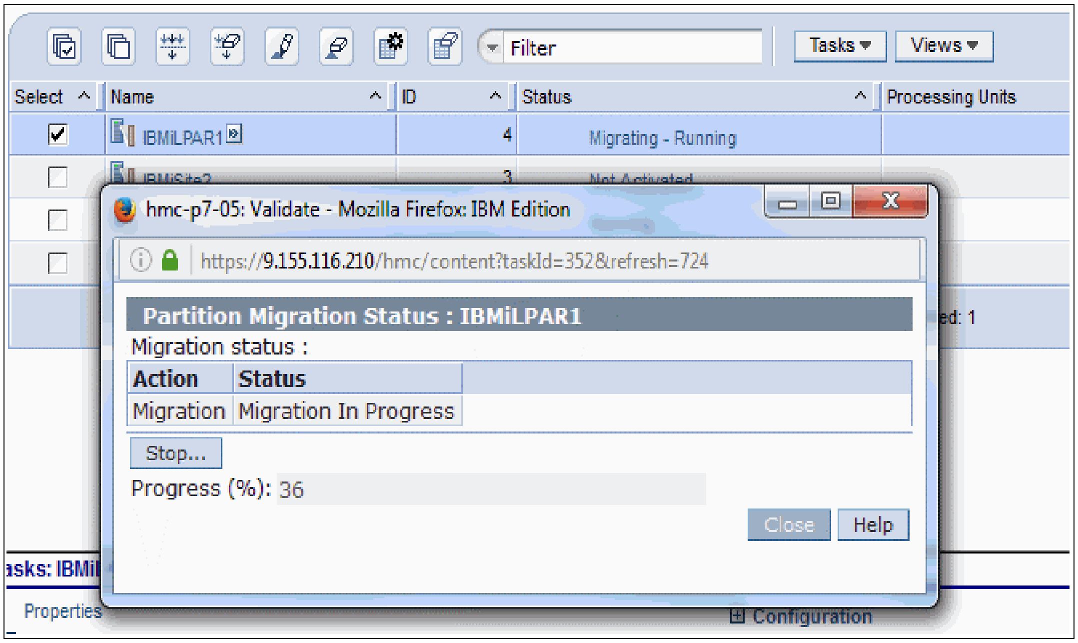


Figure 5-22 Progress of migration back to site 1

I/O rate on the Storwize nodes, as captured by IBM Spectrum Control, is shown on Figure 5-23. At LPM, migrating to site 2 the I/O rate transfers from the nodes on site 1 to the nodes on site 2. When migrating back, the I/O rate transfers to the nodes on site 1.

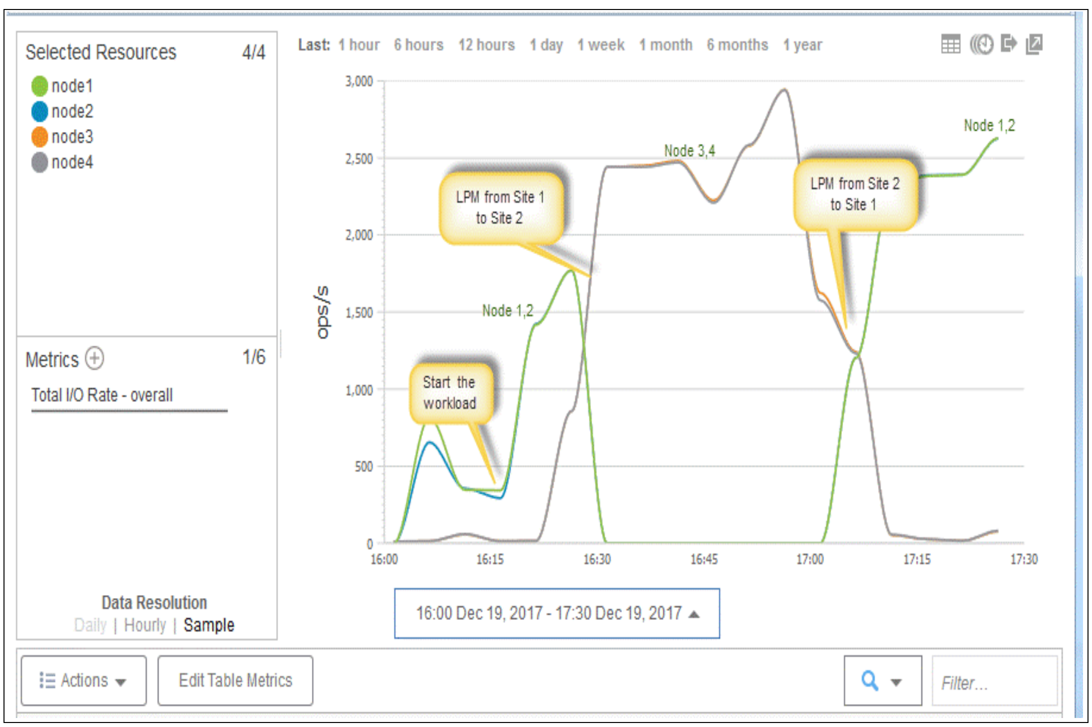


Figure 5-23 IBM Spectrum Control graph of I/O rate on nodes

6

PowerHA SystemMirror for IBM i with Storwize HyperSwap

In this chapter we describe solutions with IBM i IASP and HyperSwap, and the solutions are managed by PowerHA for i.

6.1 Implementation of PowerHA for i and Storwize HyperSwap

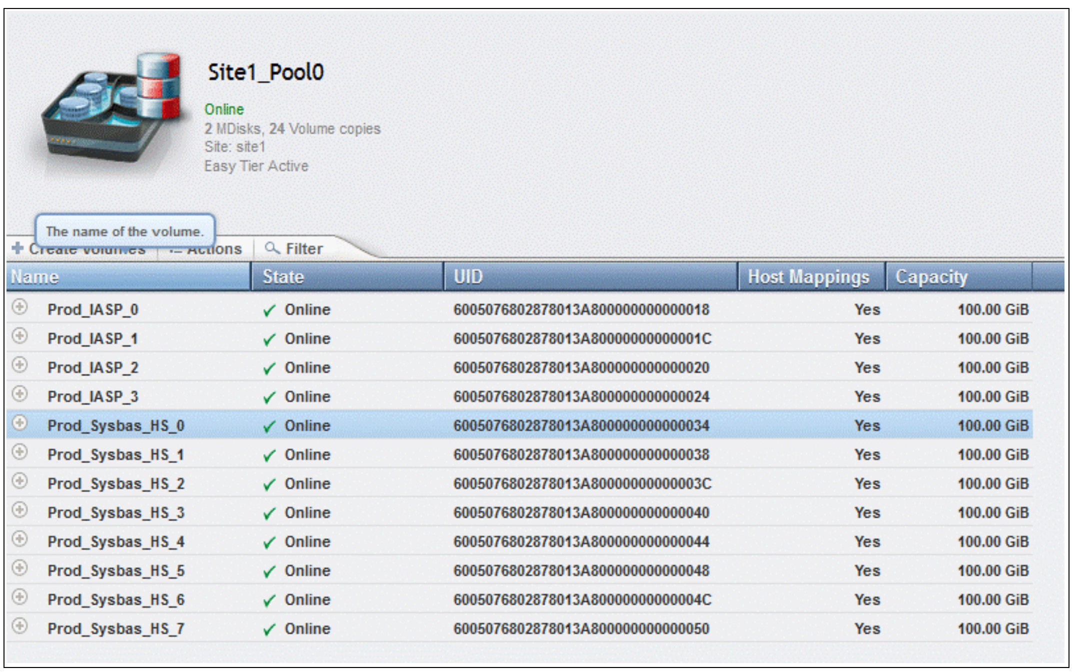
For test purposes, we use only *one* fabric for the setup of PowerHA for IBM i with Storwize HyperSwap. However, we strongly advise you to use two fabrics in actual implementations.

We create eight HyperSwap LUNs at Site 1, connect them to the production partition `IBMiLPAR1`, we install IBM i in the LPAR; these are Sysbas LUNs on site 1. We create eight HyperSwap LUNs at Site 2, connect them to disaster recovery partition `IBMiDR1`, and install IBM i in this LPAR; these are Sysbas LUNs at site 2.

We create four HyperSwap LUNs for Independent Auxiliary Storage Pool (IASP) and connect them to partition `IBMiLPAR1`. For PowerHA for i implementations the IASP LUNs must be assigned to different virtual FC adapter than Sysbas (system LUNs, other than IASP), and the virtual FC adapter of IASP must be mapped to a different port in VIOS than the virtual FC of Sysbas.

> **Note:** Both Sysbas LUNs and IASP LUNs are HyperSwap LUNs. Production Sysbas LUNs and IASP LUNs are created in the disk pool at Site 1.

Figure 6-1 shows Sysbas and IASP LUNs at site 1, the production site.



Name	State	UID	Host Mappings	Capacity
Prod_IASP_0	Online	6005076802878013A800000000000018	Yes	100.00 GiB
Prod_IASP_1	Online	6005076802878013A80000000000001C	Yes	100.00 GiB
Prod_IASP_2	Online	6005076802878013A800000000000020	Yes	100.00 GiB
Prod_IASP_3	Online	6005076802878013A800000000000024	Yes	100.00 GiB
Prod_Sysbas_HS_0	Online	6005076802878013A800000000000034	Yes	100.00 GiB
Prod_Sysbas_HS_1	Online	6005076802878013A800000000000038	Yes	100.00 GiB
Prod_Sysbas_HS_2	Online	6005076802878013A80000000000003C	Yes	100.00 GiB
Prod_Sysbas_HS_3	Online	6005076802878013A800000000000040	Yes	100.00 GiB
Prod_Sysbas_HS_4	Online	6005076802878013A800000000000044	Yes	100.00 GiB
Prod_Sysbas_HS_5	Online	6005076802878013A800000000000048	Yes	100.00 GiB
Prod_Sysbas_HS_6	Online	6005076802878013A80000000000004C	Yes	100.00 GiB
Prod_Sysbas_HS_7	Online	6005076802878013A800000000000050	Yes	100.00 GiB

Figure 6-1 Sysbas and IASP LUNs at site 1

Figure 6-2 shows Sysbas LUNs at site 2, the DR site.

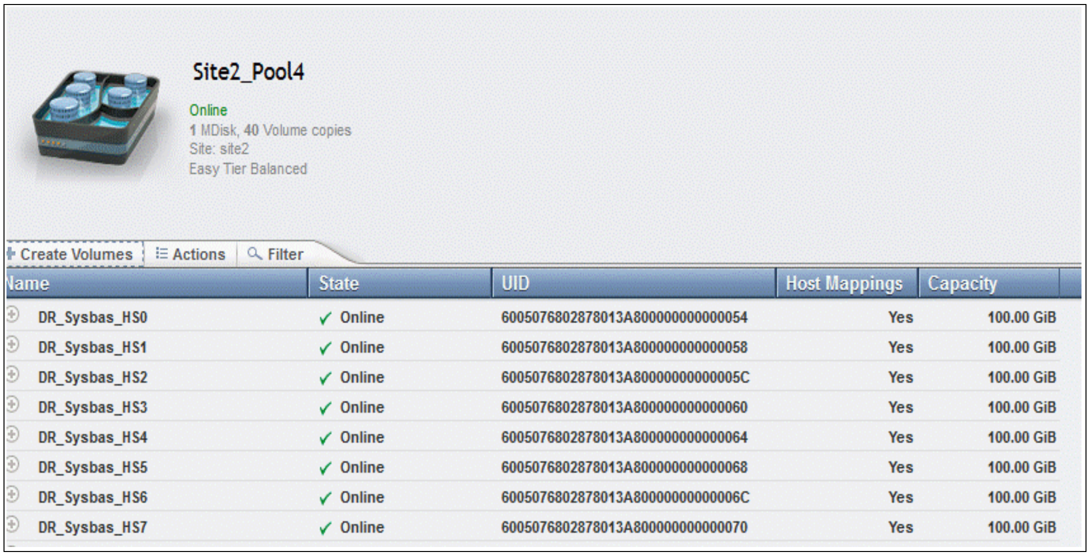
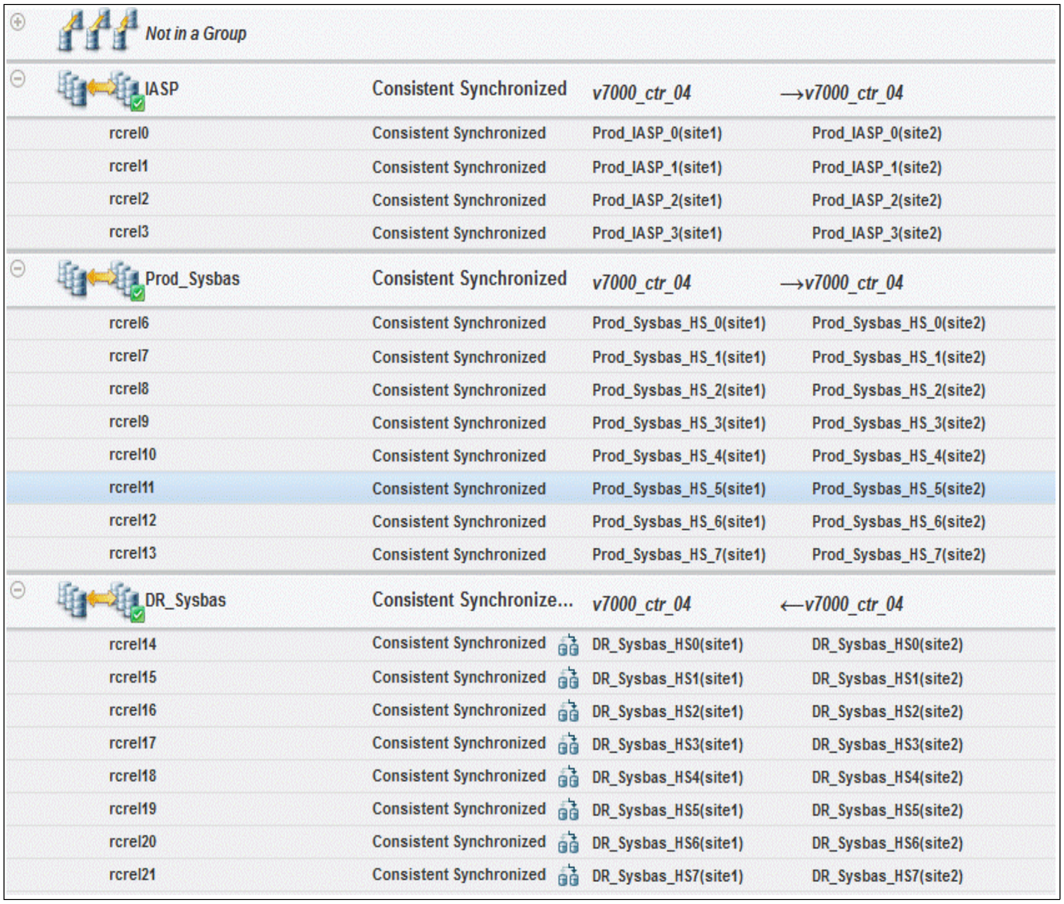


Figure 6-2 Sysbas LUNs at site 2

Metro Mirror of HyperSwap LUNs is automatically started. Figure 6-3 shows the Metro Mirror relationship of Sysbas and IASP LUNs at both sites. Note that the Metro Mirror direction of HyperSwap LUNs on site 2, is from site 2 to site 1, as is denoted by icons in the *Status* column.



Not in a Group			
IASP	Consistent Synchronized	*v7000_ctr_04*	*→v7000_ctr_04*
rcrel0	Consistent Synchronized	Prod_IASP_0(site1)	Prod_IASP_0(site2)
rcrel1	Consistent Synchronized	Prod_IASP_1(site1)	Prod_IASP_1(site2)
rcrel2	Consistent Synchronized	Prod_IASP_2(site1)	Prod_IASP_2(site2)
rcrel3	Consistent Synchronized	Prod_IASP_3(site1)	Prod_IASP_3(site2)
Prod_Sysbas	Consistent Synchronized	*v7000_ctr_04*	*→v7000_ctr_04*
rcrel6	Consistent Synchronized	Prod_Sysbas_HS_0(site1)	Prod_Sysbas_HS_0(site2)
rcrel7	Consistent Synchronized	Prod_Sysbas_HS_1(site1)	Prod_Sysbas_HS_1(site2)
rcrel8	Consistent Synchronized	Prod_Sysbas_HS_2(site1)	Prod_Sysbas_HS_2(site2)
rcrel9	Consistent Synchronized	Prod_Sysbas_HS_3(site1)	Prod_Sysbas_HS_3(site2)
rcrel10	Consistent Synchronized	Prod_Sysbas_HS_4(site1)	Prod_Sysbas_HS_4(site2)
rcrel11	Consistent Synchronized	Prod_Sysbas_HS_5(site1)	Prod_Sysbas_HS_5(site2)
rcrel12	Consistent Synchronized	Prod_Sysbas_HS_6(site1)	Prod_Sysbas_HS_6(site2)
rcrel13	Consistent Synchronized	Prod_Sysbas_HS_7(site1)	Prod_Sysbas_HS_7(site2)
DR_Sysbas	Consistent Synchronize...	*v7000_ctr_04*	*←v7000_ctr_04*
rcrel14	Consistent Synchronized	DR_Sysbas_HS0(site1)	DR_Sysbas_HS0(site2)
rcrel15	Consistent Synchronized	DR_Sysbas_HS1(site1)	DR_Sysbas_HS1(site2)
rcrel16	Consistent Synchronized	DR_Sysbas_HS2(site1)	DR_Sysbas_HS2(site2)
rcrel17	Consistent Synchronized	DR_Sysbas_HS3(site1)	DR_Sysbas_HS3(site2)
rcrel18	Consistent Synchronized	DR_Sysbas_HS4(site1)	DR_Sysbas_HS4(site2)
rcrel19	Consistent Synchronized	DR_Sysbas_HS5(site1)	DR_Sysbas_HS5(site2)
rcrel20	Consistent Synchronized	DR_Sysbas_HS6(site1)	DR_Sysbas_HS6(site2)
rcrel21	Consistent Synchronized	DR_Sysbas_HS7(site1)	DR_Sysbas_HS7(site2)

Figure 6-3 Metro Mirror relationships

We create Storwize hosts for both production and disaster recovery (DR) IBM i. The IASP LUNs are initially mapped to the production IBM i host, as can be seen in Figure 6-4. During the switch of IASP to the DR site, PowerHA remaps the LUNs to the DR Host.

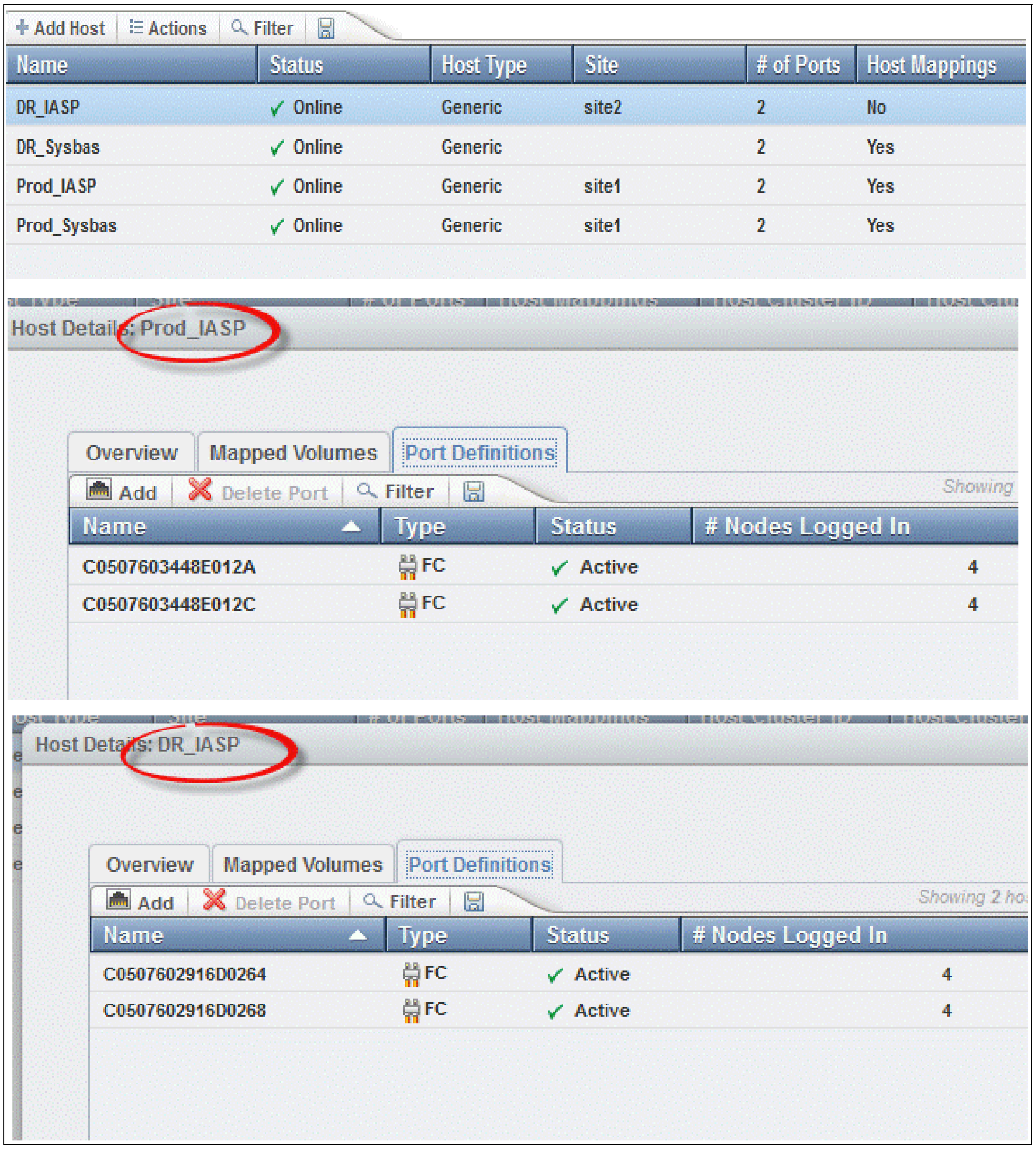


Figure 6-4 Hosts of HyperSwap LUNs

Zoning is done so that every IBM i virtual FC is zoned with all four nodes in the Storwize V7000 cluster. The two zones shown in Figure 6-5 are for the virtual FC adapters of the IASP connection of the production partition.

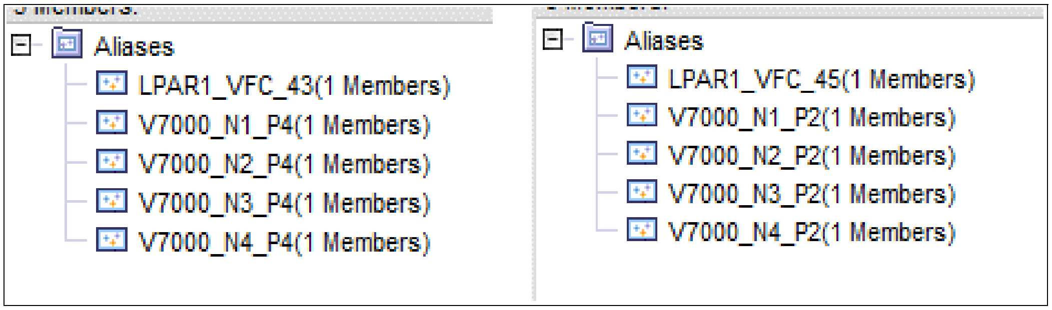


Figure 6-5 Zoning of Virtual FCs of IASP

We set up the environment for PowerHA for i by completing the following steps:

1. In both IBM i partitions `IBMiLPAR1` and `IBMiDR1`, start the TCP server `INETD`, by using the **`STRTCPSVR SERVER(*INETD)`** command.
2. In either partition `IBMiLPAR1` or `IBMiDR1`, we change the network attribute to allow the node to be added to the cluster, by using the **`CHGNETA ALWADDCLU(*ANY)`** command.
3. In partition `IBMiLPAR1` we create a cluster with IBM i LPARs `IBMiLPAR1`, system name `PROD` on Site 1, and `IBMiDR1`, system name `DISREC` on site 2, by using the IBM i **`CRTCLU CLUSTER(PWRHA_CLU) NODE((PROD ('x.x.x.x')) (DISREC ('x.x.x.x'))) CLUMSGQ(QSYS/QSYSOPR) FLVWAITTIM(1)`** command.
4. We start both nodes of the cluster, by using the following commands:
   a. In the production partition: **`STRCLUNOD CLUSTER(PWRHA_CLU) NODE(PROD)`**
   b. In the DR partition: **`STRCLUNOD CLUSTER(PWRHA_CLU) NODE(DISREC)`**
5. We create the device domain by using the following commands:
   a. In the production partition: **`ADDDEVDMNE CLUSTER(pwrha_clu) DEVDMN(pwrha_dmn) NODE(PROD)`**
   b. In the DR partition: **`ADDDEVDMNE CLUSTER(pwrha_clu) DEVDMN(pwrha_dmn) NODE(DISREC)`**
6. In the production partition `PROD`, we create the IASP of the HyperSwap LUNs by using the **`CFGDEVASP ASPDEV(PWRHA_IASP) ACTION(*CREATE)`** command.

   Because the IASP LUNs are connected with two virtual FC adapters, each of them zoned with four Storwize V7000 nodes, the IASP disk units have eight paths, two of them active and four passive, as can be seen in Figure 6-6.

Figure 6-6 Paths of IASP LUNs

7. On the node DISREC, we create the description of the IASP by using the following command:

```
CRTDEVASP DEVD(pwrha_iasp) RSRCNAME(pwrha_iasp) RDB(*GEN)
```

8. In partition PROD we generated the device CRG by using the following command:

```
CRTCRG CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG) CRGTYPE(*DEV) EXITPGM(*NONE)
USRPRF(*NONE) RCYDMN((PROD *PRIMARY *LAST SITE1 *NONE) (DISREC *BACKUP *LAST
SITE1 *NONE)) CFGOBJ((PWRHA_IASP *DEVD *ONLINE))
```

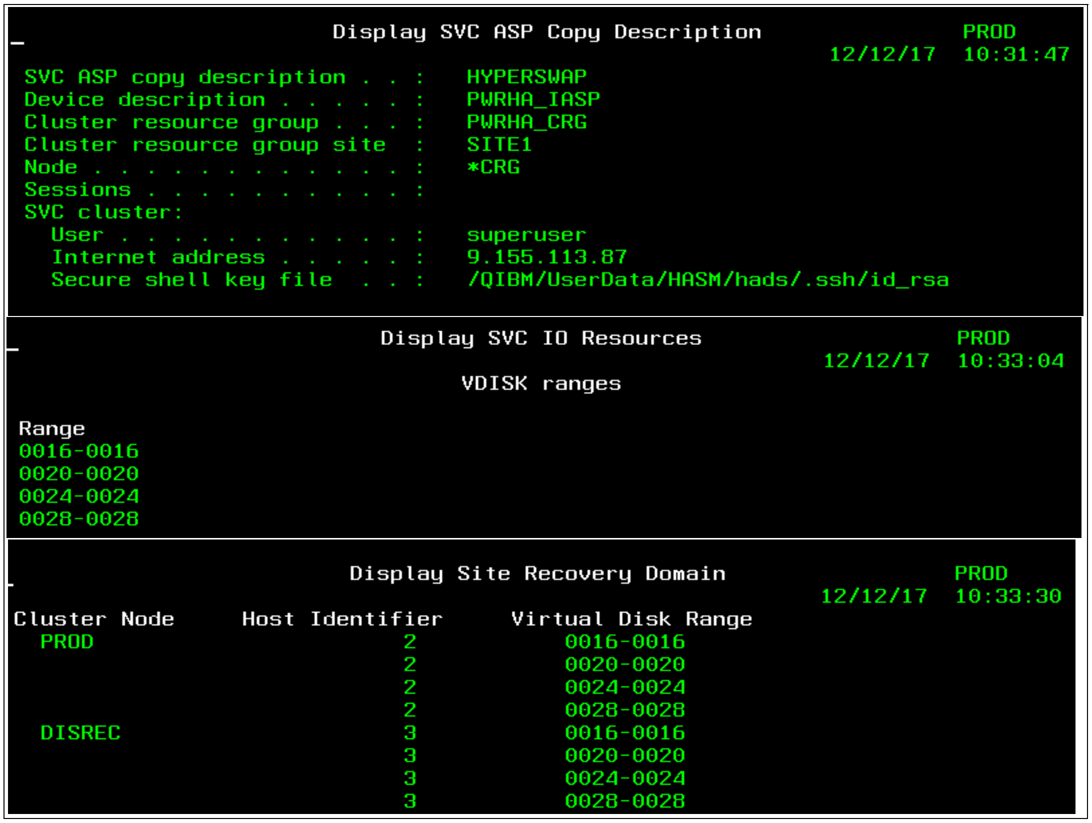
> **Note:** This site specified in the recovery domain is meaningful for PowerHA for IBM i and has nothing to do with the HyperSwap site. Both nodes in the recovery domain use site 1, because there is one Storwize V7000 cluster accessed by both partitions PROD and DISEC.

9. In partition PROD, we generate SSH keys for communication between IBM i and Storwize to enable PowerHA to use CLI commands to Storwize. For this we use the following steps:
   a. Start Qshell by using the **QSH** command.
   b. Use the **cd /QIBM/UserData/HASM/hads/ -> mkdir .ssh** command.
   c. Use the **cd /QIBM/UserData/HASM/hads/.ssh** command.
   d. Use the **ssh-keygen -t rsa -f id_rsa -N ''** command.
10. We transfer the private key to all IBM i nodes in the cluster using FTP to directory /QIBM/UserData/HASM/hads/.ssh, which we had created before. We transferred the public key to the Storwize V7000.
11. In both the PROD and DISREC partitions, we change the authority of PowerHA user QHAUSRPRF for the file with the private key, by using the **CHGAUT OBJ('/QIBM/UserData/HASM/hads/.ssh/id_rsa') USER(QHAUSRPRF) DTAAUT(*R)** command.
12. In partition PROD, we create the PowerHA Copy description for the switchable IASP for SVC LUN-level switching. For this we used the following command:

```
ADDSVCCPYD ASPCPY(HYPERSWAP) ASPDEV(PWRHA_IASP) CRG(PWRHA_CRG) SITE(SITE1)
NODE(*CRG) SVCHOST(superuser '/QIBM/UserData/HASM/hads/.ssh/id_rsa' 'x.x.x.x')
VRTDSKRNG((16 16 *ALL) (20 20 *ALL) (24 24 *ALL) (28 28 *ALL)) RCYDMN((PROD
(2)) (DISREC (3)))
```

    The IDs 16, 20, 24, and 28 in the VRTDSKRNG option are the Storwize LUN IDs of the IASP LUNs. IDs 2 and 3 in the RCYDMN option are Storwize IDs of the Host connections to the relevant LPARs.

Figure 6-7 shows the created copy description.


```
                     Display SVC ASP Copy Description                PROD
                                                           12/12/17  10:31:47
 SVC ASP copy description . . . :   HYPERSWAP
 Device description . . . . . . :   PWRHA_IASP
 Cluster resource group . . . . :   PWRHA_CRG
 Cluster resource group site    :   SITE1
 Node . . . . . . . . . . . . . :   *CRG
 Sessions . . . . . . . . . . . :
 SVC cluster:
   User . . . . . . . . . . . . :   superuser
   Internet address . . . . . . :   9.155.113.87
   Secure shell key file    . . :   /QIBM/UserData/HASM/hads/.ssh/id_rsa

                         Display SVC IO Resources                    PROD
                                                           12/12/17  10:33:04
                               VDISK ranges

Range
0016-0016
0020-0020
0024-0024
0028-0028

                      Display Site Recovery Domain                   PROD
                                                           12/12/17  10:33:30
Cluster Node       Host Identifier      Virtual Disk Range
  PROD                          2           0016-0016
                                2           0020-0020
                                2           0024-0024
                                2           0028-0028
  DISREC                        3           0016-0016
                                3           0020-0020
                                3           0024-0024
                                3           0028-0028
```

Figure 6-7 Display PowerHA Copy description

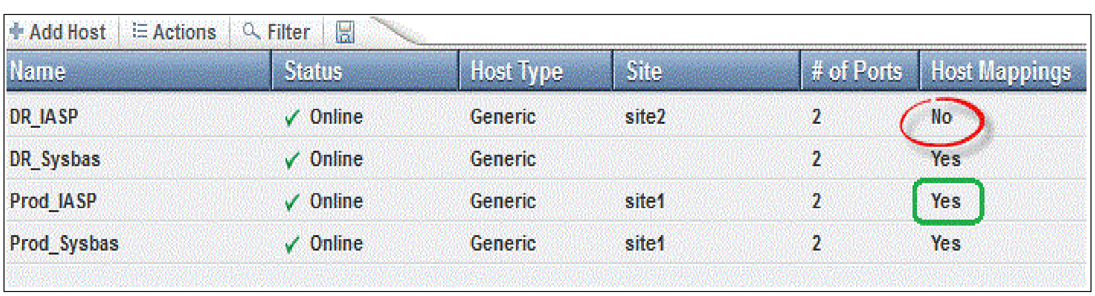
13. We started CRG by using the following command in partition PROD:

```
STRCRG CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG)
```

After performing these steps, the environment of PowerHA for IBM i with HyperSwap is ready. Next, we describe three tests of business continuity with PowerHA for i HyperSwap.

6.2 Planned outage of production IBM i

Before the outage, the Storwize V7000 host connection of production IASP LUNs is online and the host of DR IASP LUNs is offline. This can be seen in Figure 6-8.



Name	Status	Host Type	Site	# of Ports	Host Mappings
DR_IASP	Online	Generic	site2	2	No
DR_Sysbas	Online	Generic		2	Yes
Prod_IASP	Online	Generic	site1	2	Yes
Prod_Sysbas	Online	Generic	site1	2	Yes

Figure 6-8 Host connections before the outage

Before the outage, the direction of HyperSwap Metro Mirror is from site 1 (production site) to site 2 (DR site), as shown in Figure 6-9.

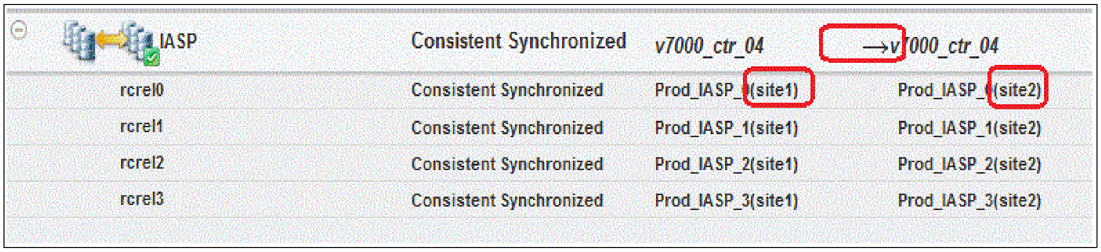


Figure 6-9 Metro Mirror before the outage

The IBM i workload is running in the production IASP. This can be seen in Figure 6-10, which shows IO operations on the IASP LUNs: disk units 4001-4004.

Unit	Type	Size (M)	% Used	I/O Rqs	Request Size (K)	Read Rqs	Write Rqs	Read (K)	Write (K)	% Busy
1	2145	95443	9.5	.7	10.6	.0	.6	24.2	8.9	0
9	2145	95443	4.1	.4	15.8	.1	.3	26.3	12.8	0
10	2145	95443	4.1	.6	15.2	.1	.5	23.0	13.0	0
11	2145	95443	4.1	.4	17.0	.1	.3	24.6	14.8	0
12	2145	95443	4.1	.4	16.3	.1	.2	20.1	14.9	0
13	2145	95443	4.1	.8	11.3	.1	.7	20.1	9.6	0
14	2145	95443	4.1	.6	12.9	.1	.5	23.0	10.8	0
15	2145	95443	4.1	.9	10.9	.1	.8	27.6	8.8	0
4001	2145	95443	6.4	440.3	11.9	.3	439.9	5.7	11.9	3
4002	2145	95443	6.4	652.1	9.5	.3	651.7	5.7	9.5	4
4003	2145	95443	6.4	332.3	14.3	.3	331.9	5.7	14.3	3
4004	2145	95443	6.4	544.8	10.4	.3	544.4	5.7	10.4	4

Figure 6-10 Workload in IASP

IBM i recovery domain shows the production IBM i node as Primary, and DR IBM i node as Backup, as shown in Figure 6-11.

```
                         Work with Recovery Domain

 Cluster resource group . . . . . . . . . . :   PWRHA_CRG
 Consistent information in cluster  . . . :   Yes

 Type options, press Enter.
   1=Add node    4=Remove node    5=Display more details    20=Dump trace

                                             Current        Preferred      Site
 Opt     Node         Status            Node Role      Node Role      Name
         ________
 __      DISREC       Active            *BACKUP   1    *BACKUP   1    SITE1
 __      PROD         Active            *PRIMARY       *PRIMARY       SITE1
```

Figure 6-11 PowerHA for i Recovery domain before switch-over

Before the planned outage, you can stop the IBM i workload.

To switch over to the DR site, issue the following command in the IBM i DR partition:

```
CHGCRGPRI CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG)
```

During switchover, the following actions are automatically performed by PowerHA for IBM i:

1. IASP in production IBM i is varied-off, as can be seen in Figure 6-12, which shows the changing status of the IASP.

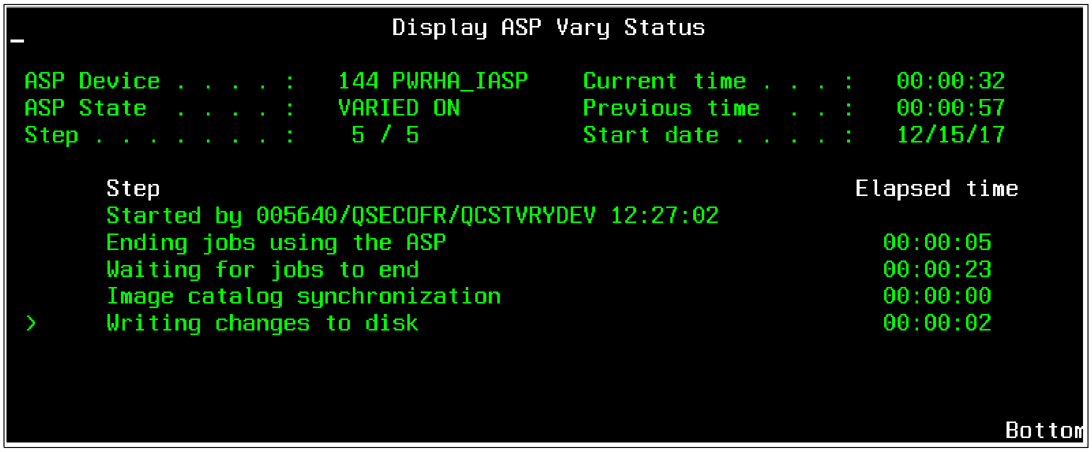


Figure 6-12 Vary off production IASP

2. The connection of IASP LUNs changes so that they are now connected through the Host to the DR site, as shown in Figure 6-13.

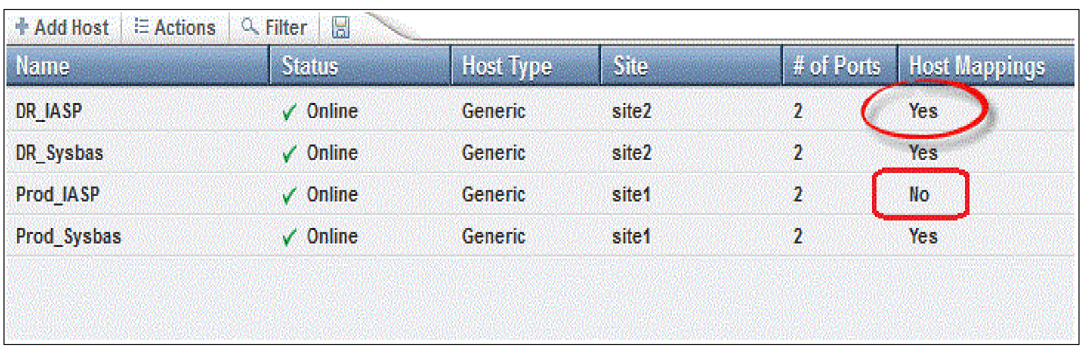


Figure 6-13 Host connections after switch-over

3. IASP in DR IBM i partition on DR site is varied on, as can be seen in Figure 6-14.

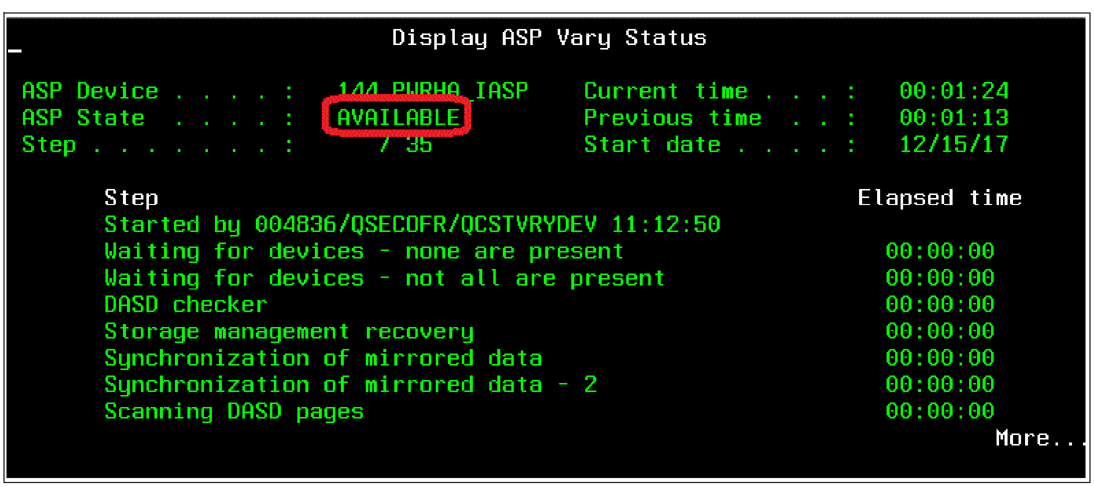


Figure 6-14 IASP is varied-on at the DR site

4. The HyperSwap Metro Mirror relation of IASP is reversed, as can be seen in Figure 6-15.

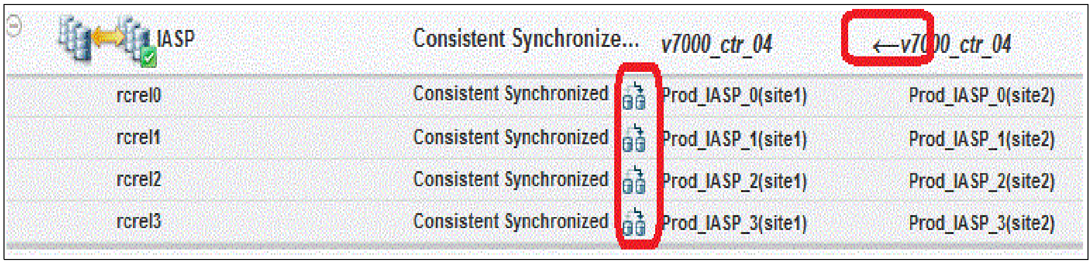


Figure 6-15 HyperSwap Metro Mirror direction after switch-over

The environment at the DR site is now ready to restart or continue the workload.

When the planned outage is finished, we switch IASP back to the production site by using the **CHGCRGPRI CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG)** command in partition PROD. During switchback, the same actions are performed as during switchover. They are done in the direction of DR site to production site. The direction of HyperSwap Metro Mirror is changed to site 1 → site 2.

At PowerHA for IBM i, switch over the I/O rate on Storwize nodes switches from Storwize nodes 1 and 2 on site 1, to nodes 3 and 4 at site 2. At PowerHA for IBM i, switch back the I/O rate to Storwize nodes 1 and 2 on site 1. This can be seen on the I/O rate in Storwize captured by IBM Spectrum Control, shown in Figure 6-16.

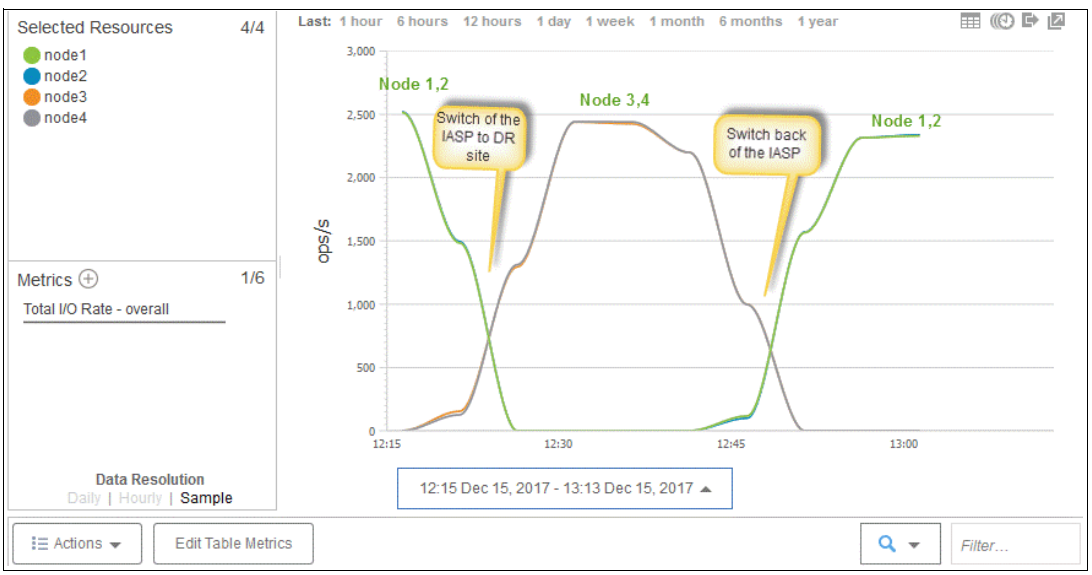
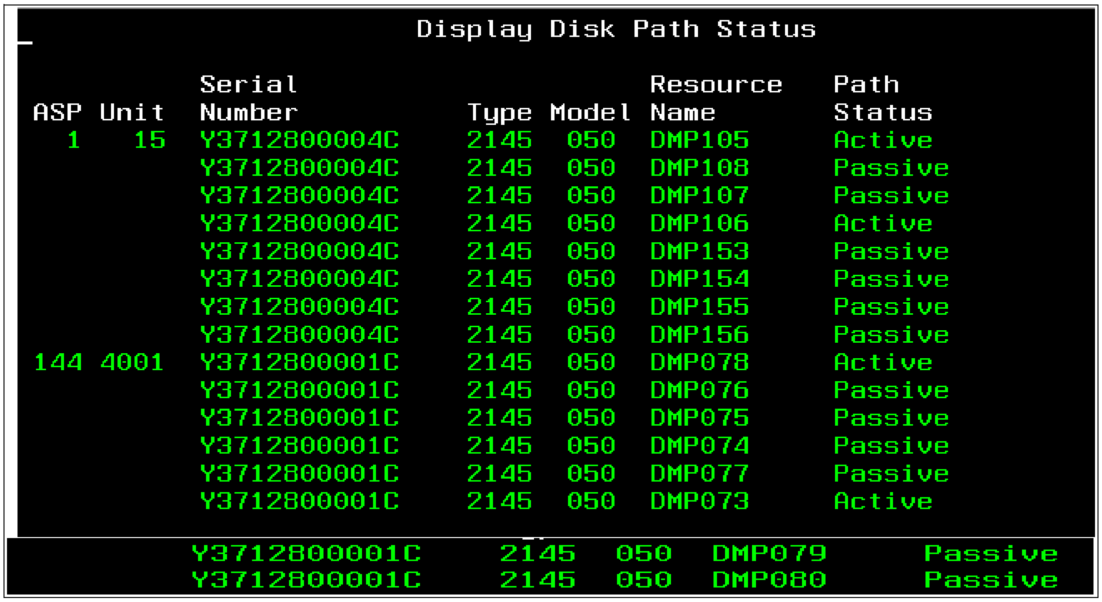


Figure 6-16 I/O rate in Storwize during planned outage

6.3 Storwize on production site fails

Before the outage, the workload in the production IBM i partition is running, each LUN of Sysbas or IASP has two active and six passive paths, as can be seen in Figure 6-17.



```
                         Display Disk Path Status

           Serial                         Resource      Path
ASP  Unit  Number           Type  Model  Name          Status
  1    15  Y3712800004C     2145   050   DMP105        Active
           Y3712800004C     2145   050   DMP108        Passive
           Y3712800004C     2145   050   DMP107        Passive
           Y3712800004C     2145   050   DMP106        Active
           Y3712800004C     2145   050   DMP153        Passive
           Y3712800004C     2145   050   DMP154        Passive
           Y3712800004C     2145   050   DMP155        Passive
           Y3712800004C     2145   050   DMP156        Passive
144  4001  Y3712800001C     2145   050   DMP078        Active
           Y3712800001C     2145   050   DMP076        Passive
           Y3712800001C     2145   050   DMP075        Passive
           Y3712800001C     2145   050   DMP074        Passive
           Y3712800001C     2145   050   DMP077        Passive
           Y3712800001C     2145   050   DMP073        Active
           Y3712800001C     2145   050   DMP079        Passive
           Y3712800001C     2145   050   DMP080        Passive
```

Figure 6-17 Paths to IBM i LUNs before failure

We simulate the outage of the two Storwize nodes on the production site by entering Service state on the nodes. IBM i system and workload keep running, and there are warning messages about failed paths, as can be seen in Figure 6-18.

```
ype reply (if required), press Enter.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
 Warning - An external storage subsystem disk unit connection has failed.
```

Figure 6-18 Warnings about failed paths

During the outage, the LUNs in the production IBM i partition have two active, two passive, and four failed paths. The failed paths belong to the failed Storwize nodes. The status of paths can be seen in Figure 6-19 on page 61.

ASP	Unit	Serial Number	Type	Model	Resource Name	Path Status
1	15	Y3712800004C	2145	050	DMP155	Active
		Y3712800004C	2145	050	DMP108	Failed
		Y3712800004C	2145	050	DMP105	Failed
		Y3712800004C	2145	050	DMP107	Failed
		Y3712800004C	2145	050	DMP106	Failed
		Y3712800004C	2145	050	DMP153	Passive
		Y3712800004C	2145	050	DMP154	Passive
		Y3712800004C	2145	050	DMP156	Active
144	4001	Y3712800001C	2145	050	DMP076	Active
		Y3712800001C	2145	050	DMP075	Passive
		Y3712800001C	2145	050	DMP074	Passive
		Y3712800001C	2145	050	DMP077	Active
		Y3712800001C	2145	050	DMP078	Failed
		Y3712800001C	2145	050	DMP073	Failed
		Y3712800001C	2145	050	DMP079	Failed
		Y3712800001C	2145	050	DMP080	Failed

Figure 6-19 Paths to IBM i LUNs after failure

The LUNs in the disaster recovery IBM i also have two active, two passive, and four failed paths, as can be seen in Figure 6-20.

ASP	Unit	Serial Number	Type	Model	Resource Name	Path Status
1	1	Y37128000070	2145	050	DMP157	Active
		Y37128000070	2145	050	DMP160	Passive
		Y37128000070	2145	050	DMP158	Passive
		Y37128000070	2145	050	DMP159	Active
		Y37128000070	2145	050	DMP189	Failed
		Y37128000070	2145	050	DMP190	Failed
		Y37128000070	2145	050	DMP191	Failed
		Y37128000070	2145	050	DMP192	Failed

Figure 6-20 Paths to the LUNs in DR partition after failure

We end the outage by coming out of Service state of the production Storwize nodes. The failed paths in IBM i partitions become active or passive again accordingly.

The I/O rate during the test of the production Storwize failure, as captured by IBM Spectrum Control, is shown in Figure 6-21.

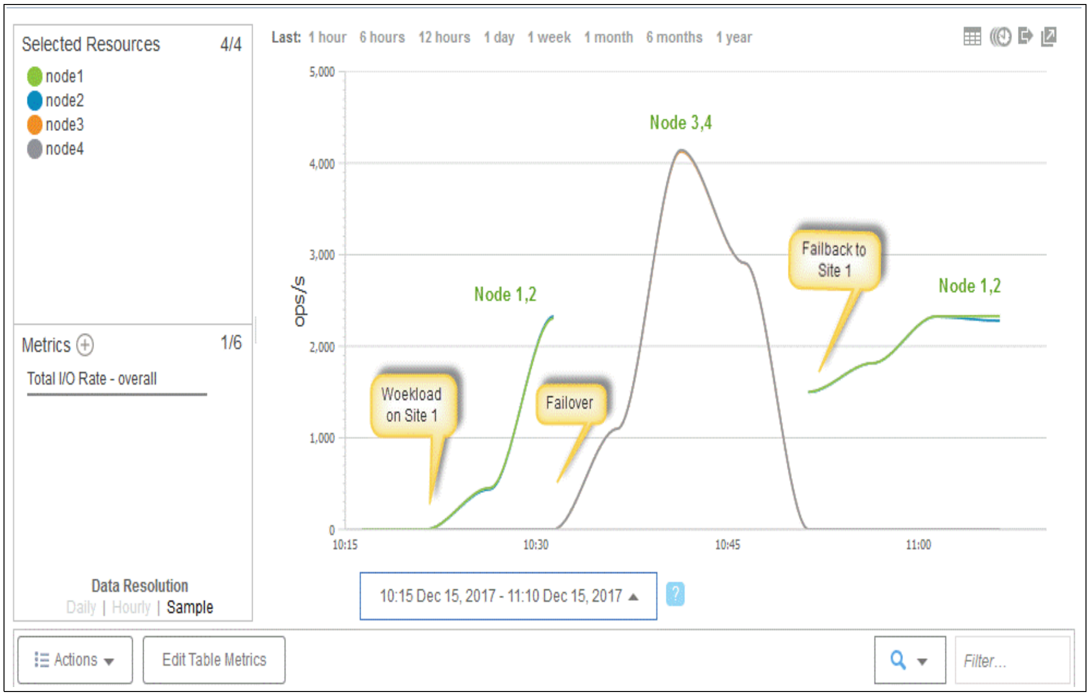


Figure 6-21 I/O rate during the failure of production Storwize

6.4 Production IBM i fails

Before the outage, the workload is running in the production IBM i IASP.

We simulate the IBM i failure by shutting down production IBM i, using the **PWRDWNSYS** command.

After the simulated failure of IBM i, PowerHA for IBM i automatically runs the following steps:

1. It sends a message to Disaster Recovery IBM i asking to confirm the further steps.

   Note that the message is sent only if PowerHA for IBM i environment is set up accordingly. If not, the actions are done without asking for confirmation.

2. Production IASP is varied off, providing that this action is possible in the failed IBM i.

3. IBM i Cluster Resource Group recovery domain changes the node roles: Disaster Recovery node becomes Primary, and Production node becomes Backup node. The changed roles can be seen in Figure 6-22.

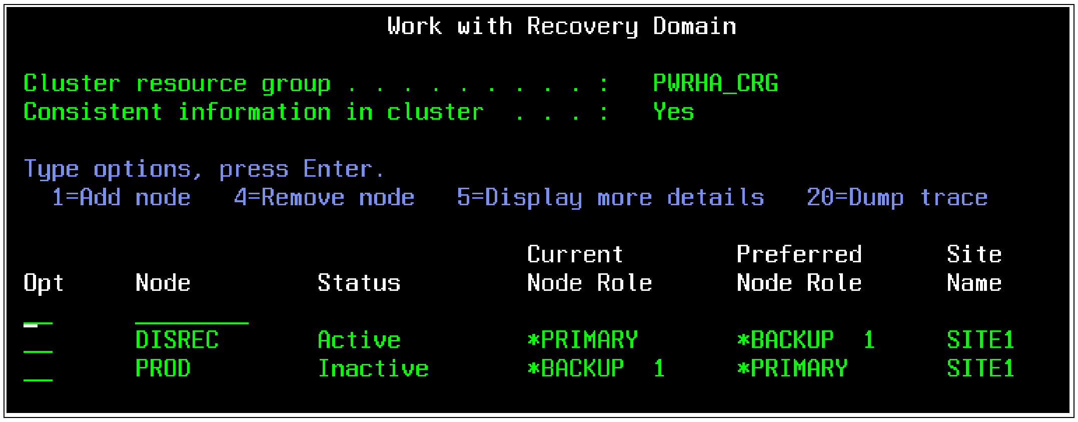


Figure 6-22 Changed node roles in the recovery domain

4. HyperSwap Metro Mirror changes direction, as shown in Figure 6-23.

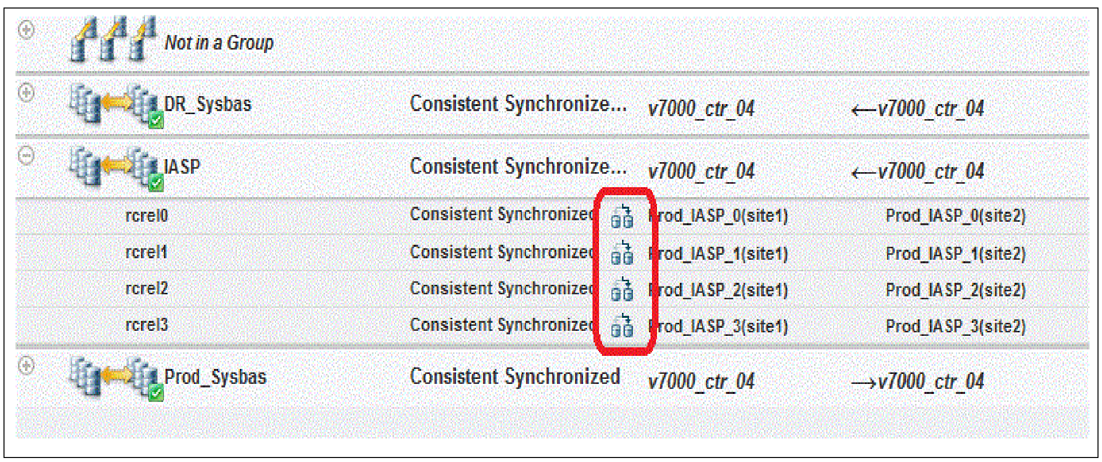


Figure 6-23 Reversed direction of Metro Mirror

5. IASP in the disaster recovery partition is varied on. The changing status of the Disaster Recovery IASP is shown in Figure 6-24.

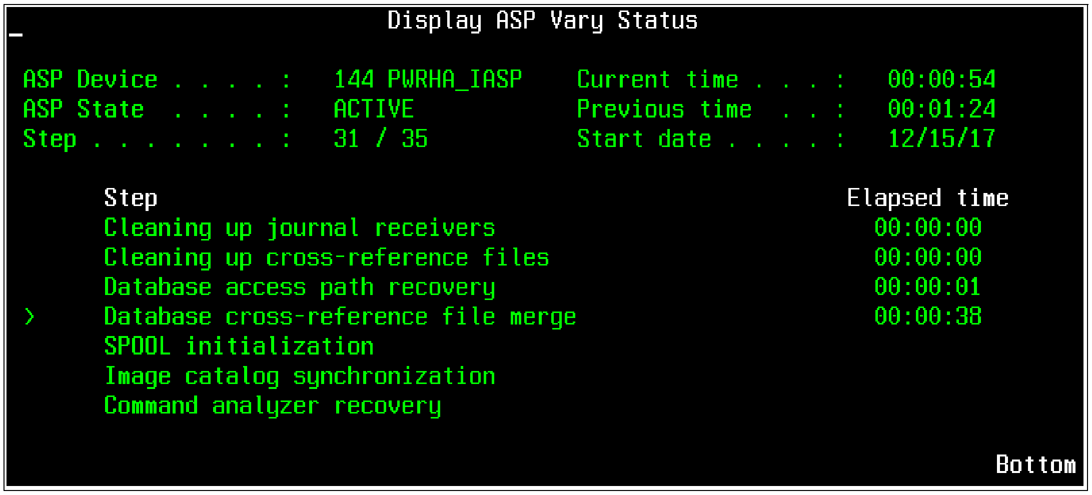


Figure 6-24 Vary-on of disaster recovery IASP

After the IASP in the DR partition varies on, we start the workload in the DR IASP.

The production node in IBM i cluster is inactive, as shown in Figure 6-25.

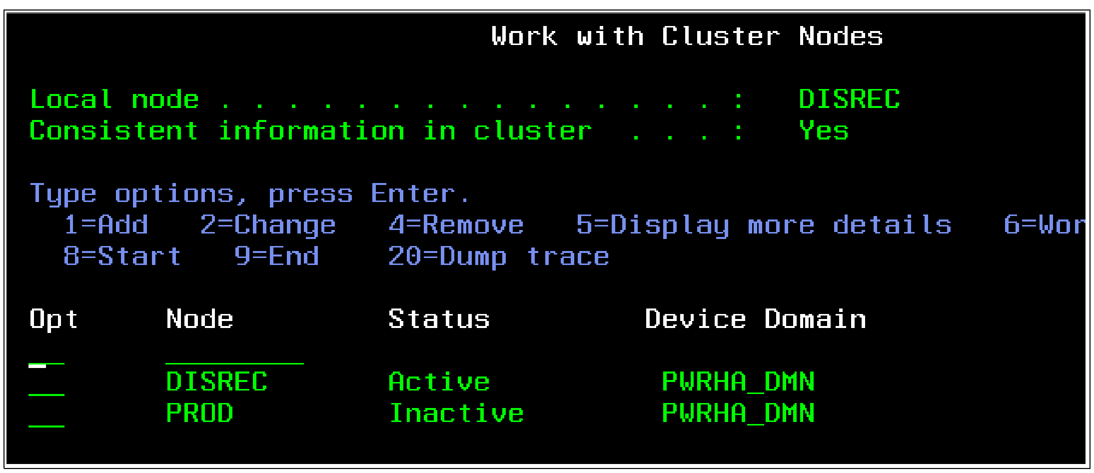


Figure 6-25 IBM i cluster nodes during outage

The I/O rate is now running on nodes 3 and 4 of the Storwize HyperSwap cluster.

To end the outage, we start the production IBM i, and we start the production node in the cluster. We switch back from DR IBM i to Production IBM i, by using the PowerHA command **`CHGCRGPRI CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG)`** *in system PROD*.

When the IASP becomes available at the production site, we restart the workload in the production IASP.

After switchback, the Metro Mirror relations in the IASP HyperSwap LUNs change direction from production → DR, as shown in Figure 6-26.

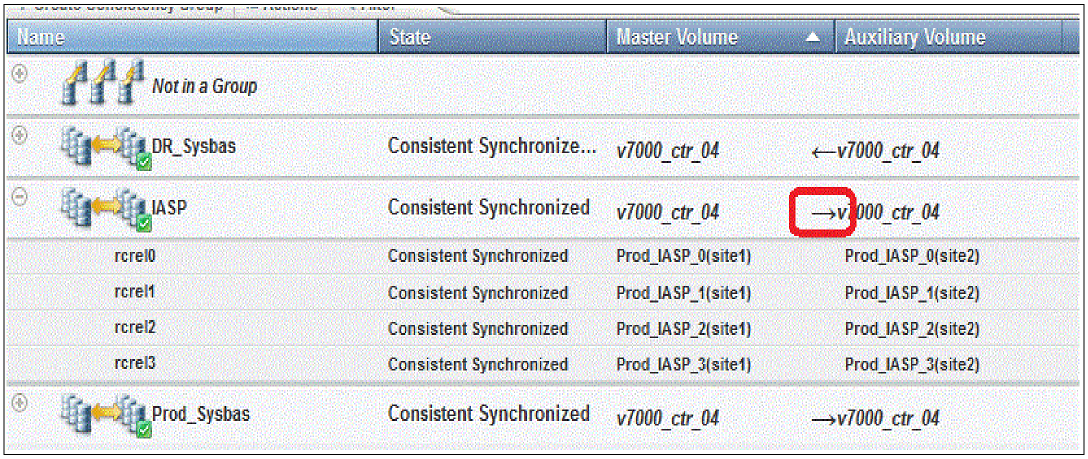


Figure 6-26 Metro Mirror direction is changed to initial

Figure 6-27 shows the I/O rate during the switch-over and switch-back, as captured by IBM Spectrum Control.

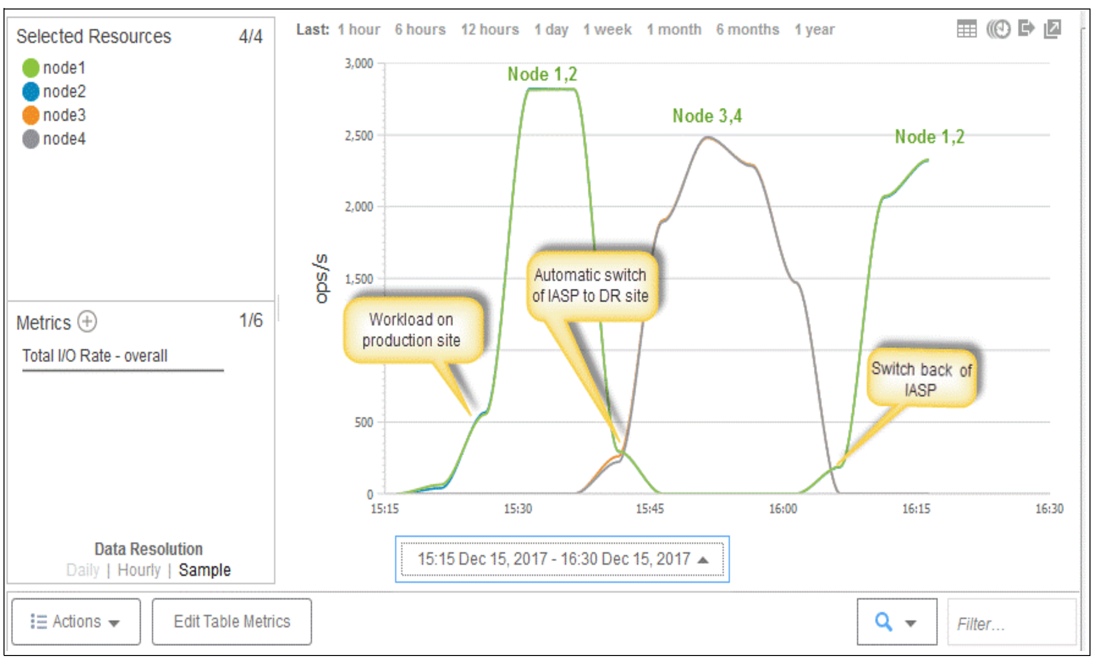


Figure 6-27 IBM Spectrum Control graph of the I/O rate

6.5 Disaster at production site

The workload is running in production IASP.

To simulate disaster at the production site (Site 1), we shut down the production IBM i and we enter Service state on both Storwize V7000 nodes at Site 1.

The following actions are automatically executed by PowerHA for i:

1. It sends a message to Disaster Recovery IBM i asking to confirm the further steps.

   Note that the message is sent only if the PowerHA for IBM i environment is set up accordingly. If not, the actions are done without asking for confirmation.
2. Production IASP is varied off, if disaster on the production site permits vary-off of the IASP.
3. IBM i Cluster Resource Group recovery domain changes the node roles: disaster recovery node becomes Primary and Production node becomes the Backup node.
4. HyperSwap Metro Mirror changes direction.
5. IASP in disaster recovery partition is varied-on.

Next, complete the following steps:

1. After IASP at the DR site becomes available, we restart the workload in DR partition at Site 2.
2. To end the disaster, we exit the Service state of Storwize V7000 nodes at Site 1, and we start production IBM i.

   After the disaster is over, the HyperSwap Metro Mirror is as shown in Figure 6-28.

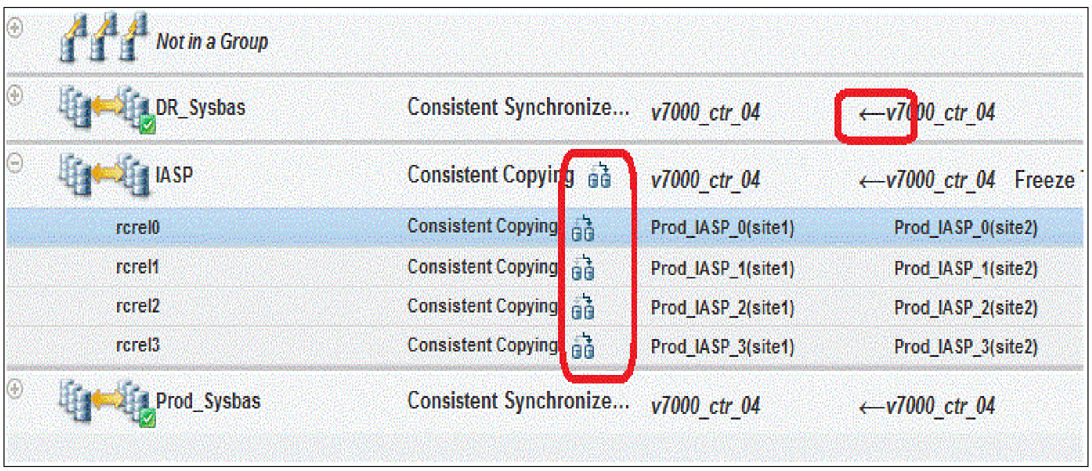


Figure 6-28 HyperSwap Metro Mirror changed direction

3. We start the cluster node on the production partition `PROD`. IBM i cluster recovery domain shows the DR node as primary and the production node as backup, as shown in Figure 6-29.

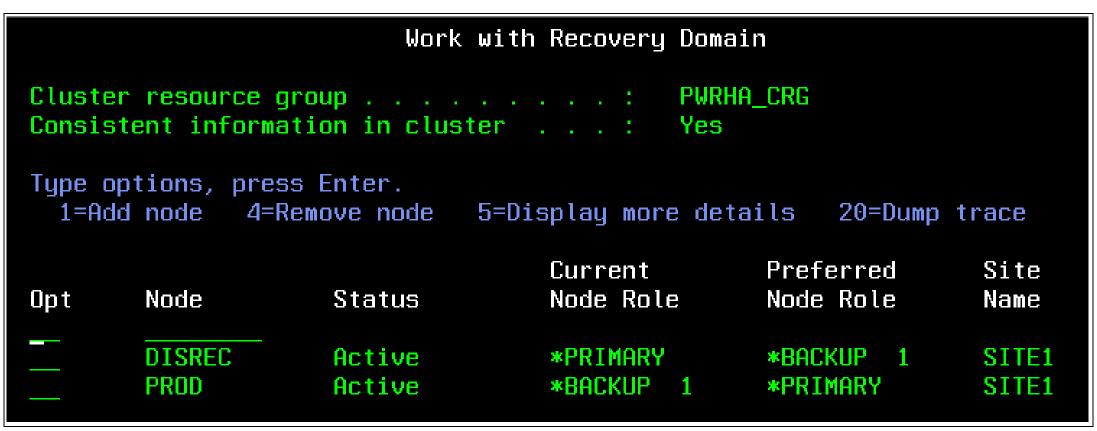


Figure 6-29 Node role after production site is back

4. We switch-back to production partition by using the command **CHGCRGPRI CLUSTER(PWRHA_CLU) CRG(PWRHA_CRG)** *in partition PROD*.
5. We restart the workload in the production partition.

During the outage of site 1, the I/O rate transfers to nodes 3 and 4 at site 2. After the outage, it transfers back to nodes 1 and 2 at site 1. This is shown in Figure 6-30 as captured by IBM Spectrum Control.

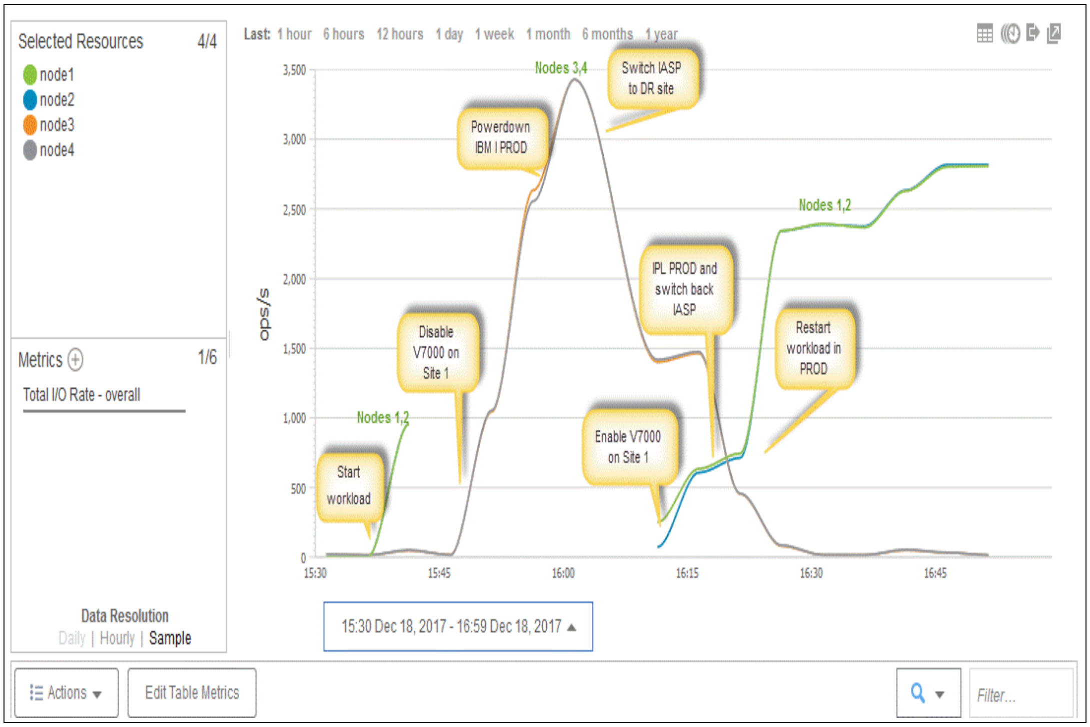


Figure 6-30 IO rate during the disaster as captured by Spectrum Control

This concludes our testing.

IBM®

REDP-5490-00

ISBN 0738456888

Printed in U.S.A.