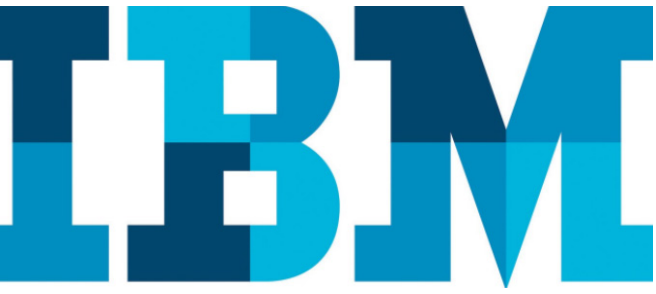
June 2020

IBM HyperSwap: An automated disaster recovery solution

Disaster recovery at different geographical sites

Table of contents



Overview

Challenge
How to configure an automated disaster recovery solution between existing IBM Storwize V7000 control enclosures at two different geographical locations

Solution
This paper talks about the IBM HyperSwap solution as a complete automated disaster recovery solution.

It also provides a checklist to validate the configuration for failures.

Introduction

This technical paper is based on a client engagement and is developed to assist IBM® Business Partners, field sales representatives, technical specialists, and IBM clients to understand when can IBM HyperSwap® be deployed on IBM Spectrum Virtualize systems and the steps to validate the configuration for failure cases. It also includes the best practices involved for the specific client engagement and can be applied in a similar environment.

Key issues in an existing environment

- Hardware refresh for IBM XIV Storage System
- Hardware refresh for IBM Storwize V5000
- High RPO

Problem statement

The client is an insurance company that was facing issues due to lack of storage space and in need to refresh the hardware. Along with this, the company had a disaster recovery site that was managed by another application which had high recovery point objective (RPO) and recovery time objective (RTO). The key requirement of having a disaster recovery solution was to reduce RPO and RTO.

The following are the issues observed with the existing setup at the customer environment:

- Customer had two different and independent IBM storage controllers such as IBM XIV® Storage System and IBM Storwize® V5000 serving Oracle and VMware workloads respectively. Both controllers were due for hardware refresh. So, they needed to replace these existing controllers with new ones.
- Additionally, they had a common disaster recovery site configured on IBM Storwize V7000 storage control enclosure, serving both workloads of IBM XIV and IBM Storwize V5000 independently. Oracle workload was replicated using their own application Oracle data guard. And VMware was replicated using the IP replication feature configured between Storwize V5000 and Storwize V7000 controllers. The observed RPO was approximately about 1 hour and the time taken for Oracle failover was approximately 3 hours. The customer was very keen on getting this duration of these tasks reduced considerably.

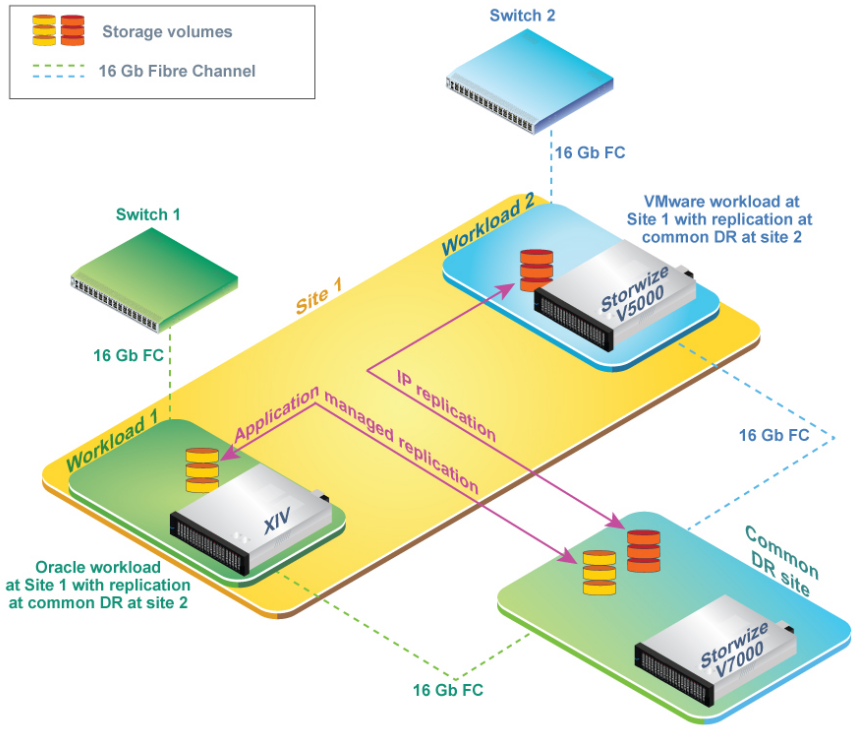


Figure 1: Existing system

Things to be considered before providing the solution:

- **Hardware refresh**: IBM Storwize V7000 is a good replacement for IBM Storwize V5000 and IBM XIV Storage System controllers as part of hardware refresh.
- **Zero RPO and zero RTO**: Application managed RPO RTO is quite high. The IBM HyperSwap® solution offers disaster recovery with minimal RPO and RTO. It enables automatic failover of storage access across domains or sites.
- **Reusing the existing hardware** along with hardware refresh. Customer already had one Storwize V7000 system as a common DR site. Replacing the old Storwize V5000 and XIV systems with the Storwize V7000 system as part of hardware refresh could be a possible solution which in-turn will enable the environment for configuring IBM HyperSwap solution for disaster recovery.
- Before finalizing the HyperSwap environment, there are few scenarios that are to be considered:
    a. **Controller failure**: Here one of the controllers (node canister) of the system fails. In such situations, its partner node or canister must take control of all the I/O operations and there shouldn't be any drop or increase in latencies.
    b. **Quorum failure**: Here the quorum disk becomes unavailable to the system for a certain period. Quorums can be of two types: FC quorum and IP quorum. In our case, it is the IP quorum application that provides an automatic tie break during link failure. Generally, it is recommended to have three IP quorums (that is, three different systems to host IP quorum for the same system) to enable better fault tolerance. Also, systems hosting IP quorum should be independent from the storage provided by the Storwize V7000 cluster they work for.

Proposed solution

Best configuration for the IBM HyperSwap solution is to have even number of I/O groups on each site. But this solution is an exception as it is based on the existing requirement and existing hardware.

After considering all above points IBM proposed its final solution:

- Considering the existing customer disaster recovery environment, IBM HyperSwap was offered as it inherently replicates data between two sites and provides continuous availability of data during outages. It provides features such as an automatic failover of storage access.
- The proposed solution included three I/O groups of IBM Storwize V7000 controllers, out of which two were introduced as part of hardware refresh (of IBM XIV and IBM Storwize V5000 controllers) on one site and one existing I/O group at another site of the complete disaster recovery solution.
- IP quorum can be used as the third site to act as a tie breaker between two HyperSwap sites.
- The following features are also included in the solution:
    - **Encryption**: IBM offers USB encryption and key server encryption with license to encrypt data before it is stored on

the drives or on cloud.

- **Data reduction pools**: For increased storage efficiency, increased performance, and reduced storage cost (especially for flash storage), IBM introduced HyperSwap feature. It does not require a new or an additional license.
- **Global Mirror with change volume replication for selected workloads**: As it periodically uses IBM FlashCopy® to establish a new set of data to replicate, establishing that set of data requires a finite length of time. This is used for replicating data between two sites.

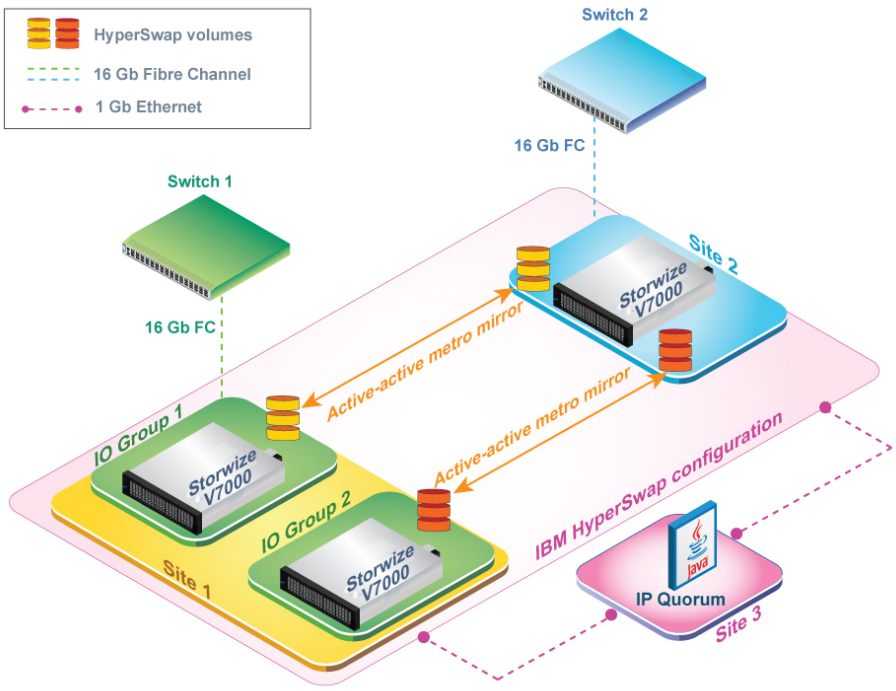


Figure 2. Proposed solution

As shown in Figure 2, we have two IBM Storwize V7000 control enclosures (contributing to two different I/O groups) at the failure domain site 1 and a single Storwize V7000 controller enclosure (a single I/O group) at the failure domain site 2 forming an active-active Metro Mirror relationship between sites. These sites reside on two different geographical locations, and an IP quorum is used for tie-breaking at a failure domain site 3.

Validating the system for HyperSwap failures

The Oracle Vdbench tool is used with the parameters indicated in the following table to simulate workload on Storwize V7000 storage volumes mapped to a system hosting the Vdbench application.

Vdbench version	Workload	Transfer size (in KB)	Interval (in sec)	Randomness (in %)
5.04.05	70% read 30% write	8	1	100

Methods to validate system failure in HyperSwap configuration:

- **Storwize V7000 control enclosure failure**: This method can be simulated by switching off the Storwize V7000 control enclosure in any of the failure domains. For testing purposes, the Failure Domain 1 control enclosures were powered off. It has been observed that Vdbench sustained the failure with a pause of 30 seconds in I/O processing during I/O path recovery by the host multipath drivers. I/O operations were queued during the pause of 30 seconds induced due to power failure and were submitted subsequently after path recovery by the multipathing drivers.

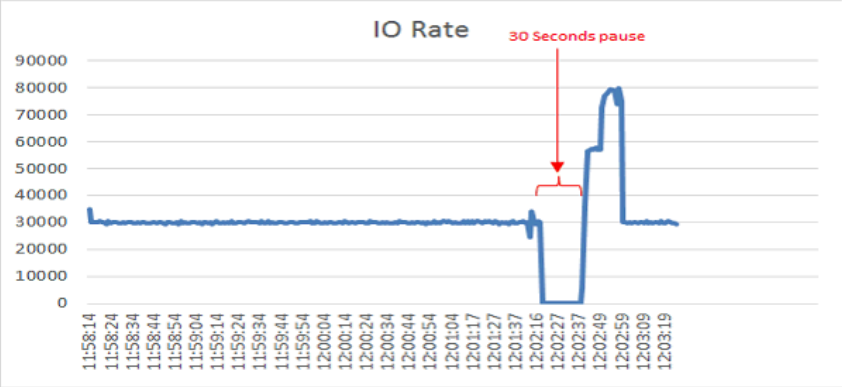


Figure 3. I/O rate output from Vdbench

Read response time returned to normal after path recovery and write response time drops due to suspension of mirroring across failure domains (as shown in figure 4).

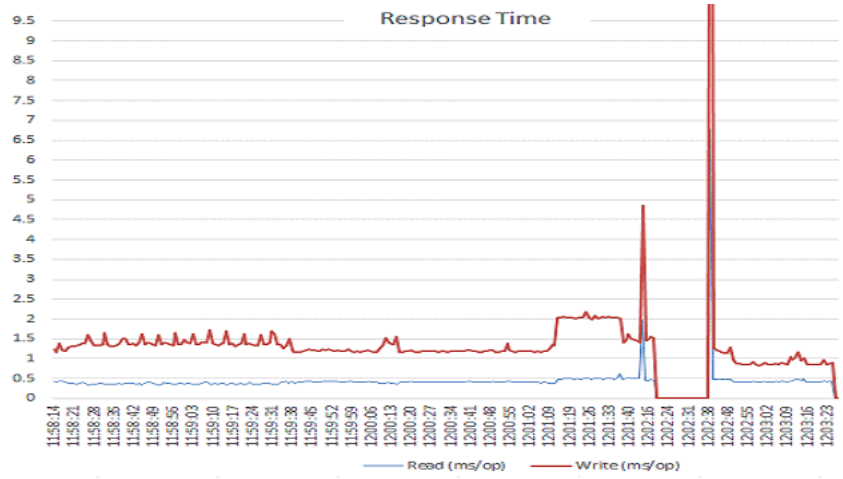


Figure 4. Response time output from Vdbench

Result: Seamless failover to Failure Domain 2 with no interruption to the I/O workload.

- **Quorum failure**: This can be simulated by terminating the IP quorum applications in Failure Domain 3. There was no drop in the I/O and Vdbench continued running without any pause.

Result: No impact was observed when quorum (Domain 3) failure was offline.

Generic best practices for IBM HyperSwap

The following best practices are recommended for HyperSwap configurations:

- Using redundant IP quorums as the third domain provides continuous protection during split-brain situations.
- Users can achieve faster failover capability by using IBM Spectrum Virtualize N_Port ID Virtualization (NPIV).
- Users can achieve traffic isolation by designating ports for each function such as internode communication, replication, and host access. Designating ports for internode communication can be evaluated by assessing the write throughput of the workload in the IBM HyperSwap configuration. It is also important to have the required inter-site link bandwidth to meet the peak write throughput.
- Never mix HyperSwap and regular volumes on the same storage system unless you have any alternate high-availability solution in place at the host or application level for the regular volumes.
- Set the buffer-to-buffer credit to 32 on storage area network (SAN) switches where Storwize V7000 ports are connected.
- There are few recommended settings for host attach operating systems (OS) for multipath drivers:

OS multipath settings/ Host OS	IBM AIX®	Microsoft® Windows®	VMware®	Linux®	XenServer
Multipath drivers	Default AIXPCM	Subsystem Device Driver Device Specific Module (SDDDSM)	Native Multipath Plug-in driver (NMP)	Device Mapper Multi pathing (DMM)	Citrix Device Mapper Multi path (DMM)
Host parameters	Algorithm: Shortest_queue Reservation policy: no_reserve	Default values	Path selection policy: VMW_PSP_RR	Default values	Default values

Summary

The IBM HyperSwap proved to be the best solution in the customer as it provides automated replication and recovery for volumes. It provides continuous availability of data during outages. Additionally, the solution came with the hardware refresh which is cost effective as well.

Get more information

To learn more about IBM Spectrum Virtualize/ IBM HyperSwap contact your IBM representative or IBM Business Partner, or visit the following websites:

- IBM FlashSystem Knowledge Center: https://www.ibm.com/support/knowledgecenter/STSLR9_8.2.1/com.ibm.fs9100_821.doc/fs9100_ichome.html
- IBM Spectrum Virtualize Knowledge Center: https://www.ibm.com/support/knowledgecenter/STVLF4_8.1.3/spectrum.virtualize.813.doc/svirt_ichome.html
- IBM SVC Knowledge Center: (HyperSwap function): https://www.ibm.com/support/knowledgecenter/STSLR9_8.2.1/com.ibm.fs9100_821.doc/svc_hyperswapovr.html
- IBM HyperSwap red paper: http://www.redbooks.ibm.com/redpapers/pdfs/redp5597.pdf

About the author

Aakanksha Mathur leads test effort for IBM Spectrum Virtualize product chain at IBM System Labs, India. She has been working in storage domain for more than 12 years including 9 years with IBM System Labs. She can be reached at aamathur@in.ibm.com

Acknowledgments

The author would like to thank Dhiraj Verma and Varun Kumar for their support on successfully completion of client engagement and would like to thank Akshat Mithal for his review and valuable feedback.

IBM Systems
3039 Cornwallis Road
RTP, NC 27709

Produced in the United States of America