Storage Center Synchronous Replication and Live Volume

Solutions Guide

Dell Compellent Technical Solutions
May 2014

A Dell Technical White Paper

Revisions

Date	Description
May 2014	Merged Sync Rep & Live Volume documents. Updated for EM2014R2 & SC6.5.

Table of contents

Executive summary

Preventing the loss of data or transactions requires a method of continuous data protection. In the scope of a disaster, data must safely reside at an alternate site. A variety of data mobility methods, including asynchronous replication, can accomplish the task of providing offsite replicas. Synchronous replication sets itself apart from the other methods by guaranteeing transactional consistency between the production site and the recovery site.

While remote replicas have traditionally provided a data protection strategy for disaster recovery, the disaster itself and the execution of a disaster recovery (DR) plan involves a period of down time for organizations. Replicas along with storage virtualization can provide other types of data mobility that fit a broader range of proactive high availability use cases without an outage.

This guide focuses on two main data protection and mobility features available in Dell Compellent Storage Center: synchronous replication and Live Volume. In this paper, each feature is discussed and sample use cases are highlighted where these technologies fit independently or together.

1 Introduction to synchronous replication

Dell Compellent historically supports both asynchronous and synchronous replication. This document focuses on synchronous replication and available features in Storage Center 6.5 (and earlier versions) as well as some use cases.

By definition, synchronous replication is a method of guaranteeing that data is written and committed to both the replication source and destination volumes in real time. The data is essentially written to both locations simultaneously. In the event that data cannot be written to either location, it will not occur at all and a failure will be issued to the storage host and application where the write request originated. Although the transaction fails, the benefit it provides is a guarantee of precise data consistency between both locations resulting in zero data loss in a recovery scenario.

Compellent advises customers to understand the types of replication available, their applications, and their business processes before designing and implementing a data protection and availability strategy.

1.1 New features in Storage Center 6.5 synchronous replication

Mode migration: Once created, a replication may be migrated to an alternate type or mode of operation without rebuilding the replication from scratch.

Live Volume support: Live Volumes may leverage any available type of replication offered in Storage Center including both modes of synchronous (high consistency or high availability) and asynchronous.

Live Volume Managed Replication: Live Volume allows an additional synchronous or asynchronous replication to a third site that can be DR activated using Enterprise Manager.

Preserve Live Volume: In the event an unplanned outage occurs impacting availability of a primary Live Volume, the secondary Live Volume can be promoted to the primary Live Volume role using Enterprise Manager.

1.2 Synchronous replication requirements

Replicating volumes between Storage Center systems requires a combination of software, licensing, storage, and fabric infrastructure. The following sections itemize each requirement.

1.2.1 Compellent Enterprise Manager

Enterprise Manager 6.3 or newer is required to create synchronous volume replication in high consistency or high availability modes. Enterprise Manager 2014 R2 or newer is required to leverage the new synchronous replication and Live Volume feature enhancements.

1.2.2 Compellent Storage Center

Storage Center 6.3 or newer is required to support synchronous volume replication in high consistency or high availability modes. Storage Center 6.5 or newer is required to leverage the new synchronous

replication and Live Volume feature enhancements such as Synchronous Live Volume. Asynchronous replication operates with older versions of Storage Center.

1.2.3 Licensing

Replication licensing, which includes Sync Remote Instant Replay and Async Remote Instant Replay, is required for each Storage Center participating in volume replication. Additionally, a Live Volume license for each Storage Center is required for Live Volume and Live Volume managed replication features.

1.2.4 Supported replication transport

Storage Centers support array-based replication using either Fibre Channel or iSCSI connectivity. A dedicated network is not required but a method of isolation for performance and/or security should be provided.

2 Data replication primer

Data replication is one of many methodologies that exist to provide data protection and integrity for businesses. The practice of replication evolved out of a necessity to address a number of drivers such as substantial data growth, shrinking backup windows, more resilient and efficient disaster recovery solutions, high availability, mobility, globalization, cloud, and regulatory requirements. The common requirement is to maintain multiple copies of data. Traditional backup methods satisfied early data protection requirements, but this feasibility diminished as data sets grew larger combined with other availability constraints. Vanishing backup windows, ecommerce, and exponential growth of transactions brought about the need for continuous data protection (CDP). Replicas are typically used to provide disaster recovery or high availability for applications and data, to minimize or eliminate loss of transactions, to provide application and data locality, or to provide a disposable data set that can be internally developed or tested. At a higher level, data protection translates to guarding the reputation of a company by protecting end user customer data.

2.1 Replication methods

There are a number of replication approaches, but two methods stand out as highly recognized options in the industry: asynchronous and synchronous. Compellent Storage Center supports a flexible variety of replication methods that fall in the category of asynchronous or synchronous.

2.1.1 Synchronous

Synchronous replication guarantees data consistency (zero data loss) between replication source and destination. This is achieved by ensuring write I/O commitments at the replication source and destination before a successful write acknowledgement is sent back to the storage host and the requesting application. If the write I/O cannot be committed at the source or destination, the write will not be committed at either location to ensure consistency. Furthermore, a write failure is sent back to the storage host and its application. Application error handling will then determine the next appropriate step for the pending transaction. By itself, synchronous replication provides CDP. Coupled with hardware redundancy, application clustering, and failover resiliency, continuous availability for applications and data can be achieved.

Because of the method used in synchronous replication to ensure data consistency, any issues impacting the source or destination storage, or the replication link in between, will adversely impact applications in terms of latency (slowness) and availability. This applies to Live Volumes built on top of synchronous replications as well. For this reason, appropriate performance sizing is paramount for the source and destination storage, as well as the replication bandwidth and any other upstream infrastructure that the storage is dependent on.

Figure 1 demonstrates the write I/O pattern sequence with synchronous replication:

1. The application or server sends a write request to the source volume.
2. The write I/O is mirrored to the destination volume.
3. The mirrored write I/O is committed to the destination volume.

4. The write commit at the destination is acknowledged back to the source.
5. The write I/O is committed to the source volume.
6. Finally, the write acknowledgement is sent to the application or server.

The process is repeated for each write I/O requested by the application or server.

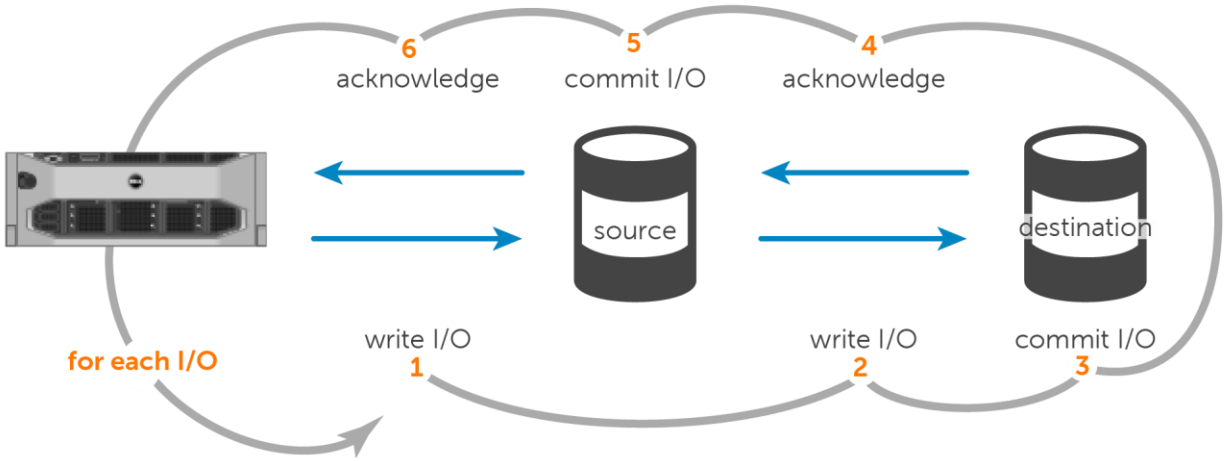

Figure 1 Synchronous replication write I/O sequence

2.1.2 Asynchronous

Asynchronous replication accomplishes the same data protection goal in that data is replicated from source storage to destination storage. However, the manner and frequency that the data is replicated differs from synchronous replication. Instead of committing a write at both replication source and destination simultaneously, the write is committed only at the source and an acknowledgement is then sent to the storage host and application. The accumulation of committed writes at the source volume are replicated to the destination volume in one batch at scheduled intervals and committed to the destination volume.

Aside from replicating the Active Replay (Semi-synchronous discussed in Section 2.1.3), Compellent Storage Center asynchronous replication is tied to the source volume Replay schedule. When a Replay is created on the source volume, and that volume is configured for asynchronous replication, the new Replay is replicated to the destination volume. Replays on a volume may be created automatically according to a schedule or manually created from a variety of integration tools. Regardless, all Replays occur on a per volume basis. As a result, volumes may adhere to their own independent replication schedule, or they may share a replication schedule with other volumes leveraging the same Replay profile. This type of replication is also referred to as Point-in-time replication which is a type of asynchronous replication that specifically leverages volume snapshots. Because asynchronously replicated transactions are not required to wait for write committals at the replica destination volume, the replication link and/or destination storage will not contribute to application or transaction latency at the source volume.

Figure 2 demonstrates the write I/O pattern sequence with respect to asynchronous replication.

1. The application or server sends a write request to the source volume.

2. The write I/O is committed to the source volume
3. Finally, the write acknowledgement is sent to the application or server.

The process is repeated for each write I/O requested by the application or server.

4. Periodically, a batch of write I/Os that have already been committed to the source volume are transferred to the destination volume,
5. Committed to the destination volume, and
6. A batch acknowledgement is sent to the source.

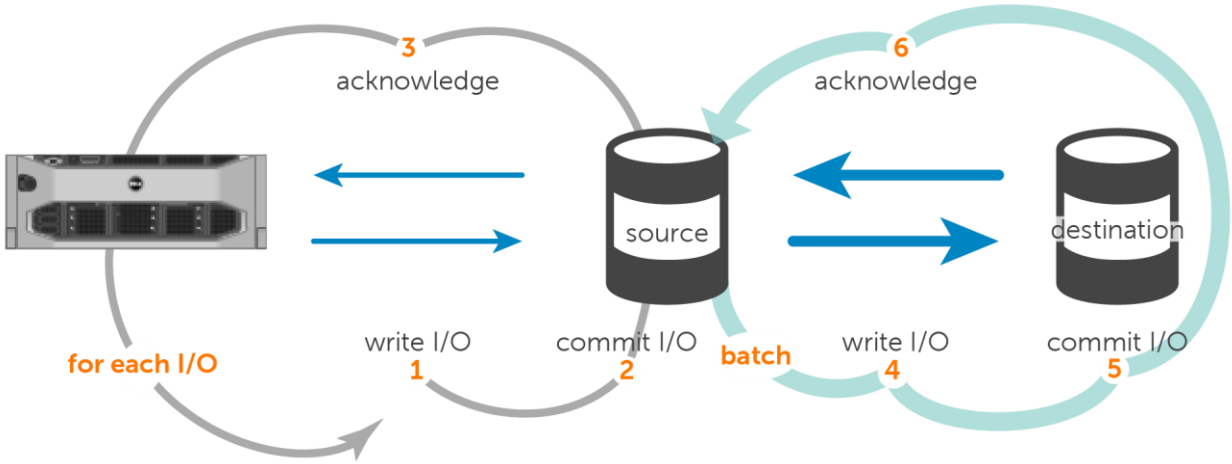

Figure 2 Asynchronous replication write I/O sequence

2.1.3 Semi-synchronous

In Compellent, semi-synchronous replication behaves like synchronous replication in that application transactions are immediately sent to the replication destination storage (assuming that the replication link and destination storage have the bandwidth to support the current rate of change). The difference is that the write I/O is committed at the source volume and an acknowledgement is sent to the storage host and application without a guarantee that the write I/O was committed at the destination storage. Semi-synchronous replication is configured in Enterprise Manager by creating Asynchronous replication between two volumes and checking the **Replicate Active Replay** box. A Replay is a Compellent Storage Center term that describes frozen data; it is similar to a snapshot. The Active Replay is a term that describes newly written or updated data that has not yet been frozen in a Replay. Semi-synchronous offers a synchronous-like Recovery Point Objective (RPO) without application latency, but the RPO and loss of data in a disaster recovery scenario cannot be guaranteed.

Figure 3 demonstrates the write I/O pattern sequence with semi-synchronous replication.

1. The application or server sends a write request to the source volume.
2. The write I/O is committed to the source volume.
3. The write acknowledgement is sent to the application or server.

The process is repeated for each write I/O requested by the application or server.

For each write I/O that completes that process, an independent and parallel process:

a. writes,
b. commits, and
c. acknowledges a mirror copy of that write to the destination volume.

The commits at the source and destination volumes are not guaranteed to be in lockstep with each other.

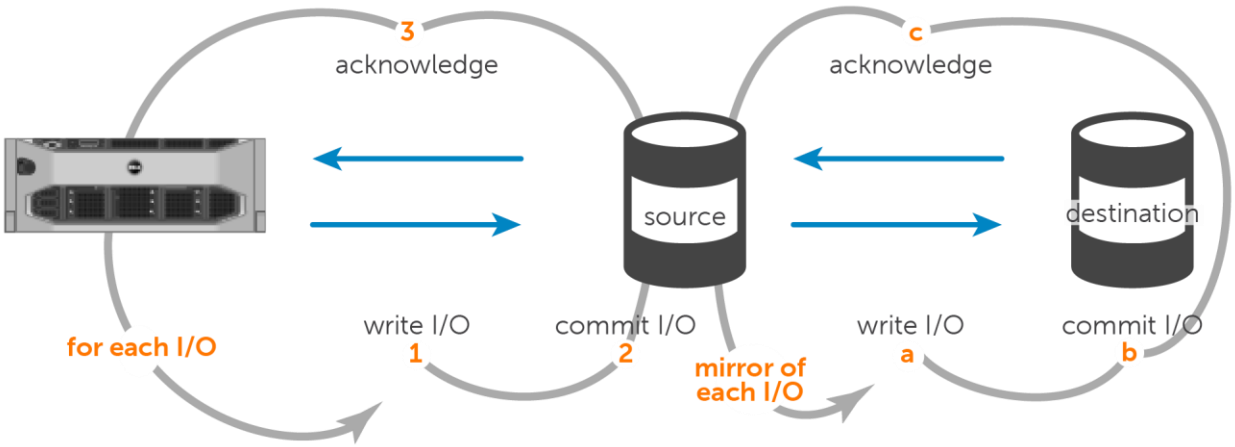

Figure 3 Semi-synchronous replication write I/O sequence

3 Synchronous replication features

Storage Center supports a wide variety of replication features. Each feature will be outlined in the following sections.

3.1 Modes of operation

A number of flexibility improvements have been added to enhance synchronous replication in Compellent Storage Center. Synchronous replication can be configured in one of two modes: High Consistency or High Availability.

3.1.1 Legacy

Synchronous replications created prior to Storage Center 6.3 are identified as Legacy after upgrading to Storage Center 6.3 and newer. Legacy synchronous replications cannot be created in Storage Center 6.3 or newer and do not possess newer synchronous replication features introduced in Storage Center 6.3 or newer. To upgrade a legacy synchronous replication to synchronous high consistency or synchronous high availability, the legacy synchronous replication must be deleted and recreated after both source and destination Storage Centers have firmware version 6.3 or newer installed. Deleting and recreating a synchronous replication will result in data inconsistency between the replication source and destination volumes until 100% of the initial and journaled replication is completed.

3.1.2 High consistency

Synchronous high consistency mode rigidly follows the storage industry specification of synchronous replication outlined earlier and shown in Figure 1. The mechanisms involved with this method of replication guarantee data consistency between replication source and destination volumes unless an administrator pauses the replication for maintenance or other reasons. Latency can impact applications at the source volume if the replication link or replication destination volume is unable to absorb the amount of data being replicated or the rate of change. Furthermore, if write transaction data cannot be committed to the destination volume, the write will not be committed on the source volume and in effect, a transaction involving a write fails. An accumulation of write failures will likely result in an application failure or outage when a tolerance threshold is crossed. For these reasons, application latency and high availability are important points to consider in a storage design proposing synchronous replication in high consistency mode.

3.1.3 High availability

Synchronous high availability mode bends the rules of synchronous replication by relaxing the requirements associated with high consistency mode. While the replication link and the replica destination storage is able to absorb the write throughput, high availability mode performs like high consistency mode (described in section 3.1.2 and illustrated in Figure 1). Data is consistently committed at both source and destination volumes and excess latency in the replication link or destination volume will be observed as application latency at the source volume.

The difference between high consistency and high availability mode is that data availability will not be sacrificed for data consistency. What this means is that if the replication link or the destination storage either becomes unavailable or exceeds a latency threshold, Storage Center will automatically remove the dual write committal requirement at the destination volume. This allows application write transactions at the source volume to continue with no downstream latency impacts instead of write I/O being halted or slowed, which is the case with high consistency mode and legacy synchronous replication. This relaxed state is referred to as being "out of date". If and when Storage Center enters the out of date state, inconsistent write I/O will be journaled at the source volume. When the destination volume becomes available within a tolerable latency threshold, journaled I/O at the source volume is flushed to the destination volume where it will be committed. During this process, incoming application writes continue to be written to the journal. After all journaled data is committed to the destination volume, the source and destination will be in sync and the data on both volumes will be consistent. When the source and destination volumes are in sync, downstream latency will return within the application at the source volume. Similar to the high consistency mode, application latency and data consistency are important points to consider in a design that incorporates synchronous replication in high availability mode.

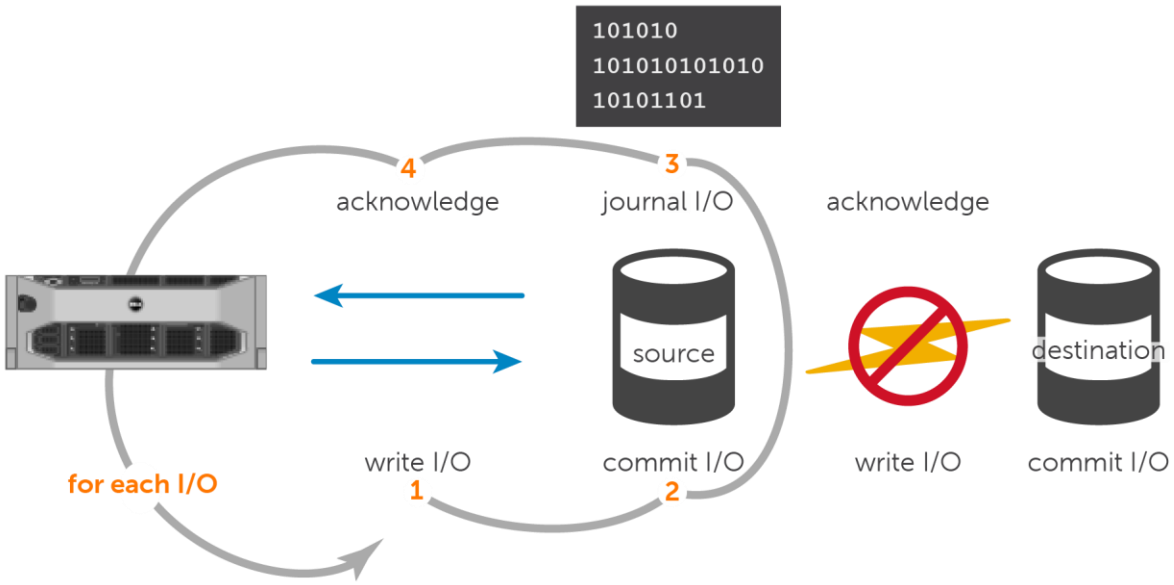

Figure 4 High availability mode synchronous replication in an out of date state

3.1.4 Mode migration

In Storage Center 6.5 or newer, replications may be migrated from one mode to another without manually having to destroy the replication and destination replica volumes, and then rebuild. For example, Asynchronous to Synchronous High Consistency, Synchronous High Consistency to Synchronous High Availability, or Synchronous High Availability to Asynchronous. Leveraging this feature can save significant time and replication bandwidth. It also reduces the data availability risk exposure associated with the time taken to destroy and rebuild a replica volume. Lastly, this method preserves predefined DR settings in Enterprise Manager that are tied to restore points and replica volumes. For all of these reasons individually or combined, it is recommended going forward to take full advantage of this feature.

Note: This feature is compatible with all replication modes except Legacy synchronous replication.

3.2 Minimal recopy

As discussed in section 2.1.1, synchronous replications configured in high availability mode allow write access to the source volume if the destination volume becomes unavailable or falls behind. While out of date, a journalizing mechanism shown in a previous figure tracks the write I/O that makes the source and destination volume inconsistent. Prior to Storage Center 6.3 with legacy replication, journaling was not performed and if the destination volume became unavailable and then later available, all data on the source volume needed to be re-replicated to the destination to get back in sync. The minimal recopy feature only needs to replicate the changed data contained in the journal to the destination volume in order to bring the source and destination volumes back in sync. This dramatically reduces the recovery time and data inconsistency risk exposure as well as the replication link bandwidth consumed to recover. Minimal recopy will also be employed in high consistency mode should the destination volume become unavailable during initial synchronization or an administrator invoked pause operation on the replication.

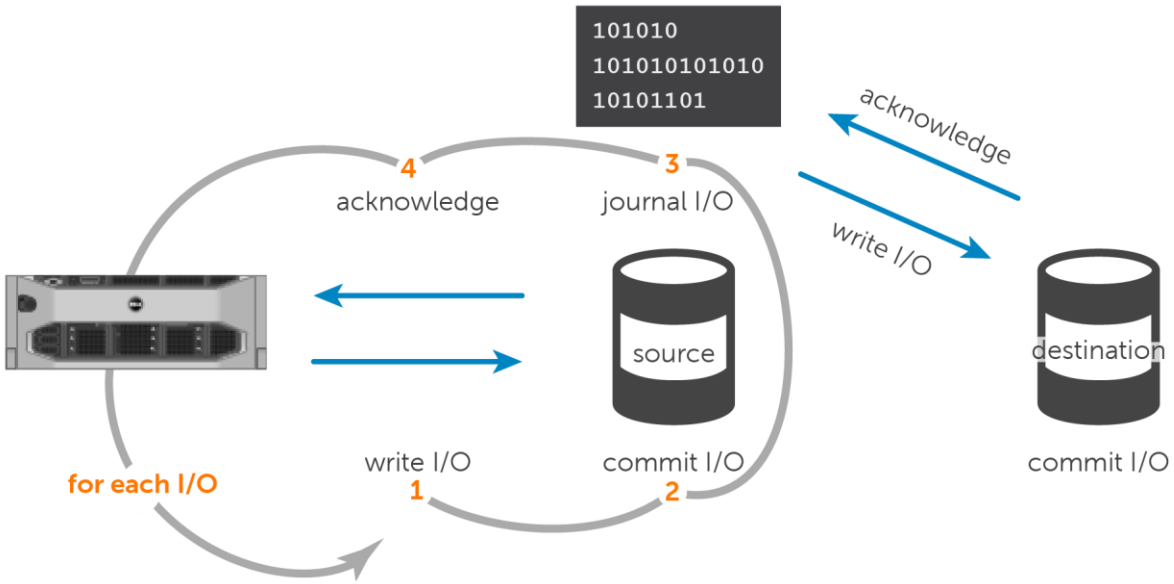

Figure 5 Flushing journaled writes to the destination volume to regain volume consistency

3.3 Asynchronous replication capabilities

Synchronous replication has seen numerous improvements over time and includes key features that were previously associated only with asynchronous replication.

3.3.1 Replays and consistency groups

The most notable asynchronous feature is the replication of Compellent Replays. In the past, only the active Replay data was replicated from source to destination. With Replays automatically replicated to the destination site, customers have more flexibility in recovery options with many historical restore points to choose from. By virtue of having Replay functionality, synchronous replication can be integrated with consistency groups and/or Replay Manager created Replays across volumes to enable snapshot interval consistency across replicated volumes. In high consistency mode, Replay consistency will be guaranteed. In high availability mode, Replay consistency is highly likely.

3.3.2 Pause

Synchronous replications configured in either high consistency or high availability modes can be paused. Pausing replication can facilitate multiple purposes. For example, it can be used to relieve replication link bandwidth utilization. In designs where replication bandwidth is shared, other processes can temporarily be given burstable priority. Pausing may also be preferred in anticipation of a scheduled replication link or fabric outage.

3.4 Multiple replication topologies

Dell Compellent now extends synchronous replication support beyond just a pair of volumes. A choice of two topologies or a hybrid combination of both is available.

3.4.1 Mixed topology

The Mixed topology, also known as 1-to-N (N=2 as of Storage Center 6.5), allows a source volume to be replicated to two destination volumes where one replication is synchronous and the additional replications are asynchronous. The maximum number of additional replications is set by the value of N. This topology is useful when data must be protected in multiple locations. If data recovery becomes necessary, a flexible choice of locations is available for recovery.

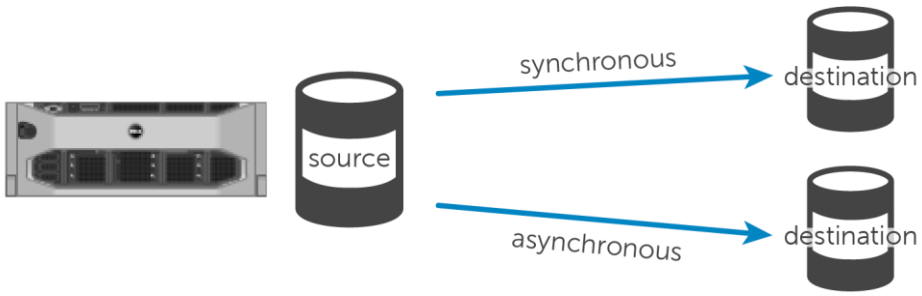

Figure 6 Mixed topology

3.4.2 Cascade topology

The Cascade topology allows asynchronous replications to be chained to synchronous or asynchronous replication destination volumes. This topology is useful in providing immediate reprotection for a recovery site. Similar to the Mixed topology, it provides a flexible choice of locations for data recovery or business continuation practices. It could also be used as a means of providing replicas of data in the same datacenter or a remote site. Copies of Microsoft SQL or Oracle databases for parallel TEST, DEV, or QA environments are two popular examples of this.

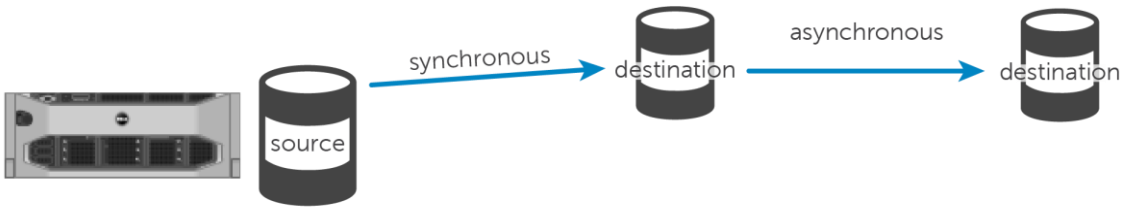

Figure 7 Cascade topology

3.4.3 Hybrid topology

A Hybrid topology can also be created by combining Mixed and Cascade topology types. This configuration is adaptable to just about any replica or data protection needs a business may require.

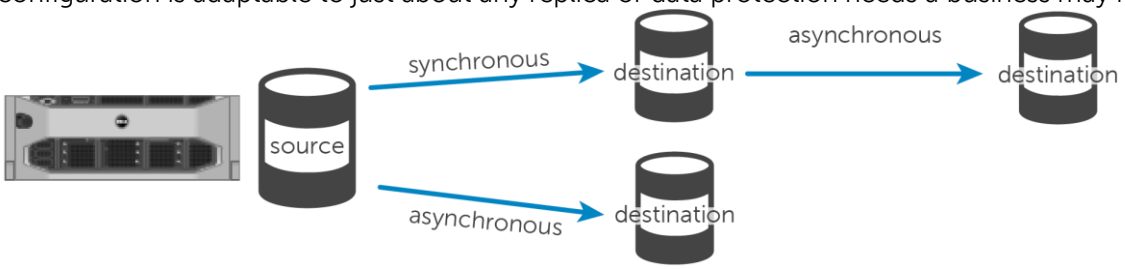

Figure 8 Hybrid topology

3.5 Live Volume

The Live Volume feature, which is discussed in detail later in this document, is built on replication. In previous versions of Storage Center, Live Volume was supported only with asynchronous replication. As of Storage Center 6.5, Live Volume is designed to work in conjunction with asynchronous and synchronous replication types. In addition, Live Volume supports many of the current synchronous replication features such as Modes of Operation and Mode Migration.

3.5.1 Preserve Live Volume

In Storage Center 6.5 or newer, recovering data from a secondary Live Volume, when the primary Live Volume is unavailable, is faster, easier, and more flexible. Secondary Live Volumes may be promoted to the primary Live Volume role that preserves volume identity and storage host mappings. Alternatively, data on a secondary Live Volume may be recovered by creating a new a View Volume and then mapping that View Volume to one or more storage hosts.

3.5.2 Live Volume managed replication

A Live Volume Managed Replication is an additional replication and replica volume that uses the primary Live Volume as its replication source. The Live Volume Managed Replication may be synchronous or asynchronous depending on the Live Volume configuration. To maintain data integrity and consistency, when a Live Volume swap role occurs automatically or manually, the Live Volume Managed Replication persistently follows the primary Live Volume as its source of replication.

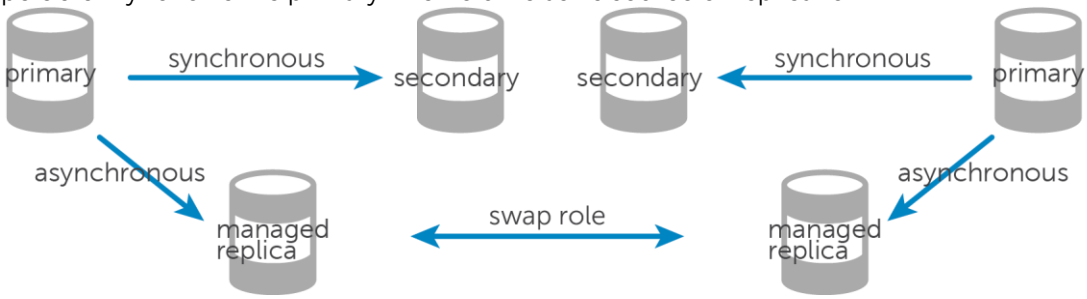

Figure 9 Live Volume Managed Replication before and after swap role

3.6 Enterprise Manager recommendations

Enterprise Manager periodically checks the status of replication and records the progress of completeness. In the event of a failure at the source site, Enterprise Manager 6.3 or newer provides a safety recommendation on the use of the destination replica. When using high consistency synchronous replication, data between source and destination must be consistent for Enterprise Manager to advise it is safe to use the destination replica for recovery.

When using high availability synchronous replication (or high consistency with the ability to pause replication), the data between source and destination volumes may or may not be consistent depending on whether the replication was in sync or out of date at the time of the failure. If at the time of failure replication was in sync, Enterprise Manager will advise that the destination replica volume is data consistent and safe to use for recovery. Conversely, if the synchronous replication was out of date, this means journaled transactions at the source volume likely have not been replicated to the destination and the destination replica is not data consistent and not recommend for use. At this point, the data recovery options would be to use a data consistent Replay as the recovery point or continue with using the inconsistent replica. In either case, the most recent transactions will have been lost at the destination but recovering from a Replay will provide a precise point in time as the recovery point.

3.7 Enterprise Manager DR recovery

Synchronous replication volumes are supported in the scope of the Enterprise Manager Predefined Disaster Recovery and DR Activation features. Those that have used this feature with asynchronously replicated volumes in the past can extend the same disaster recovery test and execution processes to synchronously replicated volumes. Enterprise Manager and its core functionality is freely available to Compellent customers making it an attractive and affordable tool for improving Recovery Time Objectives. Note that DR settings cannot be predefined for Live Volumes nor can Live Volume restore points be Test Activated.

3.8 Support for VMware SRM

Asynchronous or synchronous (either mode) replication types can be leveraged by Site Recovery Manager protection groups, recovery plans, and reprotection. Live Volume and Live Volume managed replications in Storage Center 6.5 are not supported together with Site Recovery Manager.

4 Synchronous Replication use cases

Replicating data can be a valuable and useful tool, but replication by itself serves no purpose without tying it to a use case to meet business goals. The following sections will highlight sample use cases for synchronous replication.

4.1 Overview

Fundamentally speaking, replication is typically used as a business continuation process to enable high availability or disaster recovery, a data protection process to enable image or file level backup and recovery, or a development tool to generate copies of data in near or remote locations for application development or testing purposes. For many business use cases, asynchronous replication provides a good balance of meeting Recovery Point Objective (RPO) and Recovery Time Objective (RTO) service level agreements without a cost-prohibitive infrastructure such as dark fibre, additional networking hardware, or storage. This is why the application of asynchronous replication is found between most datacenters particularly where longer distances are involved.

However, there are some business verticals and processes where a design requirement or constraint is placed on customer transactions. The constraint is that data representing customer or business transactions cannot be lost. Regardless of where the need originates, the method of replication that satisfies zero transaction loss is synchronous. The next few sections provide a selection of platform integration examples of synchronous replication with a focus on high consistency for zero data loss or high availability for relaxed data consistency requirements.

4.2 High consistency

By far, the primary use case for synchronous replication is the guarantee of zero loss of data or absolute data consistency between the source and destination replica volume. As mentioned earlier, this constraint may be introduced in a number of different ways. Zero data loss is a demand where Compellent synchronous replication, high consistency mode is a definite asset.

4.2.1 VMware and Hyper-V

When virtualized, server workloads in the datacenter are encapsulated into a small set of files that represent the virtual BIOS, virtual hardware resources, and the virtual disks that contain regularly updated data, depending on the virtual machine role. Virtual machines work particularly well with replication because they are portable and hardware independent by nature. Their inherent mobility, combined with replication, allows them to easily migrate from one site to another, with comparatively little effort required to bring them online at the destination site. Virtual machines may be relocated for load balancing or disaster avoidance/recovery purposes. Whatever the reason for relocation, high consistency synchronous replication will ensure that the contents of the virtual machine at the source and destination match. In the event the vSphere or Hyper-V virtual machine needs to be migrated to a host or cluster of hosts at the destination site, data consistency of the virtual machine being brought up at the destination site is guaranteed. Disaster recovery will be covered in detail in section 4.5.

For more information on configuring a pre-defined Compellent Enterprise Manager DR plan for Hyper-V, please refer to the *Dell Compellent Storage Center Enterprise Manager DR Plan for Hyper-V Demo Video* on Dell TechCenter at http://en.community.dell.com/techcenter/storage/w/wiki/5089.compellent-video-content.aspx.

Please note that Enterprise Manager DR plans cannot be predefined with Live Volumes. Predefined DR plans are supported with regular (asynchronous or synchronous) volume replications or with a managed (cascaded or hybrid) asynchronous replication from a Live Volume.

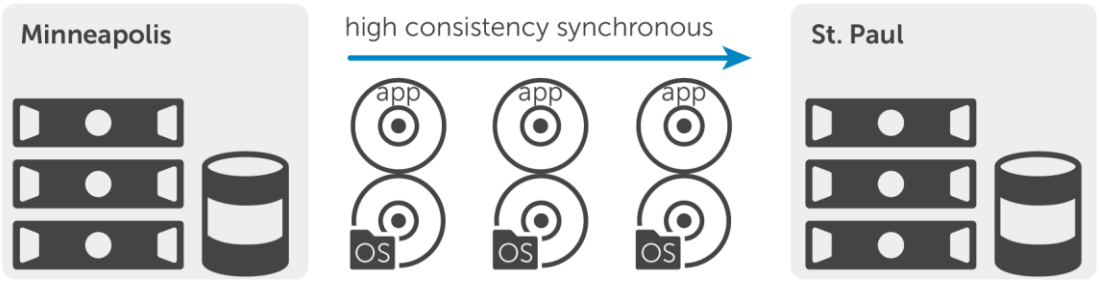

Figure 10 High consistency with consolidated vSphere or Hyper-V sites

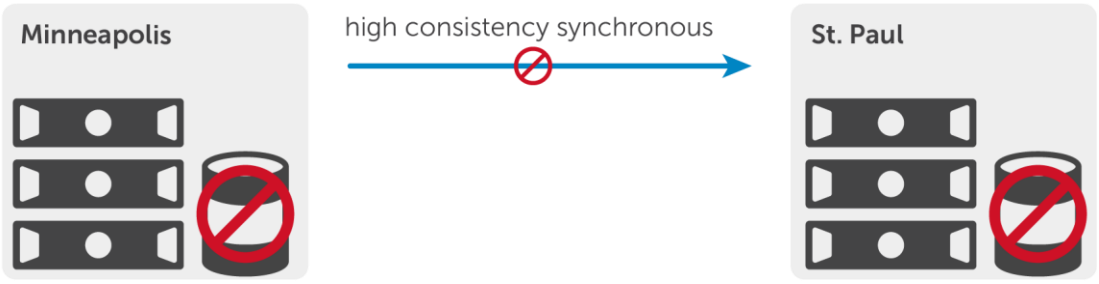

Figure 11 A replication link or destination volume issue in St. Paul results in a VM outage in Minneapolis

4.2.2 Microsoft SQL Server and Oracle Database/Oracle RAC

Database servers and clusters in critical environments are often designed to provide highly available, large throughput, and low latency access to data for application tier servers and sometimes directly to application developers or end users. Database servers differ from virtual machines in that protection of the database volumes is paramount while protection of the operating system is not required for data recovery. However, booting from SAN and replicating that SAN volume to a remote site with compatible hardware can drastically improve RTO. Similar to virtual environments, the critical data may be spread across multiple volumes. When designing for performance, application, or instance isolation, this is often the case. Unless the replication is paused, consistency between volumes is guaranteed in high consistency mode because the write order at the destination will mirror the write order at the source, otherwise the write will not happen in either location. This is the fundamental premise of high consistency mode detailed in section 3.1.2.

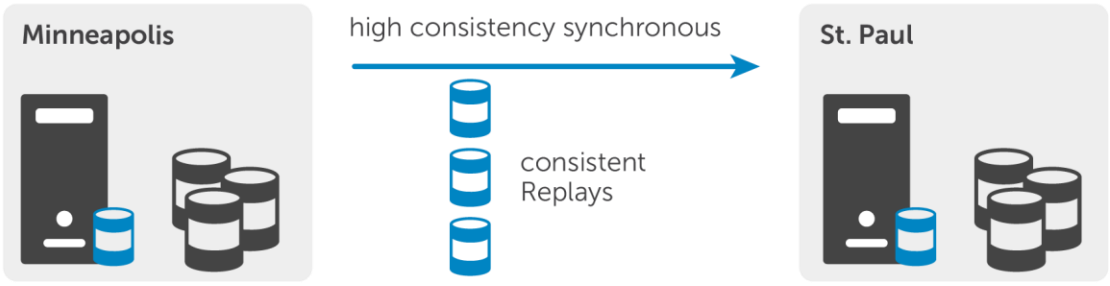

Figure 12 High consistency with databases

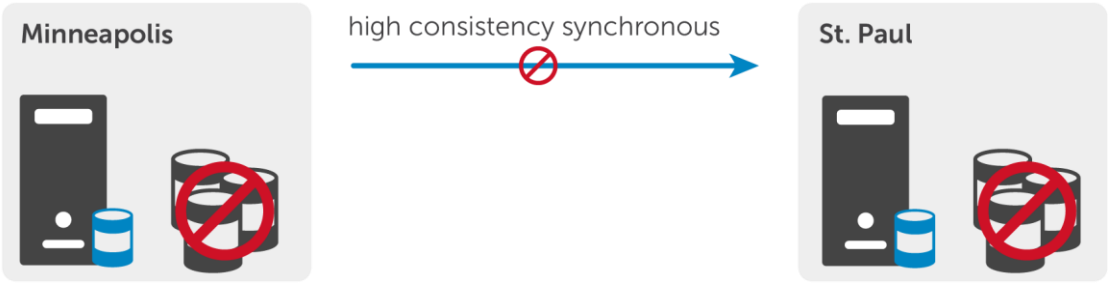

Figure 13 A replication link or destination volume issue in St. Paul results in database outage in Minneapolis

To summarize, there are high consistency use cases that can be integrated with virtualization as well as database platforms. The key benefit being provided is data consistency and zero transaction loss. Keep in mind that the infrastructure supporting synchronous replication between sites must perform sufficiently. In the case of high consistency, the supporting infrastructure must be highly redundant and immune to outages for slowness or an outage of the replication link or the destination site is reflected equally at the source site where the end user applications are running.

An important factor when considering the type of replication to be implemented is that the infrastructure required to keep two sites well connected, particularly at greater distances, often comes at a premium. Stakeholders may be skeptical about implementing a design that can have such an impact on application availability and favor asynchronous replication over synchronous replication. However, with the high availability synchronous replication offered in Storage Center, customers have additional flexibility compared to legacy synchronous.

4.3 High availability

Most organizations prefer asynchronous replication either for its cost effectiveness or its significant reduction in risk of an application outage, should the destination storage become unavailable. With the introduction of high availability synchronous replication mode in Storage Center 6.3, data consistency can be achieved under normal operating circumstances. However, if unexpected circumstances arise resulting in a degradation or outage of the replication link or destination storage, latency or loss of production application connectivity at the replication source is not at risk. While this deviates slightly from the industry recognized definition of synchronous replication, it adds flexibility that is not found in high

consistency synchronous by blending desirable features of both synchronous and asynchronous replication. In addition, Storage Center automatically adapts to shifting destination replica availability.

4.3.1 VMware and Hyper-V

Encapsulated virtual machines will be replicated in a data consistent manner just as they are using high consistency mode replication. The difference of behavior comes into play if and when the replication link (or the destination replica volume) becomes burdened with excess latency or unavailable. Instead of failing writes from the hypervisor, the writes will be committed and journaled at the source volume allowing applications to continue functioning but at the expense of a temporary lack of data consistency while the destination volume is unavailable.

In the following examples, note that using high availability mode in place of high consistency mode does not necessarily allow the design to be stretched over further distances. High availability mode is still a form of synchronous replication and should not be confused with asynchronous replication. Adding significant distance between sites generally results in latency increases which will still be observed in the applications at the source side for as long as the high availability replication is in sync.

Finally, if virtual machines are deployed in a configuration that spans multiple volumes, strongly consider Replay Manager or consistency groups. Replay Manager is covered in section 11.6.

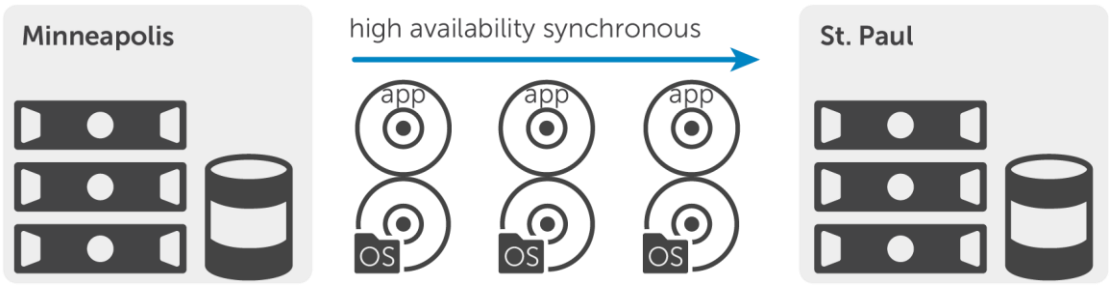

Figure 14 High availability with consolidated vSphere or Hyper-V sites

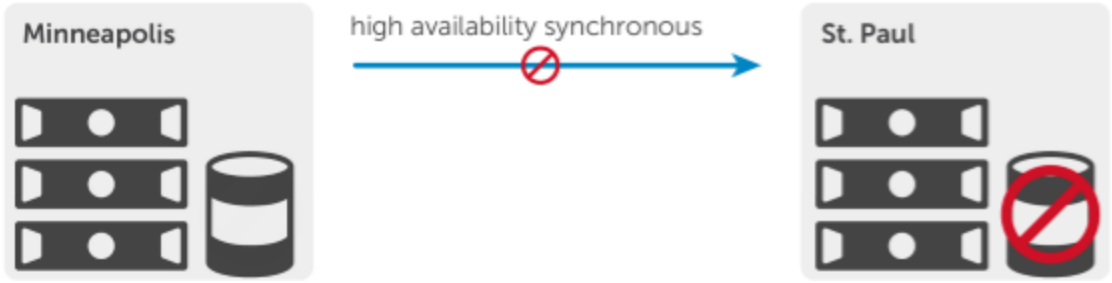

Figure 15 A replication link or destination volume issue in St. Paul results in no VM outage in Minneapolis

4.3.2 Microsoft SQL Server and Oracle Database/Oracle RAC

As discussed in section 4.3.1, the behavioral delta between high availability and high consistency is minimal until extreme latency or an outage impacts the destination volume availability. If high availability synchronous replication falls into an out-of-date state, write I/O at the source volume is journaled and the destination volume becomes inconsistent. In terms of recovery, this may or may not be acceptable. A

feature in Enterprise Manager advises customers on whether or not the active Replay on the destination volume is safe to recover from at a data consistency level. In the event that Enterprise Manager detects the data is not consistent, the recommendation is to revert to the most recent consistent Replay associated with the destination volume (another new feature in synchronous replication – replication of Replays).

For storage hosts with data confined to a single volume, special considerations are not necessary. However, if the host has application data spread across multiple volumes (for example, a VM with multiple virtual machine disk files, or a database server with instance or performance isolation of data, logs, and other files) then it becomes fairly critical to ensure Replay consistency for the replicated data that will be used as a restore point. Ensuring all volumes of a dataset are quiesced and then snapped at precisely 8 o'clock for example, provides a data consistent restore point across volumes supporting the dataset. This Replay consistency is accomplished creating Replays with Replay Manager (especially recommended for Microsoft products through VSS integration) or by containerizing volumes by use of consistency groups.

To create consistency across Relays using consistency groups, a Replay profile is created with a Replay Creation Method of Consistent. This Replay profile is then applied to all volumes containing the dataset. For virtual machines, the volumes would contain the virtual disks representing the c: drive, d: drive, or Linux mount points such as /, /tmp, etc. For Microsoft SQL database servers, the volumes may represent system databases, application databases, transaction logs, and tempdb. For Oracle Database/Oracle RAC servers, the volumes may contain data or database files (raw data, index, data dictionary), control files, online redo logs, OCR files, or voting disk. For either database platform, separate volumes for hot dumps, archived redo logs, or boot from SAN may exist but typically would not need to be included in a consistency group with the key database files.

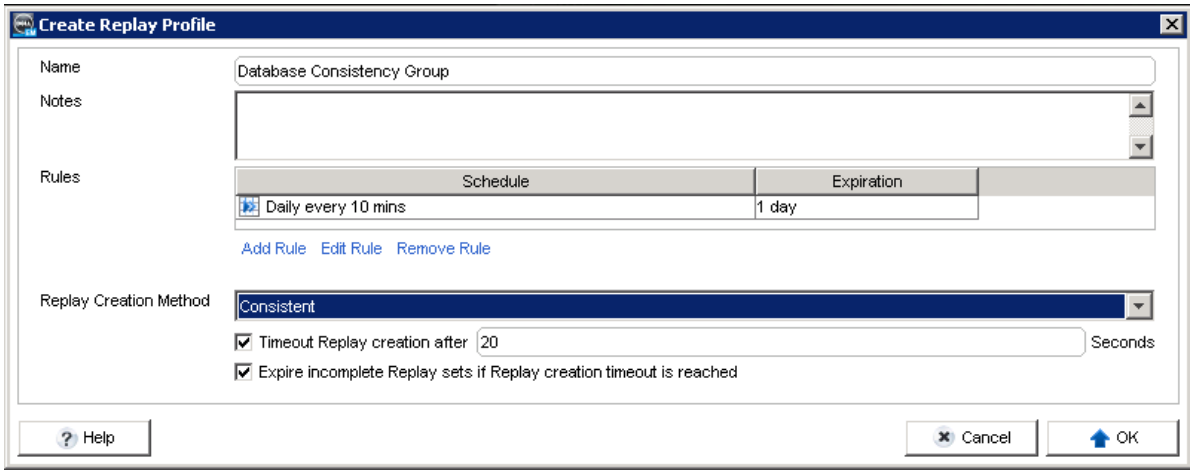

Figure 16 Creating a consistency group in Enterprise Manager

Another method of capturing consistency in Replays across volumes (and perhaps more useful for customers with Microsoft Windows, SQL, Exchange, Hyper-V, or VMware vSphere) would be to use Compellent Replay Manager. Replay Manager has the underlying storage integration and VSS awareness required to create application consistent Replays, across volumes if necessary, which can then be replicated synchronously (either mode) or asynchronously.

Once data is frozen with consistency across volumes using Replay Manager or consistency groups, those Replays will be replicated to the destination volume where they can serve as historical restore points for high availability mode recovery, disaster recovery, or remote replicas which will be discussed in the coming sections.

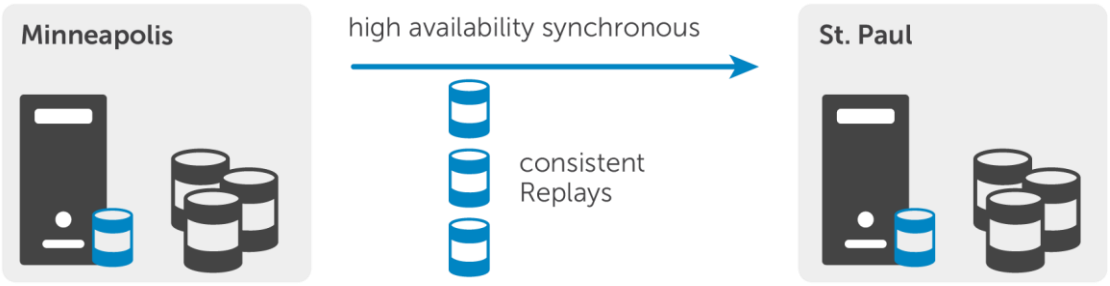

Figure 17 High availability with databases

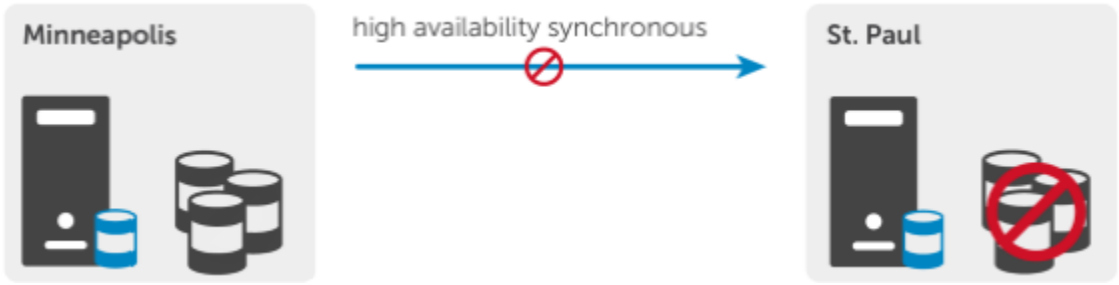

Figure 18 A replication link or destination volume issue in St. Paul results in no database outage in Minneapolis

4.4 Remote database replicas

One practice commonly found in organizations with Microsoft SQL or Oracle database technologies is to create copies of databases. There are several reasons to clone databases and most of them stem from the common principle of minimal or no disruption to the production database and thus to the application and end users. To identify a few examples, at least one separate copy of a database is maintained for application developers to develop and test code against. A separate database copy is maintained for DBA staff to test index changes, queries, and for troubleshooting areas such as performance. A synchronous or asynchronous copy of the production database may be maintained for I/O intensive queries or reporting. Compellent Storage Center Replays and view volumes are a natural tactical fit for fulfilling database replica needs locally on the same Storage Center array.

However, if the replica is to be stored on a different Storage Center array, whether or not it is in the same building or geographic region, replication or portable volume must be used to seed the data remotely, and replication should be used to refresh the data as needed. For the purposes of developer or DBA testing, asynchronous replication may be timely enough. However, for reporting purposes, synchronous replication will ensure data consistency. The choice of providing zero data loss through high consistency mode or a more flexible high availability mode should be decided ahead of time with the impacts of each mode well understood.

Dell Compellent Replays, as well as asynchronous and synchronous replication, are natively space and bandwidth efficient on storage and replication links respectively. Only the changed data is frozen in a Replay and replicated to remote Storage Centers. In the following figure, notice the use of high availability synchronous replication within the Minneapolis datacenter. Although the two Storage Centers are well connected, the risk of internal reporting database inconsistency does not warrant a production outage for the organization.

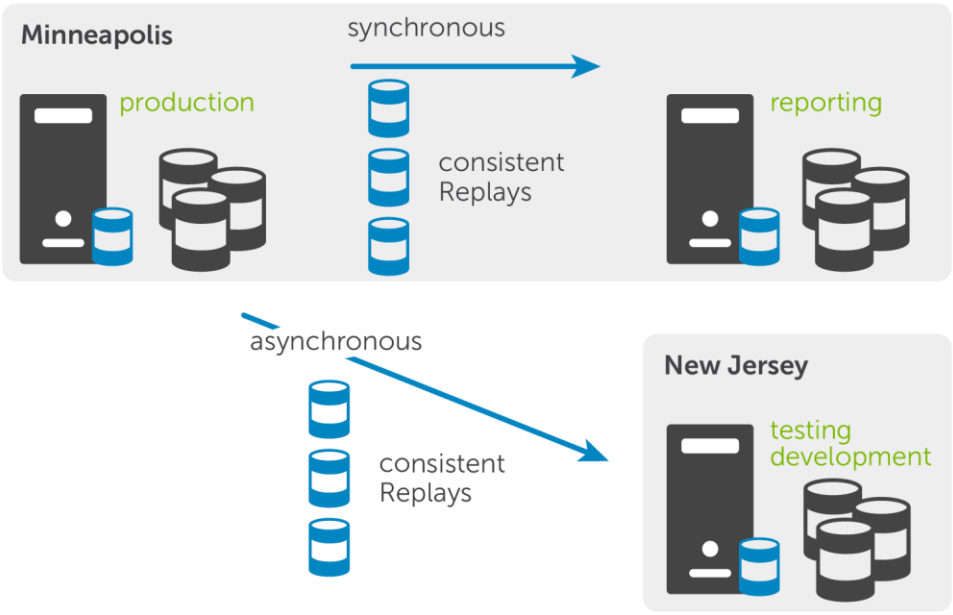

Figure 19 Database replicas distributed in a mixed topology

4.5 Disaster recovery

With data footprints growing exponentially, backup and maintenance windows shrinking along with the cost of storage, and the impact of downtime gnawing on the conscience of businesses, migrating to online storage based data protection strategies is becoming a common theme for medium to enterprise sized organizations. Legacy processes which dumped data to tape were cost effective and acceptable early on but the convergence of key decision making factors has prompted a shift from yesterday's nearline storage to the more affordable and efficient online storage of today. Data replication within or between sites is the ubiquitous backbone for much more scalable data protection strategies. With replication in place as the data mover, an assortment of vendor and platform provided tools and methods can be coupled to replication to form a manual or electronically documented and reliable recovery process. The Compellent support for multiple replication topologies really comes into play in the disaster recovery conversation because it adds a lot of flexibility for customers wanting to provide data protection for multiple or distributed site architectures. Before getting into platform specific examples, two fundamental disaster recovery metrics need to be understood as they will be referenced throughout business continuation planning discussions.

Recovery Point Objective: Also known as RPO. The point in time at which the data is recovered from. An RPO is negotiated with business units and predefined in a disaster recovery plan. In terms of replication,

the keys to achieving an RPO target are choosing the appropriate replication type, making sure replication is current (as opposed to out of date or behind), and knowing the tools and processes required to recovery from a given restore point.

Recovery Time Objective: Also known as RTO. The elapsed recovery time allowed to bring up a working production environment. Just like RPO, RTO is negotiated with business units and predefined in a disaster recovery plan and may also be included in a Service Level Agreement (SLA). The keys to achieving targeted RTO may vary from datacenter to datacenter but they all revolve around process efficiency and automation tools wherever possible. Replication is the quintessential contributor to meeting RTO, especially at large scale.

By leveraging replication, aggressive RPOs and RTOs can be targeted. Data footprint and rate of change growth may be continuous, but feasible RPO and RTO goals do not linearly diminish as long as the replication circuit can scale to support the amount of data being replicated and the rate of change.

4.5.1 Hyper-V and VMware

As discussed in previous sections, replicating the file objects that form the construct of a virtual machine takes advantage of the intrinsic encapsulation and portability attributes of a VM. Along with hardware independence, these attributes essentially mean the VM can be moved to any location where a supported hypervisor exists, and a VM or group of VMs can be quickly and easily be registered and powered on depending on the hypervisor and the automation tools used to perform the cutover. Compare this to legacy methods of disaster recovery where, at the recovery site, physical or virtual servers are built from the ground up, applications needed to be installed and configured, and then large amounts of data needed to be restored from tape. At the time a disaster is declared, virtual machines and their configured applications are essentially ready to be added to the hypervisor's inventory and powered on. Virtualization and replication shave off massive amounts of recovery time, which helps achieve targeted RTO. When the VMs are powered on, their application data payload from the most recently completed replication is already present meeting the RPO component of the DR plan.

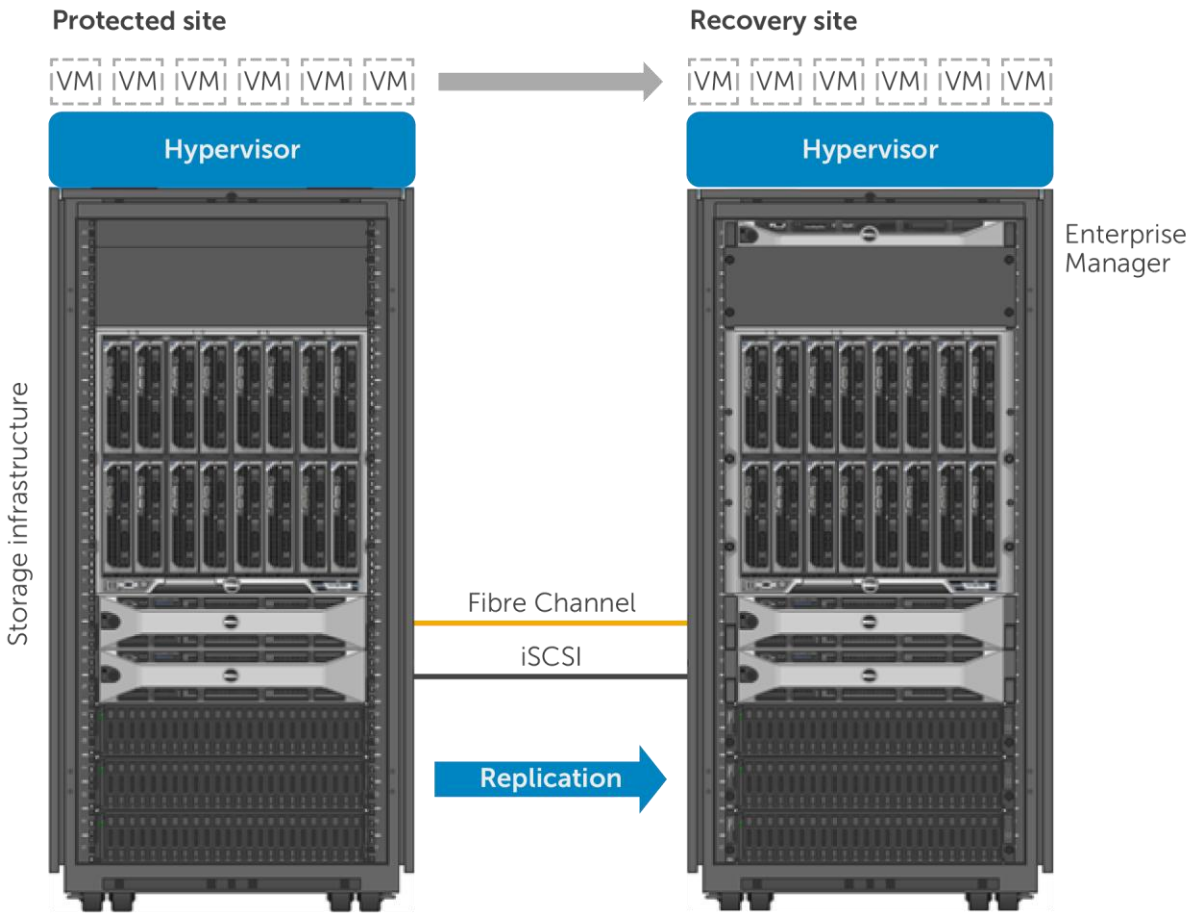

Figure 20 Virtualization and replication combined meet aggressive RTOs and RPOs

4.5.2 Microsoft SQL Server and Oracle Database/Oracle RAC

Aside from infrastructure servers such as Active Directory, LDAP, DNS, WINS, and DHCP, database servers are among the first assets to be recovered. Typically, databases are classified as Tier 1 infrastructure (Tier 1 assets included in a DR plan receive first recovery priority) and database servers are the first tier that must be brought online in a multi-tier application architecture. The next online is the application tier servers, and then last the application front end (on either client desktops or a load balanced web portal).

RTO is a paramount metric to meet in testing and executing a live business continuation plan. In a DR plan, all steps are predefined and executed in order according to the plan; some steps may be carried out in parallel. Successfully recovered database servers are a required dependency beginning early in the DR plan. This includes bringing up application and web servers that have a critical tie to the database server. The more databases a shared database server hosts, the broader the impact because the number of dependent application and front end tiers fan out.

Industry analysis reflects data growing at alarming rates across many verticals. Providing performance and capacity is not a challenge with current technology, but protecting the data is. Data growth drives changes in technology and strategy so that SLAs, RTOs, and RPOs can still be maintained even though

they were defined when data was a fraction of the size it is today. Restoring 10TB of data from tape is probably not going to satisfy a 24 hour RTO. Data growth on tapes means that there is a growing number of sequential-access, long-seek time tapes for restoration. This diminishes the chances of meeting RTO, and increases the chances that one bad tape will cause data recovery to fail. Data replication is a major player in meeting RTO.

Intra-volume consistency is extremely important in a distributed virtual machine disk or database volume architecture. Comparing the synchronous replication modes, high consistency guarantees data consistency between sites across all high consistency replicated volumes. Unfortunately, this is at the cost of destination site latency, or worse, downtime of the production application if the destination volume becomes unavailable or exceeds latency thresholds.

Outside of use cases that require the textbook definition of synchronous replication, high availability mode (or asynchronous) may be a lot more attractive for DR purposes. This mode offers data consistency in the proper conditions, as well as some allowance for latency while in sync. However, consistency is not guaranteed if production application uptime is jeopardized should the destination volume become unavailable.

Because consistency cannot be guaranteed in high availability mode, it is important to implement VSS integrated Replay Manager Replays or consistency groups with high availability synchronous replication where a multiple volume relationship exists (this is commonly found in both SQL and Oracle environments). While this will not guarantee active Replay consistency across volumes, the next set of frozen Replays that have been replicated to the remote array should be consistent across volumes.

4.5.3 Preparing and executing volume and data recovery

With the appropriate hypervisor, tools, and automation in the DR plan, powering on a virtual machine is relatively simple. Likewise, preparing the volumes for use and getting the database servers up and running requires a quick process compared to legacy methods. This is especially true when replicating database server boot from SAN (BFS) volumes with similar hardware at the DR site.

When choosing to access volumes at the DR site, it is important to consider the purpose. Is this a validation test of the DR plan, or is this an actual declared disaster? As an active replication destination target (regardless of asynchronous, synchronous, mode, or topology), destination volumes cannot be mounted for read/write use to a storage host at the DR site (For circumstances involving Live Volume, refer to section 7). To perform a test of the DR plan, present view volumes from the Replays of each test volume to the storage hosts. Replays and view volumes are available for both asynchronous and synchronous replications in either high consistency or high availability mode. Replays and view volumes are beneficial during DR testing because replication continues between the source and destination volumes to maintain RPO in case an actual disaster occurs during the test. Conversely, if a disaster is being declared and the Activate Disaster Recovery feature is invoked, then replication from source to destination needs to be halted (if it has not been already by the disaster) if the active volume at the destination site is intended for data recovery in the DR plan.

Dell Compellent Enterprise Manager is a unified management suite available to Storage Center and FS8600 customers. Enterprise Manager has disaster recovery features built into it that can create, manage, and monitor replications, as well as automate the testing and execution of a predefined DR plan.

Figure 21 Enterprise Manager bundles DR automation tools to help meet RTO requirements

For virtualization and database use cases alike, Enterprise Manager is used to create asynchronous or synchronous replications. These replications predefine the destination volumes that will be presented to storage hosts for disaster recovery.

Note: The destination volumes are for DR purposes only and cannot be used actively in a Microsoft or VMware vSphere stretched cluster design.

In most cases, predefined destination volumes are data volumes. Where physical hosts are involved, boot from SAN volumes can also be included in the pre-defined DR plan to quickly and effortlessly recover physical host and applications as opposed to rebuilding and installing applications from scratch. Rebuilding takes significant time, is error prone, and may require subject matter expert knowledge of platforms, applications, and the business depending on how well detailed the build process is in the DR plan. Confusion or errors during test or DR execution lead to high visibility failure. Detailed and current DR documentation provides clarity at the DR site. Process inconsistency and errors are mitigated by automation or closely following DR documentation. With these points in mind, the benefit of automating a DR plan with Enterprise Manager (or a similar tool) is clear. From the moment a disaster is declared by the business, the RTO is in jeopardy; automation of tasks saves time and provides process consistency.

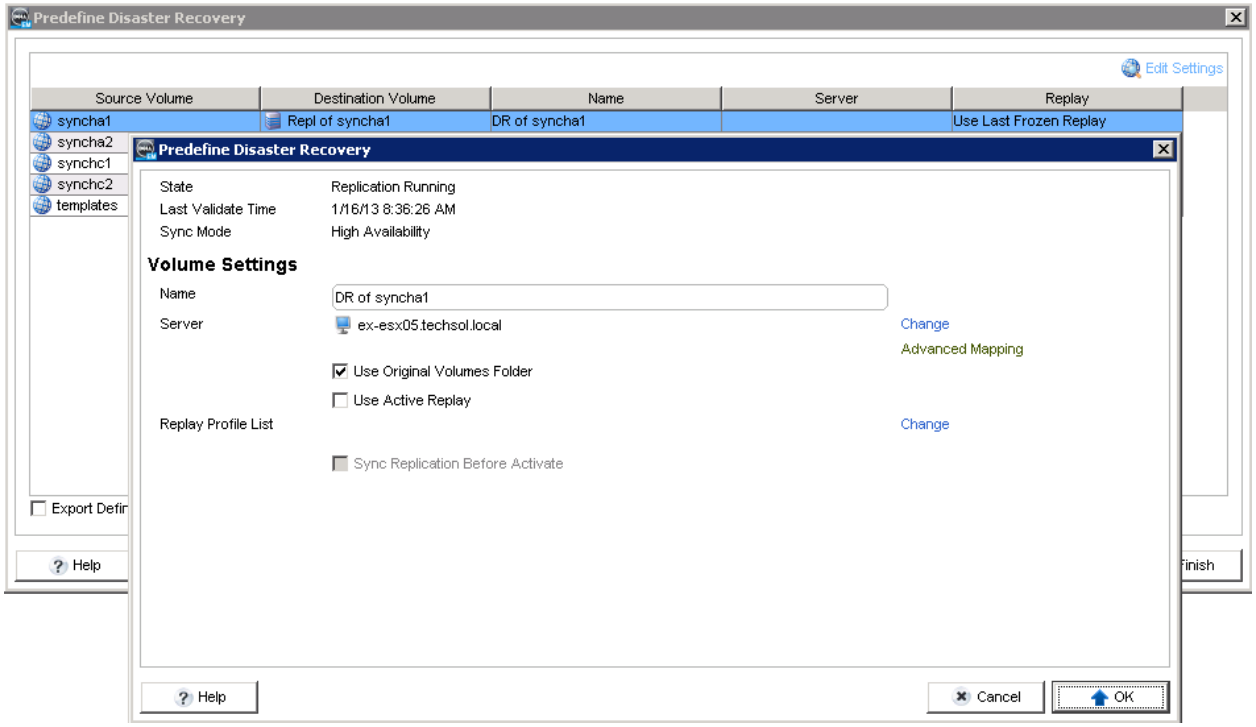

Figure 22 Predefining a DR plan in Enterprise Manager

Once the volumes are prepared and presented manually or automatically by Enterprise Manager or PowerShell scripting using Compellent cmdlets, the process of data recovery continues. For SQL and Oracle database servers, databases are attached and various scripts are run to prepare the database server and applications for production use (such as to resync login accounts). For VMware vSphere and Hyper-V hosts, VM datastores are now visible to the hosts and VMs need to be added to the inventory so that they can be allocated as compute and storage resources by the hypervisor and then powered on.

In Hyper-V 2008R2, the configuration file for each virtual machine must be generated with the correct number of processors, memory, network card, and attached virtual machine disk files. This is a process that is documented or scripted prior to the DR event. Hyper-V 2012 includes a virtual machine import wizard that is able to import the existing virtual machine configuration located on replicated Compellent storage, rather than generating a new configuration for each VM from scratch. Once the VMs are added to inventory in Hyper-V 2008R2 or Hyper-V 2012, they can be powered on. All versions of VMware vSphere have the same capability as Hyper-V 2012 in that once datastores are presented to the vSphere hosts, the datastores can be browsed and the virtual machine configuration file that is located in each VM folder can be added to inventory and then powered on. A manual DR process, especially in an environment with hundreds or thousands of virtual machines, quickly eats into RTO. The automation of discovering and adding virtual machines to inventory is covered in the next section.

VMware vSphere Site Recovery Manager is a disaster recovery and planned migration tool for virtual machines. It bolts onto an existing vSphere environment and leverages Compellent certified storage and array based replication. Both synchronous and asynchronous replication are supported as well as each of their native features with the exception of Live Volume. With this support, customers can strive for RPOs

that are more aggressive and maintain compatibility with third party automation tools, like SRM, to maintain RTOs in large VMware virtualized environments.

For vSphere environments, SRM invokes the commands necessary for tasks (such as managing replication, creating Replays, creating view volumes, and presenting and removing volumes from vSphere hosts) to be performed at the storage layer without removing Enterprise Manager from the architecture. The storage related commands from SRM flow to the Storage Replication Adapter (SRA) and then to the Enterprise Manager server. For this reason, an Enterprise Manager server needs to remain available at the recovery site for the automation to be carried out. Aside from SRM, in a heterogeneous datacenter, Enterprise Manager or PowerShell scripting would be needed to carry out the DR automation for Hyper-V or physical hosts.

Beyond the scope of storage, SRM automates other processes of DR testing, DR recovery, and planned migrations making it a major contributor to meeting RTO goals. SRM takes care of important, time-consuming tasks such as adding virtual machines to inventory at the DR site, modifying TCP/IP address configurations, VM dependency, power on order, and reprotection of virtual machines.

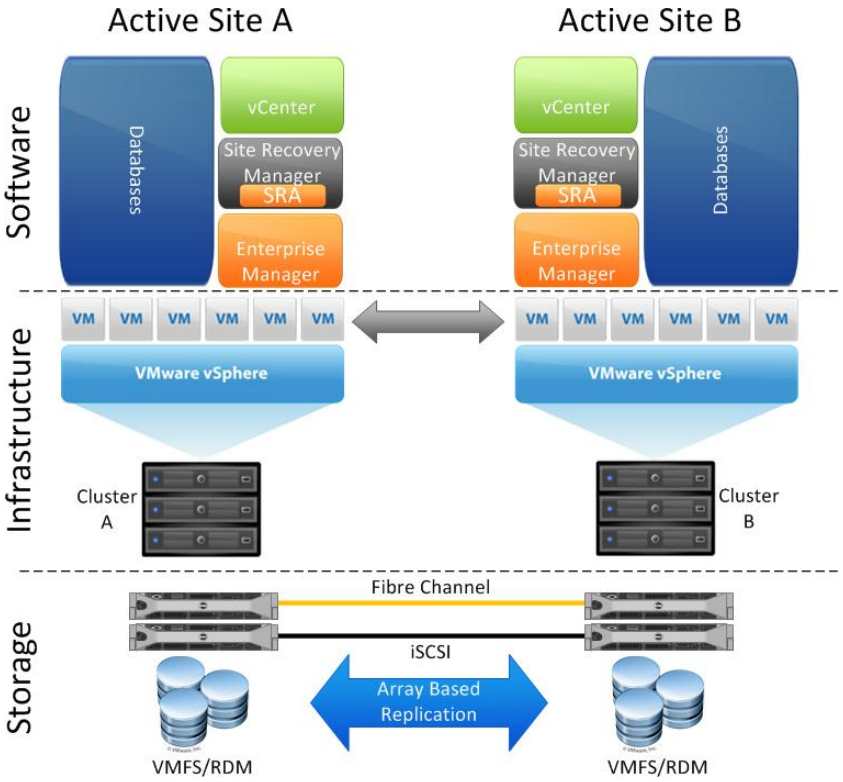

Figure 23 VMware Site Recovery Manager and Dell Compellent Storage Center active/active architecture

Enterprise Manager server must be available to perform DR testing or an actual DR cutover for an automated DR solution involving Compellent Storage Center replication. This means making sure that at least one Enterprise Manager server resides at the recovery site so that it can be engaged when needed for DR plan execution. The Enterprise Manager server labeled as a physical server in Figure 20, also represents

a virtualization candidate if it is already powered on and not required to automate the recovery of the vSphere infrastructure where it resides.

4.5.4 Topologies and Modes

Replication of data between or within datacenters is the fastest, most efficient, and automated method for moving large amounts of data in order to provide replica data, data protection, and business continuation in the event of a disaster. This section provides examples of the different topologies and modes available with synchronous replication in addition to where asynchronous replication is appropriate.

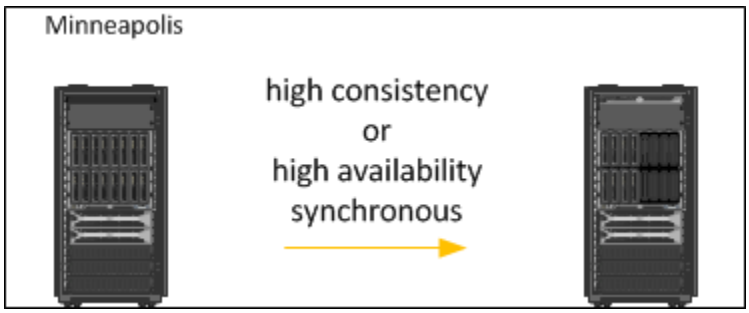

Figure 24 DR within a campus – Standard topology – fibre channel replication

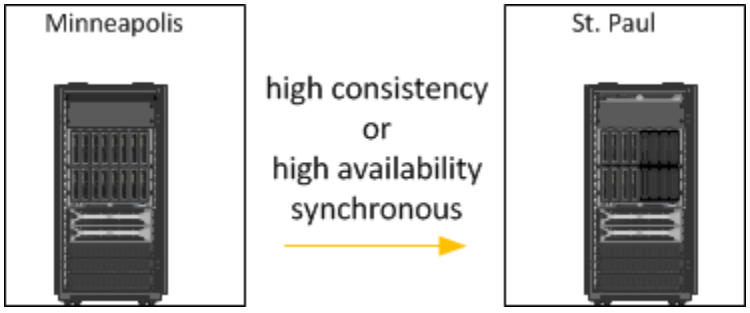

Figure 25 Metro DR site – Standard topology – fibre channel replication

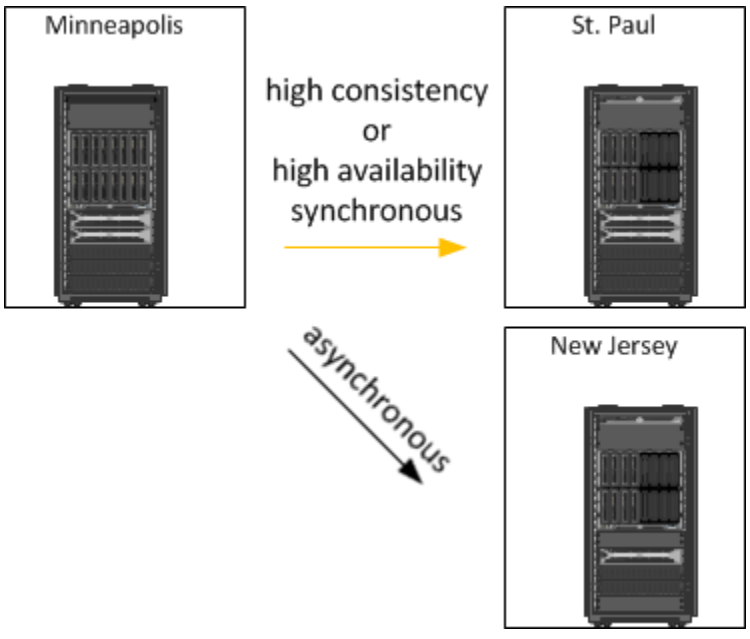

Figure 26 Metro and remote DR sites – Mixed topology (1-to-N) – fibre channel and iSCSI replication

Figure 27 Intra campus and metro DR sites – Cascade topology – fibre channel and iSCSI replication

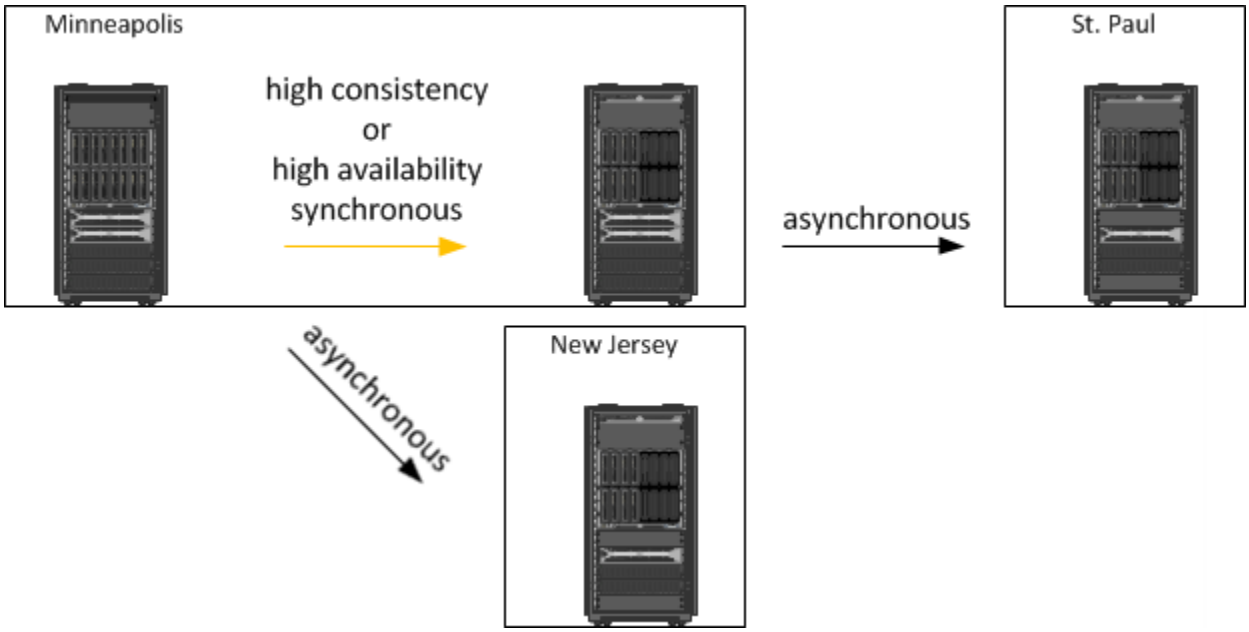

Figure 28 Intra campus, metro, and remote DR sites - Hybrid topology – fibre channel and iSCSI replication

The examples in Figure 24 through Figure 28 serve to represent physical hosts or virtual machines. Some notable details in the examples are:

- A 1U Enterprise Manager server is racked at each recovery site to facilitate DR testing and cutover automation through Enterprise Manager, VMware Site Recovery Manager, or both. The Enterprise Manager shown is for logical representation only. Enterprise Manager may be hosted by a virtual machine.
- Compellent Storage Center supports Fibre Channel and iSCSI based replication.
- Compellent Storage Center supports asynchronous replication as well as multiple modes of synchronous replication.
- Either mode of synchronous replication latency will impact production applications at the source side. Size the configuration for adequate replication bandwidth, controllers, and spindles at the recovery site to efficiently absorb throughput.

- Provide adequate hardware and datacenter redundancy at the recovery site when implementing high consistency synchronous replication. Aside from a pause operation, unavailability of a destination replica volume leads to unavailability of the source volume and will result in a production application outage.

5 Live Volume overview

Live Volume is a unique data mobility feature for Compellent Storage Center that builds on the Dell Fluid Data architecture. It enables non-disruptive data access, data migration, and stretched volumes between Storage Center arrays.

5.1 Reference architecture

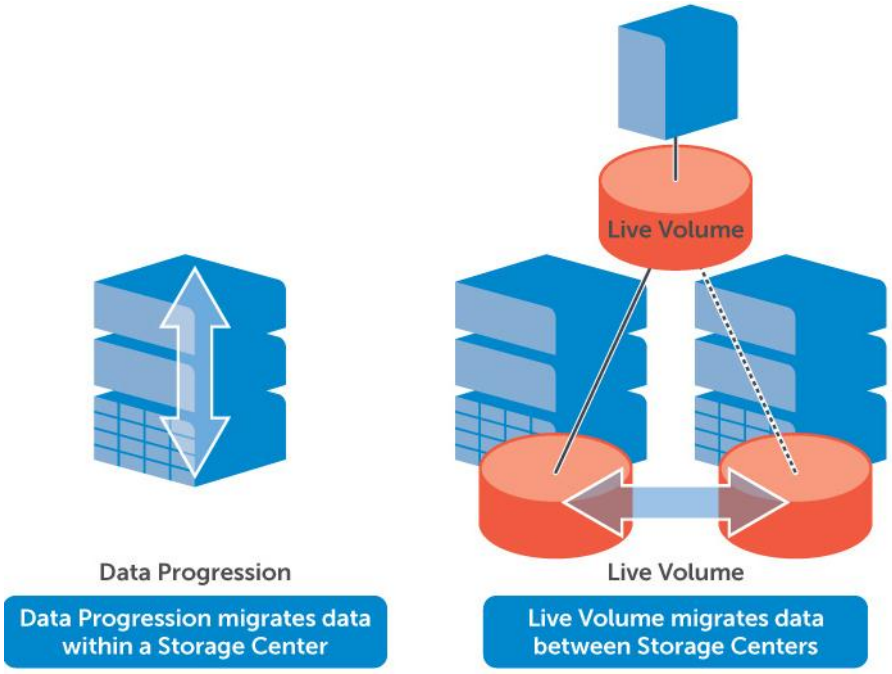

Figure 29 Live Volume migration

Live Volume is a software-defined solution that is integrated into the Compellent Storage Center Controllers. It is designed to operate in a production environment that allows storage hosts and applications to remain operational during volume migration or data movement regardless of the geographical distance between arrays.

Live Volume increases operational efficiency, reduces planned outages, and enables planned migrations as well as disaster avoidance. The Live Volume feature provides the following powerful options.

- Storage follows mobile virtual machines and applications in virtualized environments
- Supports automatic or manual mechanisms to migrate virtual machine storage as virtual machines are migrated within or across hypervisor clusters
- Zero application downtime for planned maintenance outages
- Enables all data to be moved non-disruptively between Storage Centers to achieve full planned site shutdown without downtime
- On-demand load balancing. Live Volume enables data to be relocated as desired to distribute workload between Storage Centers

- Stretch Microsoft clustered volumes between geographically disperse locations
- Live Volume allows VMware vSphere and Microsoft Clusters to see the same disk signature on the volume between data centers and allows the volume to be clustered across Storage Centers
- Support with asynchronous or synchronous replication and included features such as:
  - Replays
  - High consistency and high availability modes
  - Mode migration
  - DR activation for Live Volume Managed Replications
- Support for an additional asynchronous or synchronous Live Volume replication to a third Storage Center created and dynamically managed by Live Volume
- Improved support for Live Volume recovery in the event of an unplanned outage at a primary Live Volume site

Live Volume is designed to fit into existing physical and virtual environments without disruption or significant changes to existing configurations or workflow. Physical and virtual servers see a consistent, unchanging virtual volume. All volume mapping is consistent and transparent before, during, and after migration. The Live Volume role swap process can be managed automatically or manually and is fully integrated into the Storage Center and Enterprise Manager. Live Volume operates asynchronously or synchronously and is primarily designed for planned migration, resource balancing, and disaster avoidance use cases where both Live Volume Storage Centers are simultaneously available.

A Live Volume can be created between two Dell Compellent Storage Centers residing in the same datacenter or between two well-connected datacenters with Fibre Channel or iSCSI replication connectivity.

Using Dell Compellent Enterprise Manager, a Live Volume and an optional Live Volume managed replication can be created from a new volume, an existing volume, or an existing replication. For more information on creating a Live Volume, see the *Dell Compellent Enterprise Manager User Guide* provided with the Enterprise Manager software.

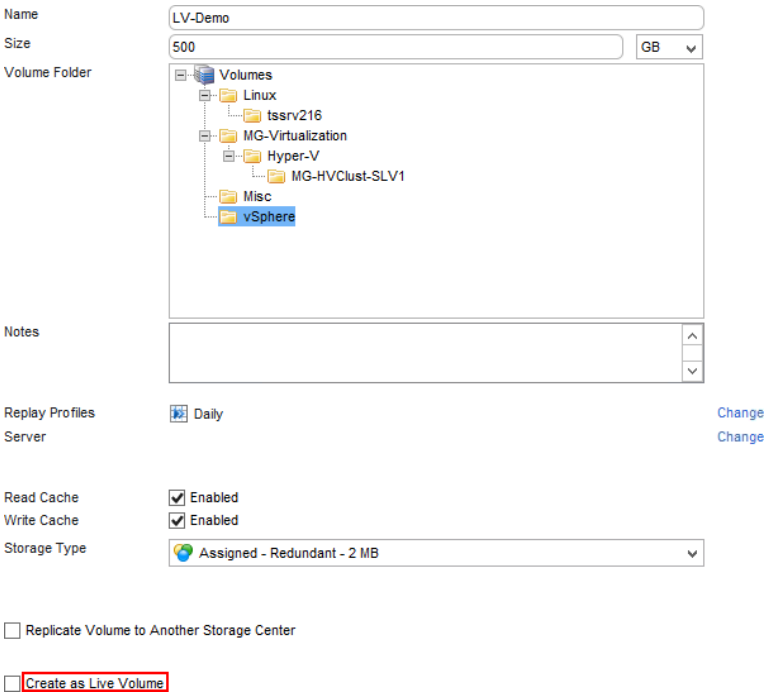

Figure 30 Creating a replica as a Live Volume

5.2 Proxy data access

A Compellent Live Volume is comprised of a pair of replication enabled volumes: a primary Live Volume and a secondary Live Volume. A Live Volume can be accessed through either Storage Center participating in a Live Volume Replication. However, the primary Live Volume role can only be held on one of the two Storage Centers. All read and write activity for a Live Volume occurs on the Storage Center hosting the primary Live Volume. If a server is accessing the Live Volume through the secondary Live Volume Storage Center, data is accessed by proxy over the Fibre Channel or iSCSI replication link to the primary Live Volume system.

In the following figure, a mapped server is accessing a Live Volume by proxy access through the secondary Live Volume system to the primary Live Volume system. This type of proxy data access requires the replication link between the two Storage Centers to have enough bandwidth and minimum latency to support the I/O operations and latency requirements of the application data access.

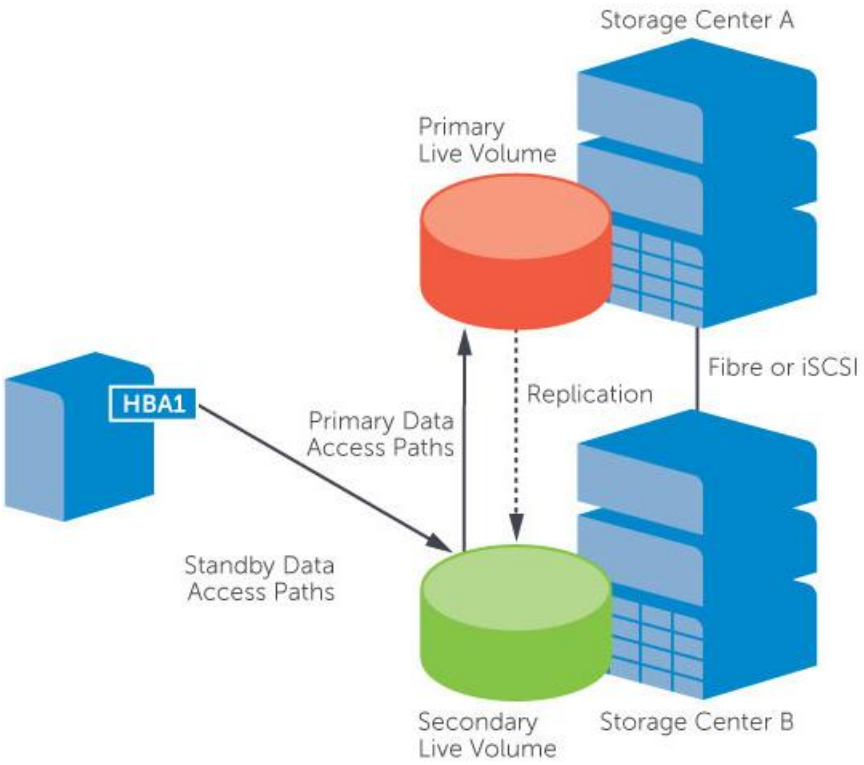

Figure 31 Proxy data access through the Secondary Live Volume

5.3 Live Volume requirements

Live Volume requirements may vary depending on intended use. For example, there are different requirements for using Live Volume to migrate a workload depending on whether or not the virtual machines are powered on during the migration.

5.3.1 Connectivity

From the Compellent Live Volume perspective, there are no firm restrictions on bandwidth or latency. However, to proxy data access from one Storage Center to another requires the Live Volume Storage Centers to be connected through a high bandwidth/low latency replication link. Some operating systems and applications require disk latency under 10ms for optimal performance. However, performance impact may not be effectively realized until disk latency reaches 25ms or greater. Some applications are more latency sensitive. This means that if average latency in the primary data center to the storage is 5ms for the volume and the connection between the two data centers averages 30ms of latency, the storage latency writing data to the primary Live Volume from the secondary Live Volume proxy across the link is probably going to be 35ms or greater. While this may be tolerable for some applications, it may not be tolerable for others.

If the Live Volume proxy communication is utilized, it is strongly recommended to use fiber connectivity between the sites to ensure consistent bandwidth and the least amount of latency. The amount of

bandwidth required for the connectivity is highly dependent on the amount of changed data that requires replication, as well as the amount of other traffic on the same wire. If a site is not planning to proxy data access between Storage Centers, then latency is not a concern.

It is recommended to have separate VLANs or fabrics to isolate storage traffic from other types of general-purpose LAN traffic, especially when spanning data centers. While this is not a requirement for Live Volume, it is a general best practice for storage.

For hypervisor virtualization products such as VMware vSphere, Microsoft Hyper-V, and Citrix XenServer, a site must have at least a 1GB connection with 10ms or less latency between servers to support vMotion Metro or live migration activities. Standard VMware vMotion requires 5ms or less latency between source and destination host.

High bandwidth, low latency

For an inter-datacenter, campus environment (or within a 60-mile radius) high-speed fibre connectivity is possible. While inter-datacenter and campus environments may be able to run fiber speeds of up to 16Gb using Multi Mode fiber connectivity, Single Mode fiber connectivity of up to 1Gb using dark fibre can assist with connecting datacenters together that may be up to 60 miles apart as an example. Minimal latency is especially key when implementing Live Volume on top of synchronous replication (in either mode). This type of connectivity is highly recommended for live migrating virtual machine workloads between Compellent Storage Centers.

Low bandwidth, high latency

If a site is planning on using Live Volume on a low bandwidth/high latency replication link, it is recommended to control swap role activities manually by shutting down the application running at site A, perform a Live Volume swap role, and then bring the application up at the remote site. This scenario prevents any storage proxy traffic from going across the link, as well as providing a pause in replication I/O for the link allowing the replication to catch up so that a Live Volume swap role can occur. Manual swap role activities can be controlled by deselecting the Automatically Swap Roles option on the Live Volume configuration. If the replication connection has characteristics of high latency, asynchronous replication is recommended for Live Volume so that applications are not adversely impacted by replication latency.

5.4 Replication and Live Volume attributes

Once a Live Volume is created, additional attributes can be viewed and modified by editing the replication properties of the Live Volume. To modify the Live Volume settings, select **Replication & Live Volumes** from Enterprise Manager, and then select **Live Volumes** as depicted in Figure 32.

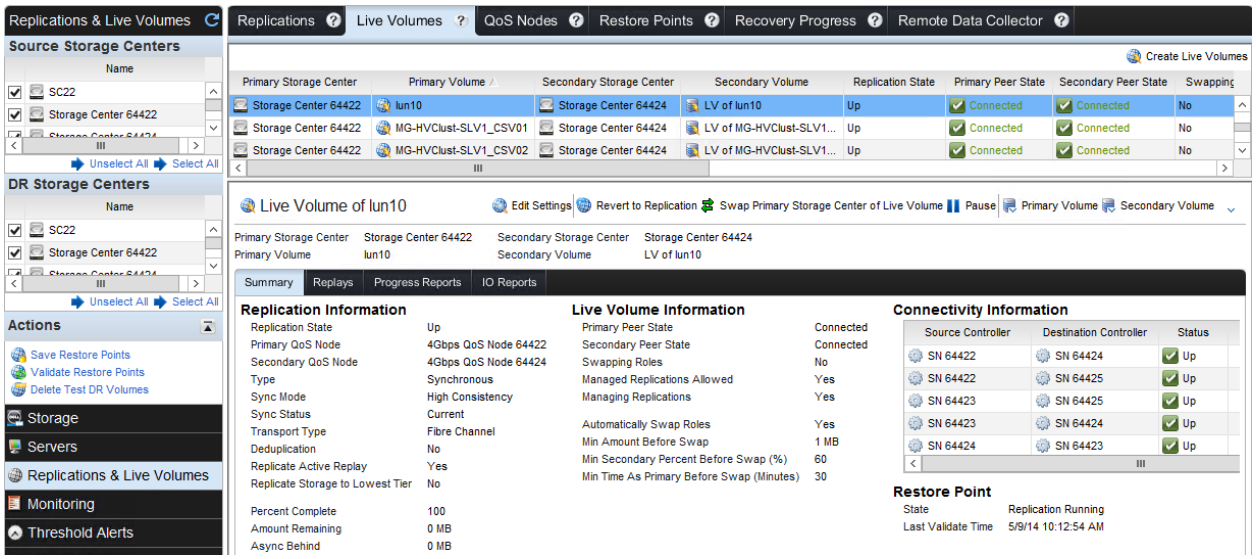

Figure 32 Live Volume settings

5.4.1 Replication Information

Live Volume is built on standard Compellent replicated volumes where that replication may be asynchronous, synchronous high availability, or synchronous high consistency. More information about these attributes can be found in earlier sections of this document and in the *Enterprise Manager Administrator's Guide*.

Type

Type refers to asynchronous or synchronous replication.

Sync Mode

If the replication type is synchronous, Sync Mode describes the mode of synchronous replication that may be either high consistency or high availability. Sync Mode is not displayed for asynchronous Live Volumes.

Sync Status

If the replication type is synchronous, Sync Status describes the current state of the synchronous replication as **Current** or **Out Of Date**. When Sync Status is Current, synchronous replication is in sync. This means that both the source and destination volumes are consistent and the cumulative latency to the secondary Live Volume will be observed at the primary Live Volume application. An out-of-date status indicates that the data on the source and destination volumes is not consistent. Changed or inconsistent

data is tracked in a journal where the primary Live Volume resides until the two volumes are once again Current. Sync Status is not displayed for asynchronous Live Volumes.

Deduplication

The deduplication feature replicates only the changed portions of the Replay history on the source volume, rather than all data captured in each Replay. While this is a more processor-intensive activity, it may reduce the amount of replication traffic and bandwidth required. If sufficient bandwidth is present on the connection, Compellent recommends that disabling deduplication for Live Volumes in order to preserve controller CPU time for other processes.

Replicate Active Replay

It is recommend that Replicate Active Replay is enabled for asynchronous Live Volumes. This ensures that data is replicated as fast as possible which decreases the amount of time required to perform a Live Volume Swap role. This feature is automatically enabled with synchronous Live Volumes.

Replicate Storage to Lower Tier

The Replicate Storage to Lowest Tier feature is automatically enabled for a new Live Volume. Disable this option to replicate data to Tier 1 on the destination Storage Center. Many users perform the initial Live Volume replication to the Lowest Tier, and then deselect this option once the initial replication completes. This strategy aids in preserving Tier 1 storage capacity, which is useful when using 15k spindles or SSD drives. For more information on Data Progression with Live Volume, see the section titled, "Data Progression and Live Volume".

QoS Nodes

A pair of QoS Nodes depicts the desired egress traffic shaping to be applied when replicating from the primary to the secondary Live Volume. Although labeled a secondary QoS Node, it does not provide ingress traffic shaping. Instead, it provides egress traffic shaping after a role swap occurs and it becomes the primary Live Volume. QoS Nodes apply to replication traffic only. The Live Volume proxy traffic between the Storage Centers is not governed by QoS Nodes. If the link between the Live Volume Storage Controllers is shared by other traffic, it may be necessary to throttle the replication traffic using QoS Nodes to prevent the it from flooding the replication link. However, throttling synchronous replication traffic will produce application latency at the primary Live Volume. For this reason, replication links and QoS Nodes should be sized appropriately taking into account the amount of data per Live Volume, rate of change, application latency requirements, and any other applications or services that may be sharing the replication link. This is especially important when using synchronous replication.

For instance, if an 8Gbps replication link exists between datacenters that is shared by all intra-data center traffic, a replication QoS could be set at 4Gbps and thereby limits the amount of bandwidth used by replication traffic to half of the pipe capacity. This allows the other non-Live Volume replication traffic to receive a reasonable share of the replication pipe, but could cause application latency if synchronous replication traffic exceeds 400MBps.

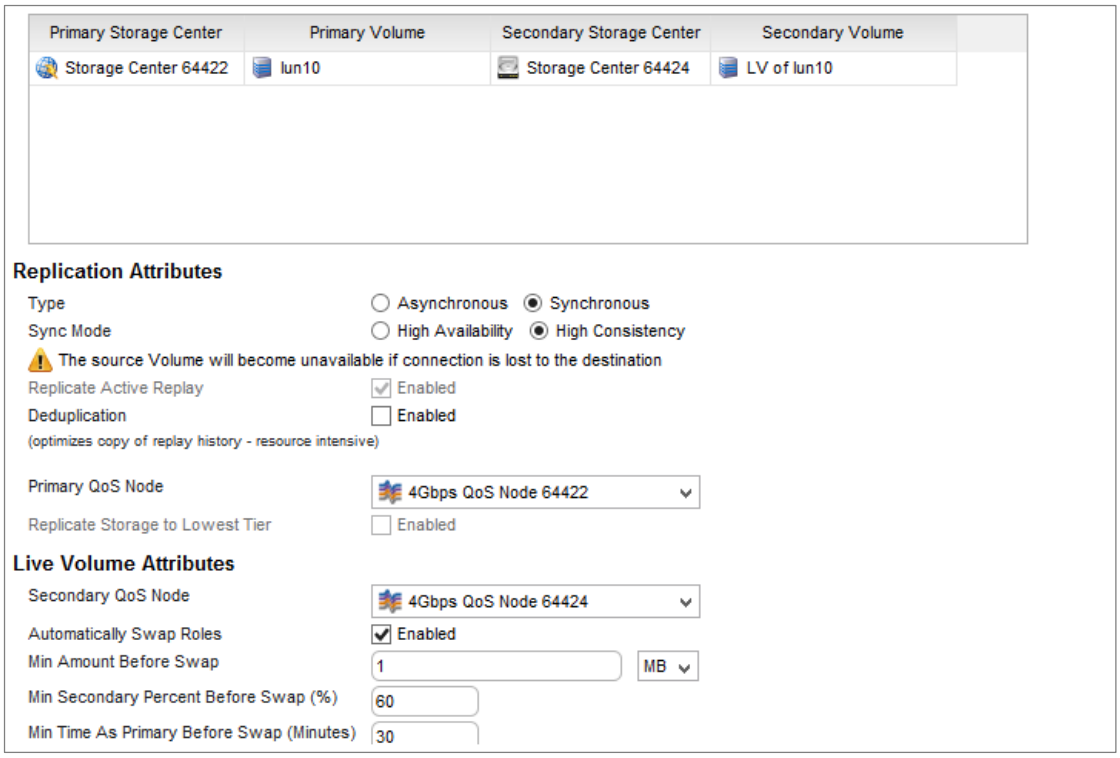

Figure 33 Creating a Live Volume with QoS nodes

As a best practice, common QoS Nodes should not be shared between a single Storage Center source and multiple Storage Center destinations, particularly where Live Volume Managed Replications are in use. For instance, if Storage Center A is replicating to Storage Center B synchronously to support a Live Volume and Storage Center A is also replicating to Storage Center C asynchronously to support a Live Volume Managed Replication, an independent QoS Node should be created and used for each of these replications.

5.4.2 Live Volume information

A Live Volume provides additional attributes that control the behavior of the Live Volume. Those attributes are listed below.

Automatically Swap Roles

When Automatically Swap Roles is selected, the Live Volume will be automatically swapped to the Storage Center with the most I/O load as long as it meets the conditions for a swap. The Live Volume logic gathers I/O samples to determine the primary access to the Live Volume (from either servers accessing it directly on the primary Live Volume Storage Center, or servers accessing from a secondary Live Volume Storage Center). Samples are taken every 30 seconds and Storage Center makes automated role swap decisions based on the last ten samples (five minutes worth). This occurs continuously on the primary Live Volume Storage Center (it does not start once the 30-minute delay timer expires).

The autoswap design is meant to make intelligent decisions on the autoswap movement of Live Volume primary systems while preventing role swap movement from occurring rapidly back and forth between arrays.

Min Amount Before Swap

The first aspect is the amount of data accessed from a secondary system. If there is light, infrequent access to a Live Volume from a secondary Storage Center, does it make sense to move the primary to that system? If so, set this value to a very small value. The criteria for this aspect are defined by the Min Amount Before Swap attribute for the Live Volume. The value specifies an amount of (read/write access)/second per sample value. If a sample shows that the secondary Storage Center access exceeds this value, this sample/aspect has been satisfied.

Min Secondary Percent Before Swap

The second aspect is the percentage of total access of a Live Volume from the secondary Storage Center on a per sample basis. The criteria for this aspect are defined by the Min Secondary % for Swap attribute for the Live Volume. If a sample shows the secondary Storage Center accessed the Live Volume more than the defined setting for this aspect, this sample/aspect has been satisfied. The default setting for this option is 60%. Compellent takes samples every 30 seconds and keeps the most recent ten samples (five minutes worth) for analysis. This means that the secondary Live Volume has to have more I/O than the primary system for six out of ten samples (60%).

Min Time As Primary Before Swap

Each Live Volume has a TimeAsPrimary (default setting of 30 minutes) timer that will prohibit an autoswap from occurring after a role swap has completed. This means that following a role swap of a Live Volume (either auto or user specified), Storage Center waits the specified amount of time before analyzing data for autoswap conditions to be met again. The purpose of this is to prevent thrashing of autoswap in environments where the primary access point could be dynamic or when a Live Volume is shared by applications that can be running on servers both at the primary and secondary sites.

6 Data Progression and Live Volume

Data Progression life cycles are managed independently on each Storage Center involved with a Live Volume. If the Live Volume is not being replicated to the lowest tier on the destination Storage Center, it will follow the volume based storage profile and Data Progression lifecycle on the destination Storage Center.

6.1 Primary and secondary Live Volume

If a Live Volume is typically always primary on Storage Center A and is not being replicated to the lowest tier on the destination Storage Center B, the data will progress down on the destination Storage Center to the next tier or RAID level every 12 days. This is because the data on the destination Storage Center is never actually being read. All reads take place on the primary Live Volume Storage Center.

For instance, if a Storage Center has two tiers of disk, 15k and 7k, and the Storage Profiles write data at RAID 10 on Tier 1, Replay Data at Tier 1 is at RAID 5, and Tier 3 is at RAID 5. The blocks of data that were written during the day will progress from Tier 1, RAID 10 to Tier 1 RAID 5 on the first night.

If a Live Volume is frequently swapped as primary between the Live Volume Storage Centers, then the Data Progression pattern will be determined by how often the data is accessed on both systems.

7 Live Volume and MPIO

By using Live Volume with a storage host that has access to both Storage Center arrays in a Live Volume configuration, multiple paths can be presented to the server through each Storage Center controller. Live Volume data access from the secondary system is proxied to the primary Storage Center. For this reason, special consideration should be taken to control the I/O path for the Live Volume.

7.1 MPIO policies for Live Volume

For Live Volume Storage Centers where a server has access to both the primary and secondary Live Volume controllers, set the MPIO policy to one that prevents primary data access through the secondary Live Volume Storage Center when possible. These types of MPIO policies are typically **Failover Only, Fixed**.

Additional information on configuring Live Volume MPIO can be found in each of the sections of this document devoted to a specific application, such as VMware vSphere, Windows/Hyper-V, Linux, Solaris, and AIX.

8 VMware vSphere and Live Volume

VMware vSphere and Live Volume can combine to provide larger scale data mobility, planned maintenance, resource balancing, disaster avoidance, and disaster recovery options for virtual environments.

8.1 MPIO

vSphere ships with three MPIO policies: **Fixed**, **Most Recently Used**, and **Round Robin**. When mapping a Live Volume through both the primary and secondary Storage Centers to a vSphere host (for instance, when all three are in the same datacenter), the MPIO policy on the host should be set to **Fixed** with the preferred path set to the primary Live Volume Storage Center controller.

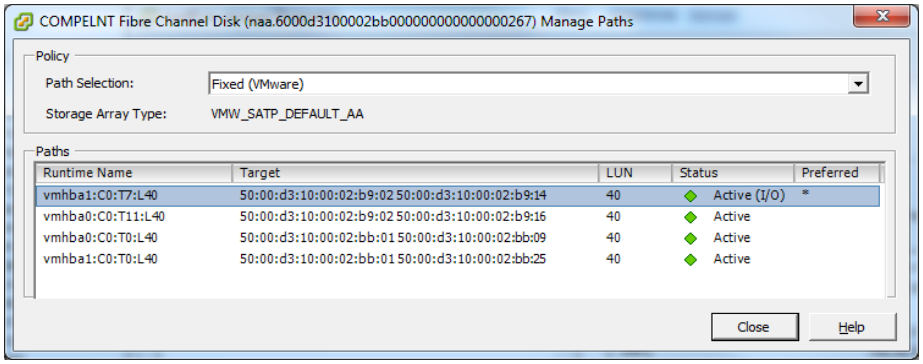

Figure 34 Fixed policy

The following figure depicts a Round Robin policy on a Compellent Live Volume going between two Storage Centers. This configuration is not optimal because 50% of the I/O traffic will have to traverse the Live Volume Replication proxy link.

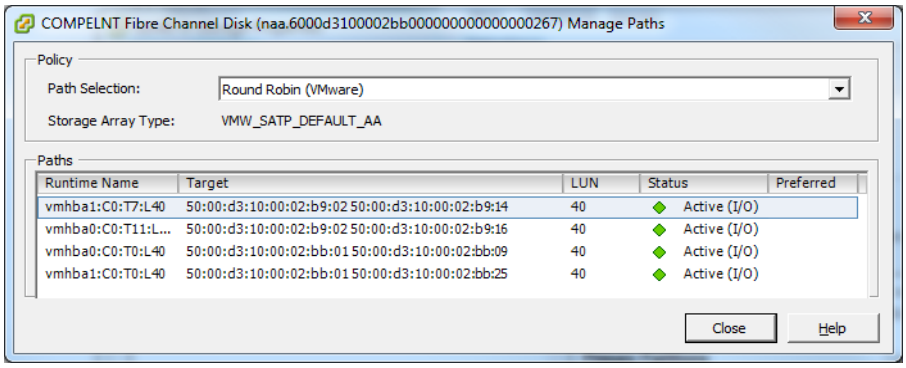

Figure 35 Round Robin policy

8.2 Single site MPIO configuration

In a single site configuration, multiple vSphere hosts may be zoned to both Storage Center arrays. If a Live Volume is mapped over both Storage Centers to the vSphere hosts, then the volume can participate in an MPIO configuration involving both Storage Centers. In this scenario, use vSphere's Fixed MPIO policy to ensure traffic is always going to the primary Live Volume Storage Center. The preferred path is always used in this policy unless the primary path fails and the Fixed policy will migrate I/O to one of the other available paths in the policy.

As depicted in Figure 36, a Live Volume replication exists between Storage Center A and Storage Center B. Two VMware vSphere hosts are connected and mapped to the Live Volume on each controller. The Primary Live Volume is located on Storage Center A, so the Fixed Preferred path (Figure 35) on each vSphere host is configured to use a path to Storage Center A as the preferred path.

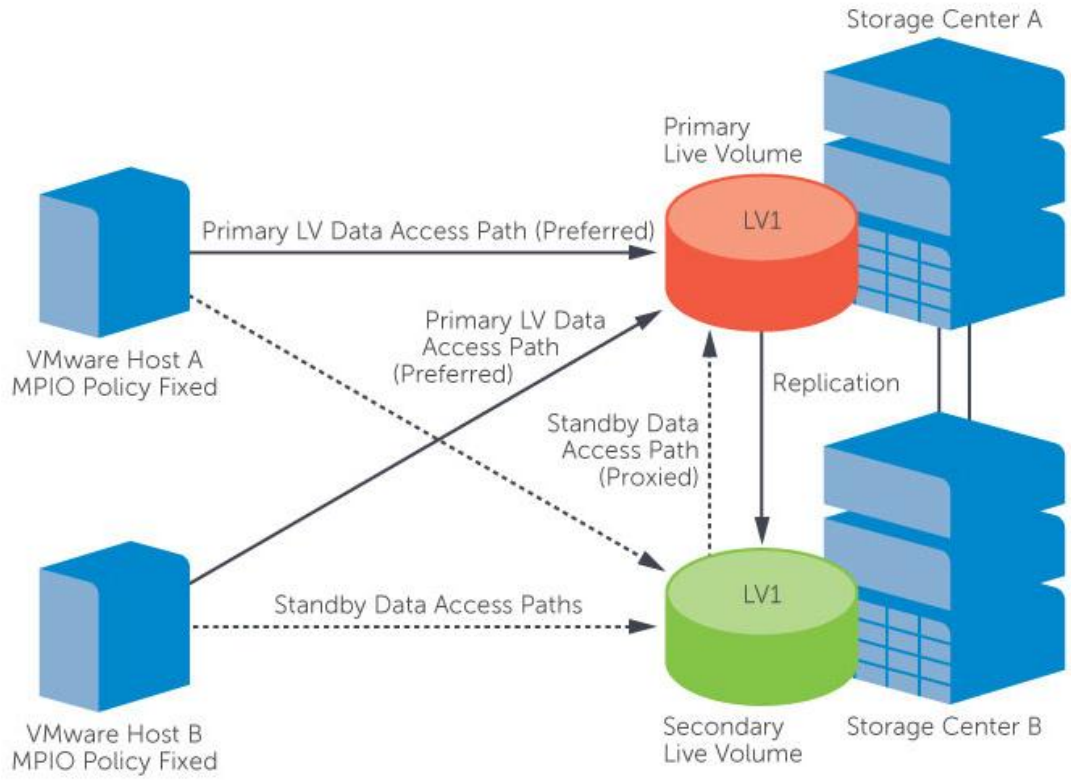

Figure 36 Live Volume replication

If maintenance is required on Storage Center A, for example, the preferred path for both vSphere hosts could be configured for Storage Center B. When the Live Volume is configured for automatic role swap, this change in preferred paths will trigger the Live Volume to swap roles making Storage Center B the Primary Live Volume controller, so that Storage Center A can be taken offline without a disruption to virtual machines on that Live Volume.

8.3 Multi-site MPIO configuration

In a multi-site or stretched configuration, the vSphere hosts are typically mapped only to their local corresponding Storage Center (see Figure 37).

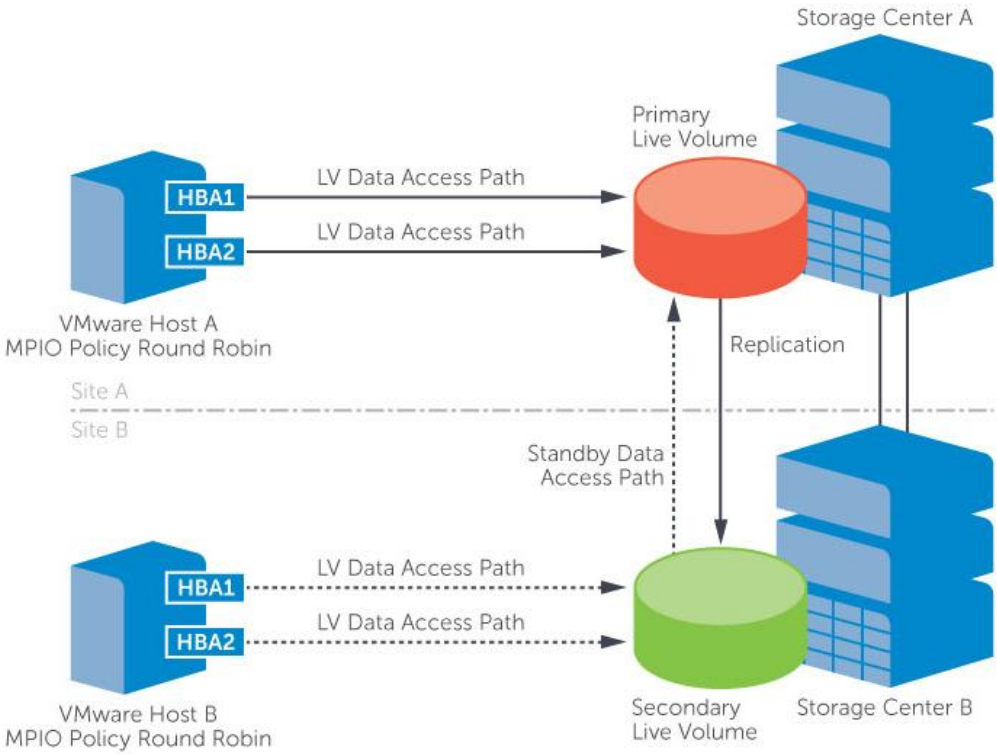

Figure 37 Multi-site MPIO configuration

In this configuration, the MPIO policy can be set to Round Robin between paths to the local Live Volume as the mappings do not include multiple Storage Centers within the same datacenter. All inter-site volume access from the secondary Live Volume is proxied to the Primary Live Volume controller via the asynchronous or synchronous replication link(s) between the Storage Centers.

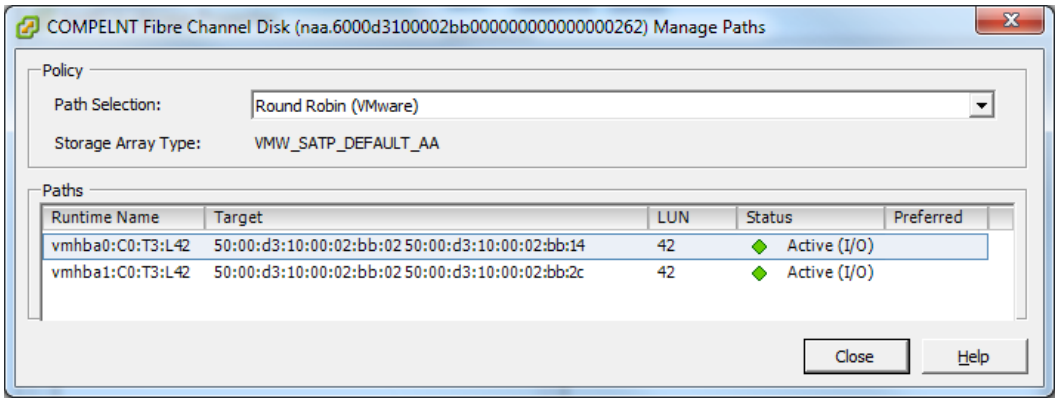

Figure 38 Round Robin to the local Live Volume

8.4 VMware vMotion and Live Volume

Another way of controlling which Storage Center is primary for the Live Volume is by migrating one or more virtual machines from one host to another using vMotion. In this scenario, vSphere host A would be mapped to Storage Center A and vSphere host B would be mapped to Storage Center B. When a virtual machine running on a Live Volume is migrated from host A to host B, the Live Volume will observe that the storage is being accessed through Storage Center B rather than Storage Center A and can automatically swap the Secondary Live Volume to become the Primary Live Volume. Once this occurs, the virtual machine disk I/O will traverse a local path to the Primary Live Volume on Storage Center B instead of going through the Secondary Live Volume proxy across the replication link. The result is evading an inevitably higher cost in terms of increased latency. For stretched clusters or datacenters, consider the vMotion and Metro vMotion latency requirements between hosts. vMotion requires round-trip latency of 5ms or less between hosts on the vMotion network. Metro vMotion, an Enterprise Plus feature introduced in vSphere 5, provides better vMotion performance over increasing distances and latency, and increases the allowable round-trip latency from 5ms to 10ms on the vMotion network.

8.5 VMware DRS/HA and Live Volume

VMware DRS technology uses vMotion to automatically move virtual machines to other nodes in a cluster. Unfortunately, vSphere clusters do not natively possess site awareness capabilities. In a multi-site Live Volume VMware cluster, it is a best practice to keep virtual machines running on the same site as their primary Live Volume. Additionally, it is best to keep virtual machines that share a common Live Volume enabled datastore together at the same site. This ensures that compute and storage resources remain local to the vSphere host running the virtual machine(s). If DRS is activated on a vSphere Cluster with nodes in each site, DRS could automatically move some of the virtual machines running on a Live Volume datastore to a host that resides in the other data center. In vSphere 4.1 and later, DRS Host Groups and VM Groups can be used in a few ways to benefit a multi-site Live Volume environment. Virtual machines that share a common Live Volume datastore can be placed into VM Groups. Movement of virtual machines and management of their respective Live Volume datastore can then be performed at a containerized group level rather than at an individual virtual machine level. Hosts that share a common site can be placed into Host Groups. Once the host groups are configured, they can represent locality for the primary Live Volume. At this point VM groups can be assigned to host groups using the DRS Groups Manager. This will ensure all virtual machines that share a common Live Volume datastore are consistently running from the same datastore. The virtual machines can be vMotioned as a group from one site to another. After a polling threshold is met, Storage Center will swap roles with the Live Volume enabled datastore to the site where the VMs were migrated.

The infrastructure can be designed in such a way where separate DRS enabled clusters exist at both sites keeping automatic migration of virtual machines within the respective site where the Primary Live Volume resides. In the event of a Live Volume role swap, all virtual machines associated with the Live Volume can be vMotioned from the Site A cluster to the Site B cluster as long as both clusters fall under the same datacenter object in vCenter. HA is a cluster centric configuration and operation. In this design, in the event of a host failure, HA will attempt to restart virtual machines only within the same cluster. This means that the VMs will always attempt start up at the same site they failed in. Virtual machines will not attempt to restart at the remote site. In the event of a host outage or isolation, vSphere HA will be able to restart

virtual machines when the primary Live Volume is available and either a direct mapping to the host or an indirect mapping to the host through the secondary Live Volume proxy exists. However, if an outage impacts primary Live Volume availability and only the secondary Live Volume remains available, HA will not be able to immediately restart virtual machines on the secondary Live Volume. In this case, the Preserve Live Volume feature should be used to promote the Live Volume from a secondary role to a primary role. vSphere hosts mapped to this volume will now be able to power on virtual machines residing on this volume as long as all other virtual machine power on conditions have been met. The following vSphere advanced tuning should also be configured for non-uniform stretched cluster configurations to allow HA to power off and migrate virtual machines during storage related availability events after the primary Live Volume has been made available again:

- vSphere 5.0u1+/5.1: Disk.terminateVMOnPDLDefault = True (/etc/vmware/settings on each host)
  or
- vSphere 5.5: VMkernel.Boot.terminateVMOnPDL = yes (advanced setting on each host)
  and
- vSphere 5.0u1+: das.maskCleanShutdownEnabled = True (Cluster advanced options)
  and
- vSphere 5.5: Disk.AutoremoveOnPDL = 0 (advanced setting on each host)

If the VMware virtual infrastructure is version 4.0 or earlier, other steps should be taken to prevent virtual machines from unexpectedly running from the secondary Live Volume. An individual VM or group of VMs may be associated with a DRS rule that keeps them together but this does not guarantee they will stay on the same host or group of hosts over a period of time where the primary Live Volume is located. As a last resort, DRS can be configured for manual mode or disabled when using Live Volume in a multi-site configuration that will prevent the automatic migration of VMs to Secondary Live Volume hosts in the same cluster.

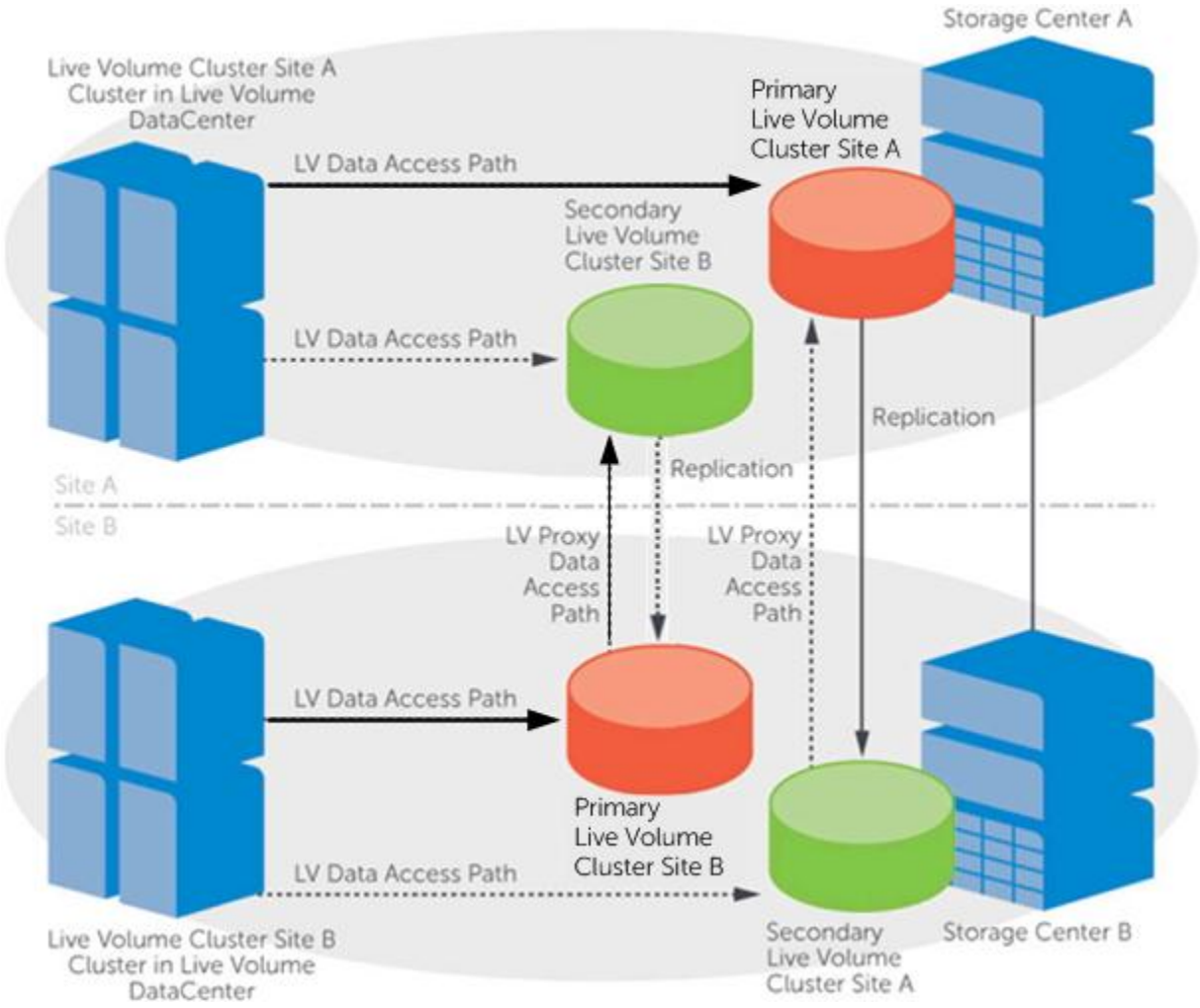

Figure 39 Live Volume data access in a vSphere environment

One particular design of a multi-site vSphere environment exhibited in Figure 40 may be to create a single vSphere datacenter object, then create a vSphere cluster for each physical site within that datacenter. In this design, each site can benefit from VMware DRS and HA features and virtual machines will only migrate within that site or cluster. Since all of the cluster nodes are in the same vSphere datacenter, the virtual machines can be manually moved using vMotion between the clusters. This provides a blend of flexibility and mobility. Virtual machines residing in clusters separated by datacenters cannot leverage vMotion. In this case, virtual machines must be powered off to migrate between datacenters.

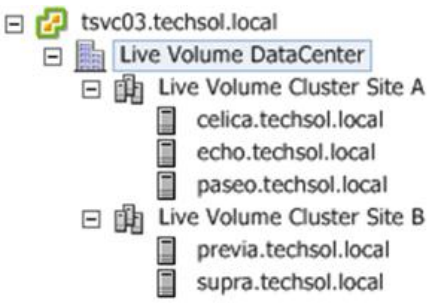

Figure 40 Live Volume cluster design options

8.6 VMware and Live Volume managed replication

Live Volume has historically met the needs of specific high availability use cases between two Storage Centers. With the release of Storage Center 6.5, a third Storage Center can be added to the site design to host a Live Volume Managed Replication. A Managed Replication is a replica of the primary Live Volume and may be either synchronous or asynchronous depending on the type of replication used between the Live Volume pair. While Live Volume provides high availability, the Managed Replication provides an additional level of data protection and recovery in the event of an outage at the site(s) where the primary and secondary Live Volumes reside.

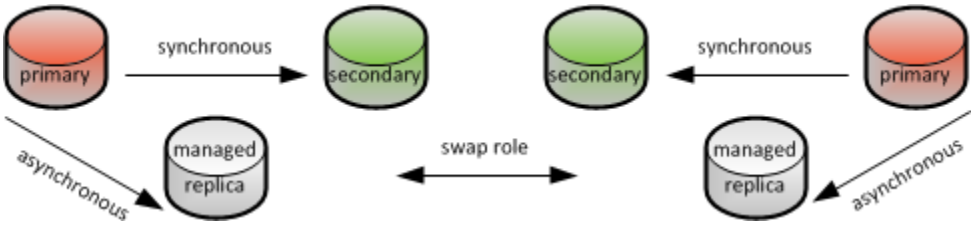

Figure 41 Synchronous Live Volume with asynchronous Live Volume managed replication

Although managed replications are controlled by Live Volume, they fundamentally work the same way as standard synchronous or asynchronous volume replication. This includes the recovery options available with View volumes and DR activation.

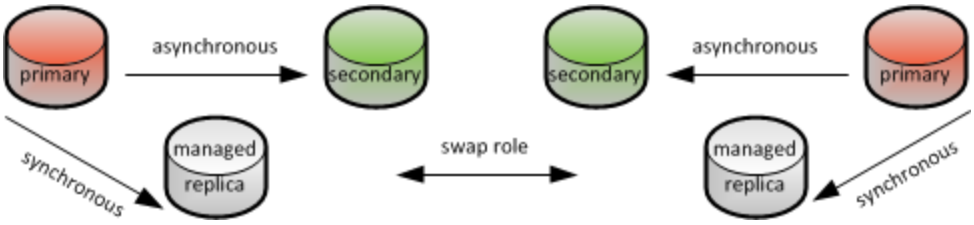

Figure 42 Asynchronous Live Volume with Synchronous Live Volume managed replication

One differentiator between standard replication and Live Volume managed replication is that the source volume of a managed replication is dynamic and changes as Live Volume role swaps occur. The easiest way to think of this is to understand that the managed replication source is always the primary Live Volume. Because the primary Live Volume role can shift manually or automatically between Storage

Centers, the flow of replicated data will also seamlessly and automatically follow this pattern. The reason for this is that the primary Live Volume is the volume where write I/O is first committed. Therefore, especially in an asynchronous Live Volume configuration, in the event of an unplanned outage or disaster that disables both Live Volumes, data should be recovered from the most transaction consistent volume to minimize loss of data. For synchronous Live Volumes, transaction inconsistency is less of an issue because both primary and secondary Live Volumes should have a sync status of current and not out of date unless the synchronous replication was paused or became out of date due to excess latency in high availability mode. Disaster Recovery at a remote site is a good use case for the Managed Replication. However, Live Volume and Live Volume Managed Replications are not currently supported with Site Recovery Manager. LUN presentation and recover of data on a managed replication can be handled automatically by Enterprise Manager but before virtual machines can be powered on, they must first be registered into vSphere inventory, which may be a manually followed DR documentation or scripted process.

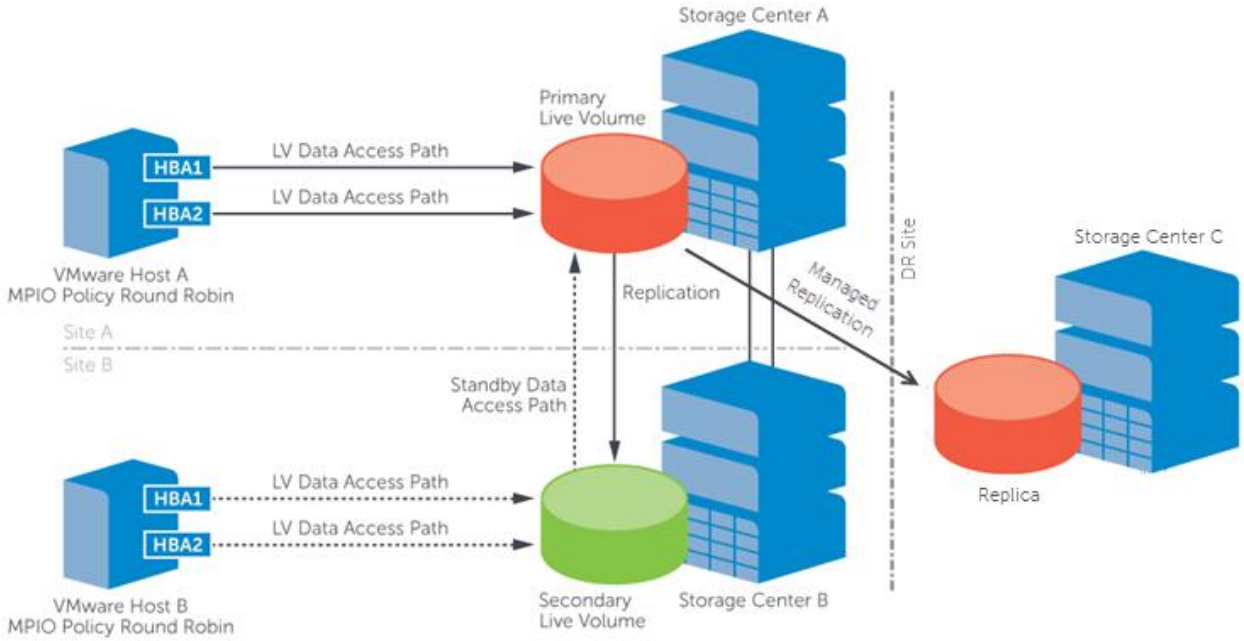

Figure 43 Live Volume pair with Managed Replication at a third site

9 Microsoft Windows MPIO

Microsoft Windows servers running on Dell Compellent storage can use the in-box Microsoft MPIO DSM. The Microsoft Windows 2008 R2 and Server2012/R2 MPIO DSM comes with the following MPIO policies: **Failover Only, Round Robin, Round Robin with Subset, Least Queue Depth, Weighted Paths**, and **Least Blocks**. The two most common Microsoft Windows MPIO policies to control Live Volume access are **Round Robin with Subset** and **Failover Only**. Both of these policies allow you to define active and standby paths. When accessing a volume being proxied from the secondary Live Volume to the primary Live Volume, it adds a little extra latency and added traffic to the replication/Live Volumes links. If using a **Round Robin** policy that contains the primary and secondary Live Volume Storage Centers, the Live Volume will never auto swap because half of the I/O will always be going through each controller. For the best performance in your environment, use an MPIO Policy of **Round Robin with Subset** or **Failover Only** for Microsoft Windows hosts.

Note: For more information about Windows Server and MPIO, consult the *Windows Server MPIO Best Practices for Dell Compellent Storage Center* guide located on Dell TechCenter at http://en.community.dell.com/techcenter/extras/m/white_papers/20437917.aspx.

9.1 Round Robin with Subset

The **Round Robin with Subset** policy uses paths from a primary pool of paths for processing requests as long as at least one of the paths is available. This DSM uses a standby path only when all the primary paths fail.

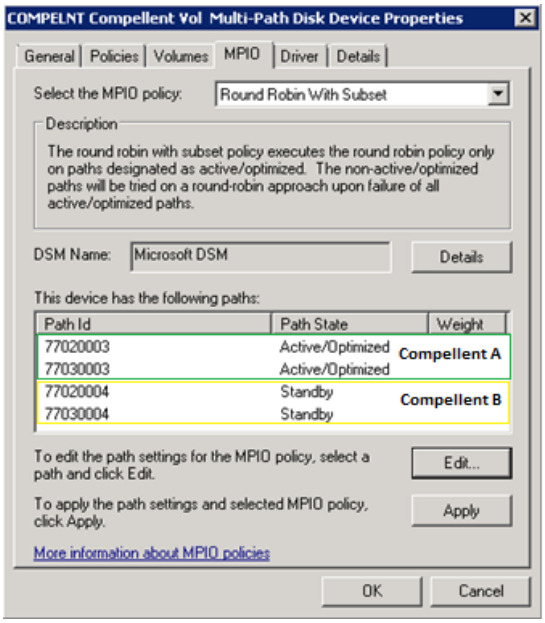

Figure 44 Round Robin with Subset policy properties

By using the **Round Robin with Subset** policy, Round Robin functionality to the primary Live Volume controller is maintained and the secondary Live Volume controller can be used as the failover path. These paths can be changed at any time on the host by modifying and applying the configuration (see Figure 44).

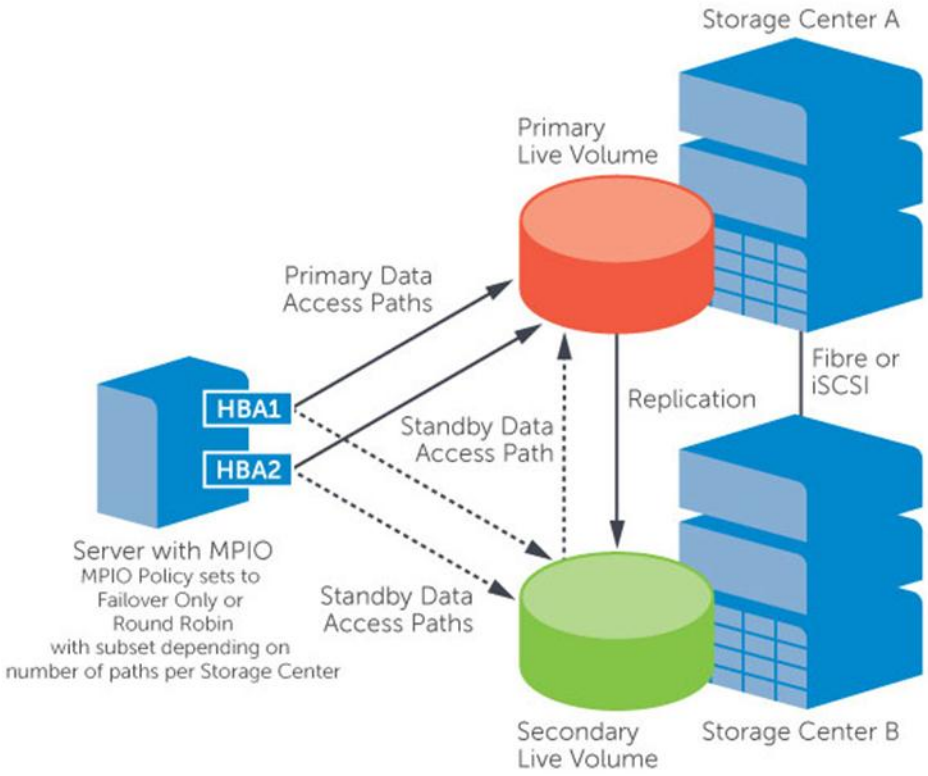

Figure 45 Round Robin with Subset MPIO policy

9.2 Failover Only

An option that works best with servers containing only one HBA or if the I/O load between paths does not need round Robin, is the MPIO policy of Failover Only. By using the Failover Only MPIO policy, the primary path is defined and all other paths are set to standby as seen in Figure 46.

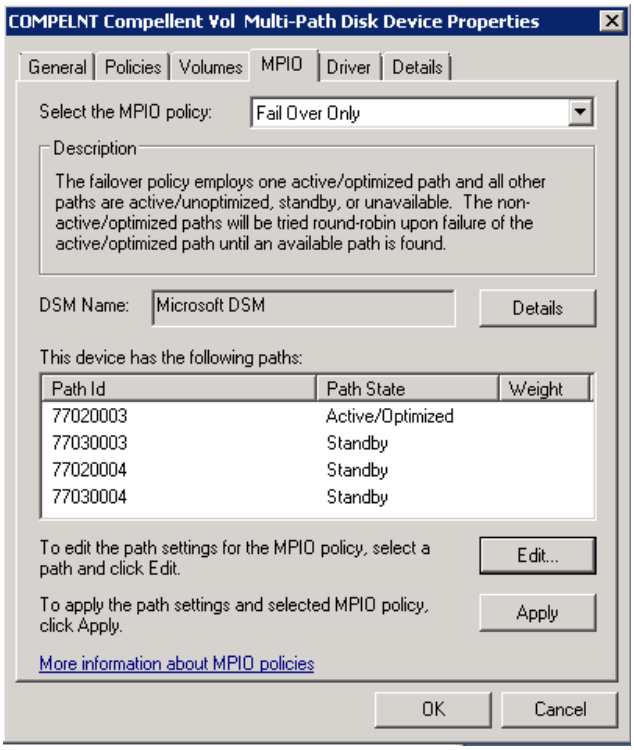

Figure 46 Fail Over Only policy properties

9.3 Sub-optimal MPIO

In Figure 47, a sub-optimal Live Volume MPIO configuration is depicted. In this scenario, the server has two adapters and is mapped to two different Storage Centers. Since all four paths are included in an MPIO Round Robin policy, about half of the storage traffic would have to traverse the proxy link between the two Storage Centers. This configuration also prevents automatically swapping roles because 50% of the traffic will always be going through each Storage Center and preventing an autoswap of the Live Volume roles.

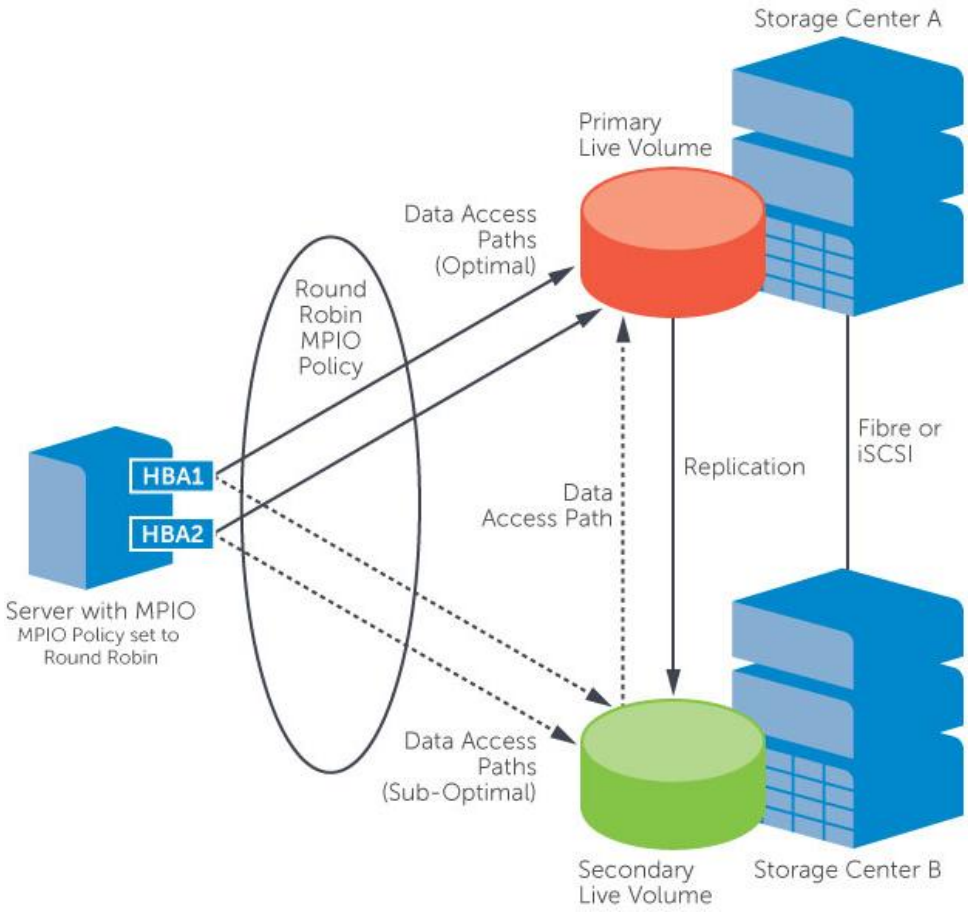

Figure 47 Sub-optimal MPIO traffic pattern

9.4 Hyper-V and Live Volume

Live Volume works well with Microsoft Hyper-V in both clustered and non-clustered scenarios.

9.5 Stand alone Hyper-V

In a non-clustered scenario, MPIO can be used to control which Storage Center is providing access to the data. It is recommended to use either the **Round Robin with Subset** or **Failover Only** MPIO policies. See the Microsoft Windows MPIO sections of this document for more information.

9.6 Clustering Hyper-V

With clustered Hyper-V servers on Microsoft Windows 2008 R2 and 2012/R2, virtual machines can be migrated from one host in a cluster to another using Live Volume. In this scenario, Node A could be mapped to Storage Center A and Node B could be mapped to Storage Center B. Therefore, when a virtual machine is migrated from Node A to Node B, the Live Volume will automatically perform a swap role making Storage Center B the primary. This configuration is most common in a multi-site cluster.

9.6.1 Single site

In a single site configuration, multiple Hyper-V servers can be connected to both Storage Centers. If a Live Volume is mapped over both Storage Centers to the Hyper-V servers, then the Live Volume can participate in an MPIO configuration. In this scenario, use a Windows **Round Robin with Subset** or **Failover** policy to ensure data access traffic is always going to the primary Live Volume Storage Center.

As depicted in the Figure 48, a Live Volume exists between Storage Center A and Storage Center B. Two Hyper-V servers are connected and mapped to the Live Volume on each controller. The primary Live Volume is running on Storage Center A, so either the **Round Robin with Subset** or **Failover Only** active path on each Hyper-V host is set to use a connection to Storage Center A as the preferred path for the said Live Volume.

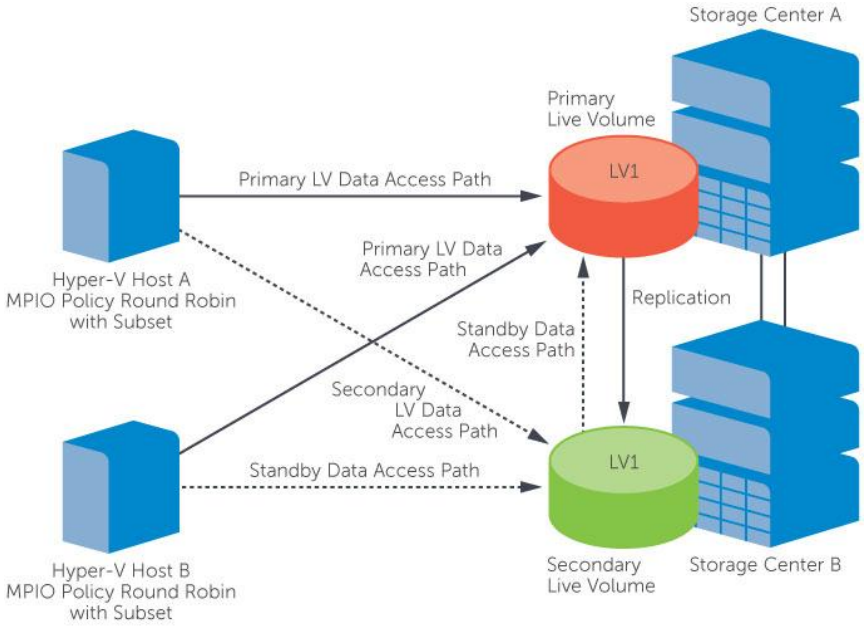

Figure 48 Single site MPIO configuration

9.6.2 Multi-site

In a multi-site configuration, typically the Hyper-V hosts are mapped only to the Storage Center in the particular site. In this scenario, the MPIO policy can be set to Round Robin for Hyper-V hosts. Virtual machine placement determines which Storage Center will host the Primary Live Volume. The scenario in Figure 49 depicts a virtual machine migrated from Host A to Host B. Storage Center B will see the primary Access for the Live Volume going through Storage Center B and will automatically swap the roles so that the Storage Center B becomes primary for the Live Volume.

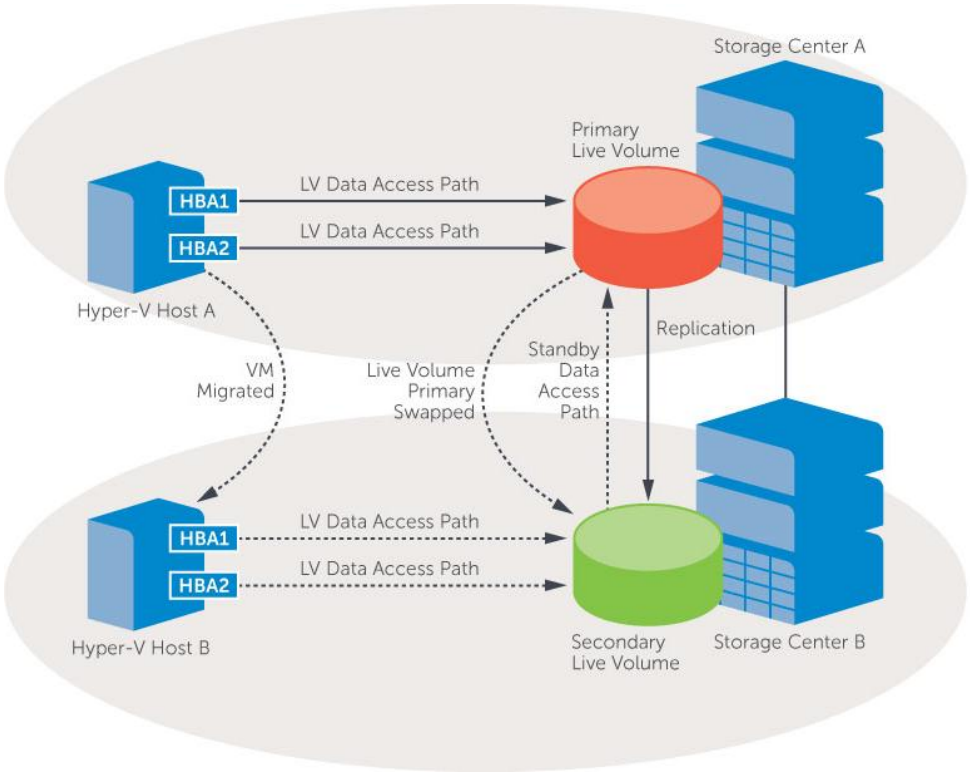

Figure 49 Virtual machine migrated from Host A to Host B

9.7 SCVMM/SCOM and Performance and Resource Optimization (PRO)

System Center Virtual Machine Manager with System Center Configuration Manager is capable of providing intelligent placement as well as automatic migrations of virtual machines from highly utilized nodes to lower utilized nodes, depending on the action setting of **Automatic** or **Manual**. If using Live Volume in a multi-site Hyper-V cluster with PRO, use the **Manual** action for virtual machine placement.

In a multi-site Live Volume Hyper-V cluster, keep the virtual machines running in the same site as their Primary Live Volume. If PRO is activated on a Hyper-V Cluster with nodes in each site, PRO could automatically migrate some of the virtual machines running on a Live Volume CSV to a server that resides in the other data center thereby splitting the I/O between data centers.

9.8 Live Volume and Cluster Shared Volumes

Hyper-V has a feature called Cluster Shared Volume (CSV) that allows administrators to place multiple virtual machines on a cluster volume. CSVs also have a feature called Network Redirection that by design makes Hyper-V cluster data access a little more fault tolerant. If the CSV is owned by a node of the cluster that has access to the volume, it can redirect data access to that volume through the network so hosts that have lost access to the volume can still communicate through the cluster volume owner to the Storage Center volume.

One of the best practices with CSVs is the controlling of the CSV owner. The CSV should be owned by a cluster node that is in the primary site and mapped directly to the Storage Center. In this way, if the CSV goes into Network Redirected Mode, the CSV owner is in the same site and downtime can be eliminated or reduced.

Figure 50 depicts a multi-site Hyper-V cluster with Live Volume. In this figure, Storage Center B was taken offline. CSV network redirection can take over and proxy all the data traffic through the CSV owner on Node A.

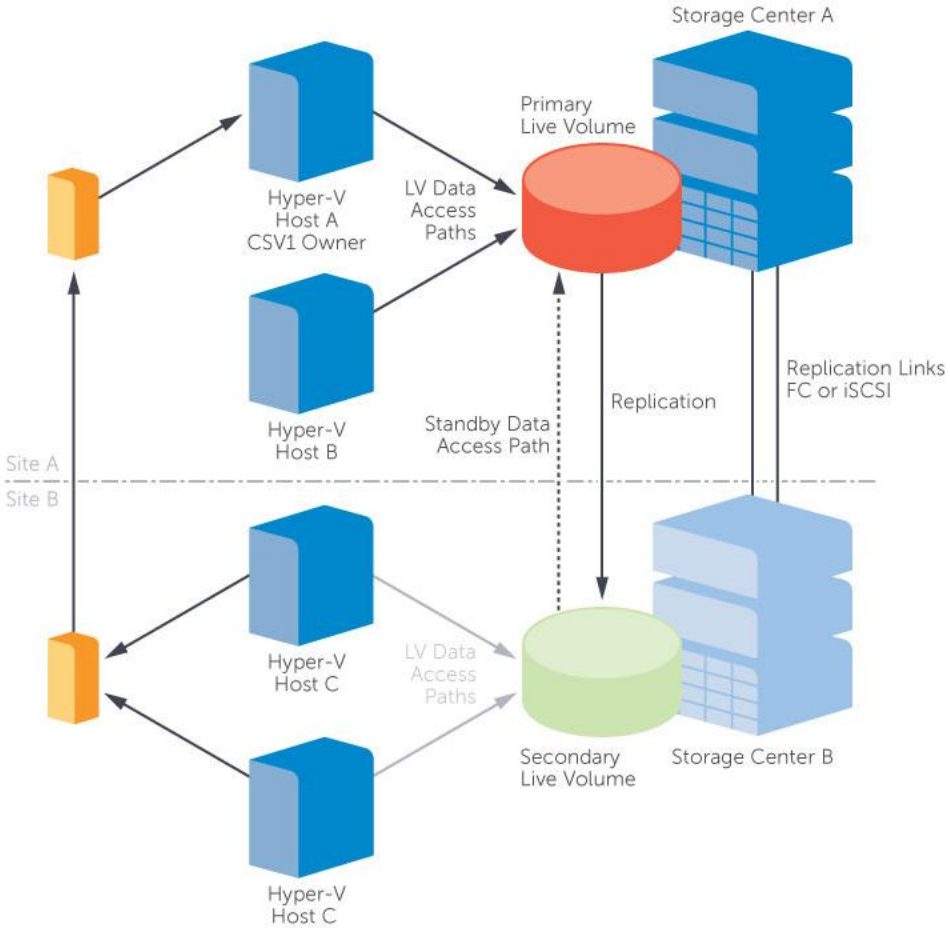

Figure 50 Multi-site Hyper-V cluster with Live Volume

As depicted in Figure 50, if a failure happens that takes down Storage Center B, Hyper-V can redirect access to the volume over the network using CSV Network Redirected Access.

Note: This was only tested with systems that had a flat Ethernet network spanned between the two sites via 1GB connectivity.

10 Live Volume and Synchronous Replication

Synchronous Replication is a feature of Compellent Storage Center that allows two copies of data to be maintained on separate Storage Center arrays using one or multiple replication relationships. These two Storage Center arrays can be in the same location, or can be geographically disperse and connected by a WAN. The integrity of these two copies of data is guaranteed. Synchronous Replication functionality is sub-divided into two user-configurable operating modes, High Consistency and High Availability; these two modes are discussed in further detail in section 3.1, "Modes of operation". Additionally, the operating mode of Synchronous Replication can be changed on the fly depending on the needs of the customer.

Note: The serial number of the replicated volume on the destination Storage Center array is unique to that of the source volume.

When Live Volume in Compellent Storage Center is coupled with Synchronous Replication, it allows the customer to deploy and manage an identical data integrity guaranteed copy of data on a secondary Storage Center array that is at another location. The use of Live Volume creates value in this scenario, by masking the serial number of the replicated volume on the destination Storage Center array to make the serial number appear identical to that of the source volume. The ability to maintain an identical serial number of the destination Storage Center simplifies and expedites the ability to recover from this volume (or volumes) into the destination application stack.

10.1 Live Volume and Synchronous Replication with Linux

The use of Live Volume in conjunction with Synchronous Replication in Red Hat Enterprise Linux deployments offers customers the ability to safely and resiliently manage, backup and recover their business-critical data across both single and multi-site deployments.

The remainder of section 10 discusses some of these use cases along with certain technical considerations in these scenarios respectively.

10.2 Live Volume Managed Replication

Live Volume Managed Replication (LVMR) is the ability of Compellent Storage Center to replicate and manage a third copy of data. It is always attached to the primary Storage Center in a Live Volume scenario. If the Storage Center roles are swapped, the LVMR volume will always follow the Storage Center that assumes the primary role. The mode of replication used for the LVMR volume is determined by the chart below, where there can always be a maximum of only one synchronous replication pair in any scenario.

Table 1 Live Volume replication modes

Live Volume at Site A	Repl. mode of LV to Site B	Repl. mode to LVMR to Site C
Live Volume	Sync	Async
Live Volume	Async	Sync

Live Volume	Async	Async

An LVMR volume (in asynchronous mode, with Active Replay enabled; or synchronous mode) can be used to manage and provide an offsite copy of business-critical data for disaster recovery or data distribution use cases (where locating data closer to the audience could significantly reduce data access latency). The figure below depicts these scenarios.

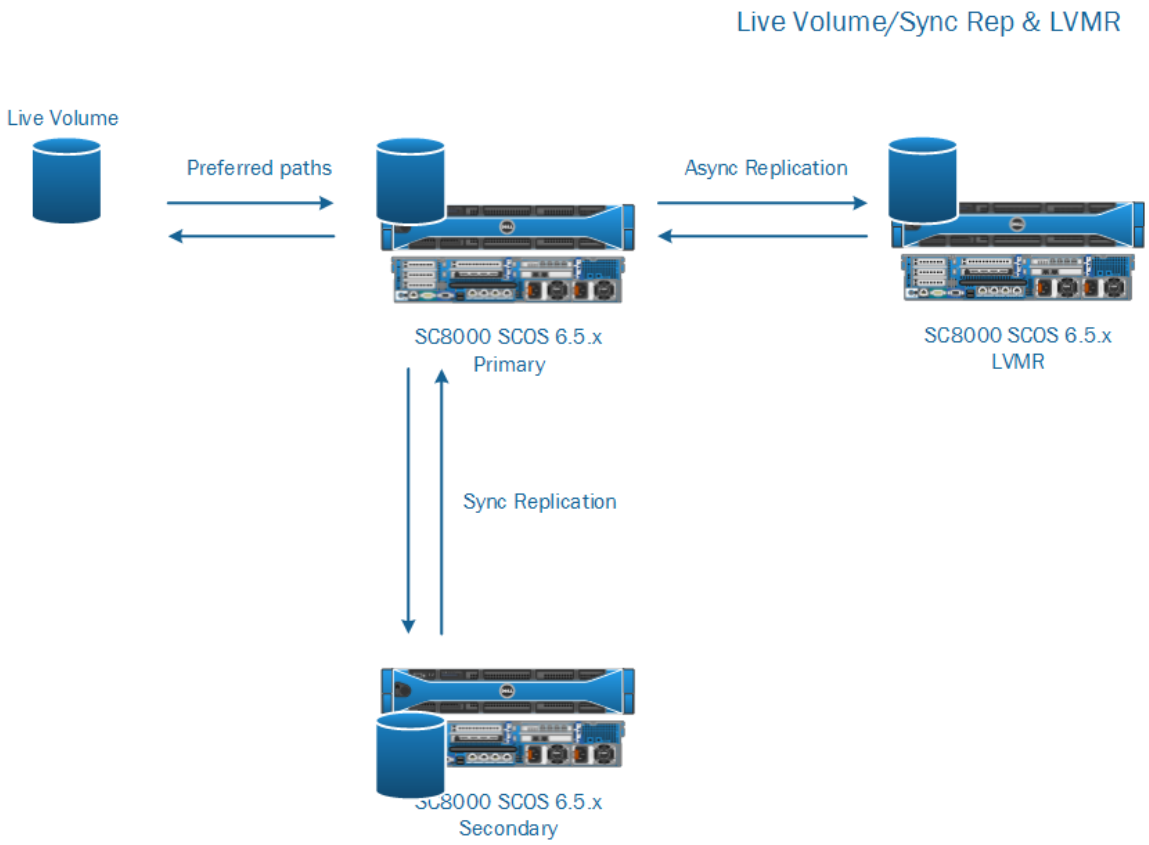

Figure 51 Live Volume/Sync Rep & LVMR

10.3 Use cases

This section discusses various use cases of Live Volume and Synchronous Replication with Linux, and highlights any technical considerations that should be considered. These uses cases are examples and starting points for consideration. It is not a complete list of scenarios where Live Volume and Synchronous Replication can be adapted. The scenarios discussed can apply to both single, as well as multi-site deployments by varying and either scaling up or scaling down the transport mechanisms (LAN, MAN, WAN) connecting Site A to Site B, or Site A to Site C.

10.3.1 Single site

In this single-site scenario, the Linux hosts and Storage Center arrays are geographically located within the same site boundaries though different buildings or labs if necessary. Figure 52 depicts this scenario.

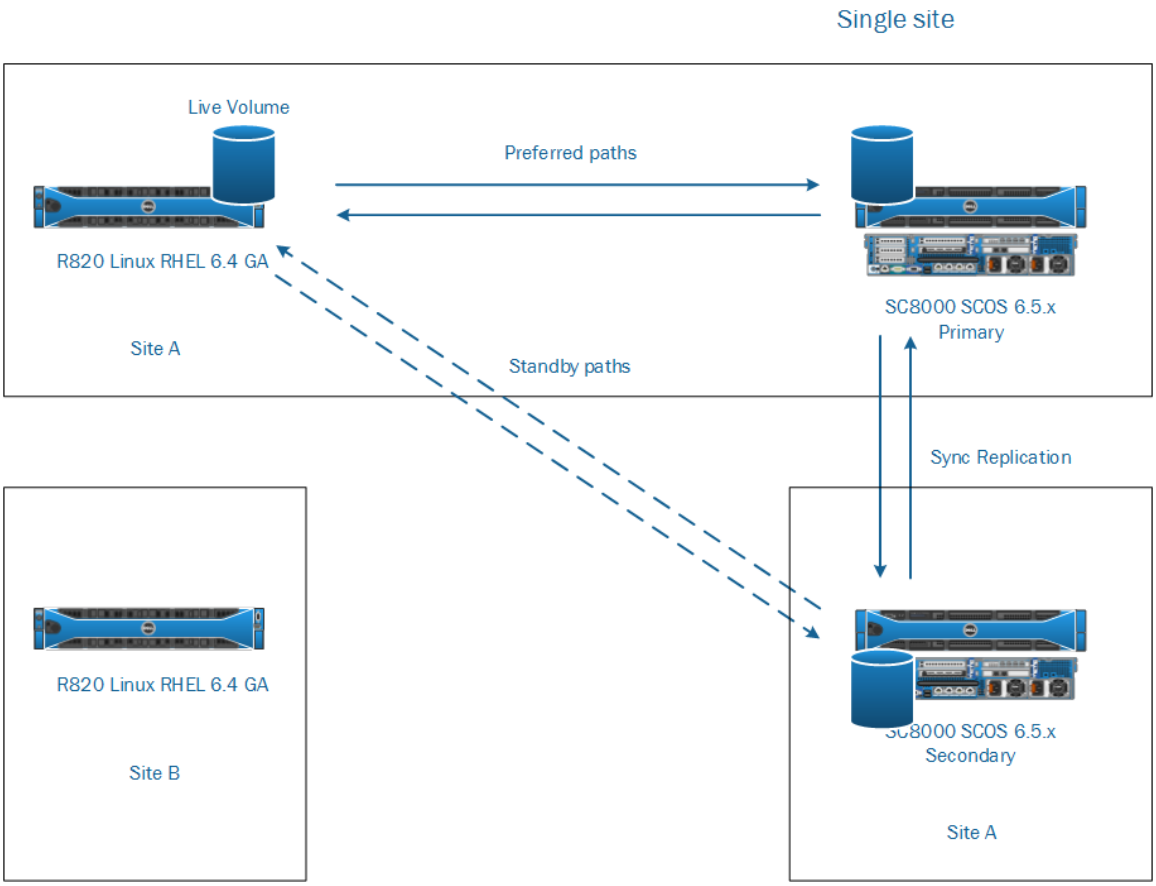

Figure 52 Live Volume/Sync Rep in Single site use case

A volume (or multiple volumes) is first created and mapped to a Linux host. The volumes are scanned, identified and brought into multipath awareness as shown below.

```
[root@tssrv216 ~]# multipath -ll
vol_02 (36000d31000fba6000000000000000015) dm-4 COMPELNT,Compellent Vol
size=10G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=1 status=active
  |- 0:0:0:2  sdd 8:48   active ready running
  |- 1:0:0:2  sde 8:64   active ready running
  |- 0:0:2:2  sdh 8:112  active ready running
  |- 1:0:2:2  sdi 8:128  active ready running
vol_01 (36000d31000fba6000000000000000014) dm-5 COMPELNT,Compellent Vol
size=10G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=1 status=active
```

```
  |- 0:0:7:1  sdj 8:144  active ready running
  |- 1:0:10:1 sdk 8:160  active ready running
  |- 0:0:9:1  sdl 8:176  active ready running
  |- 1:0:12:1 sdm 8:192  active ready running
vol_00 (36000d31000fba6000000000000000013) dm-3 COMPELNT,Compellent Vol
size=10G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=1 status=active
  |- 1:0:0:1  sdc 8:32   active ready running
  |- 0:0:0:1  sdb 8:16   active ready running
  |- 0:0:2:1  sdf 8:80   active ready running
  |- 1:0:2:1  sdg 8:96   active ready running
```

These volumes are then converted into Live Volumes and synchronously (either HA or HC modes) replicated to the alternate Storage Center array. The volumes on the alternate Storage Center array are then mapped back to the same Linux host. The volumes are once again scanned, identified and brought into multipath awareness as shown below. Each volume is now represented by four additional paths, these paths are the mappings from the alternate Storage Center array.

```
[root@tssrv216 ~]# multipath -ll
vol_02 (36000d31000fba6000000000000000015) dm-4 COMPELNT,Compellent Vol
size=10G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=1 status=active
  |- 0:0:0:2  sdd 8:48   active ready running
  |- 1:0:0:2  sde 8:64   active ready running
  |- 0:0:2:2  sdh 8:112  active ready running
  |- 1:0:2:2  sdi 8:128  active ready running
  |- 0:0:5:2  sdo 8:224  active ready running
  |- 0:0:8:2  sdq 65:0   active ready running
  |- 1:0:4:2  sdu 65:64  active ready running
  `- 1:0:8:2  sdw 65:96  active ready running
vol_01 (36000d31000fba6000000000000000014) dm-5 COMPELNT,Compellent Vol
size=10G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=1 status=active
  |- 0:0:7:1  sdj 8:144  active ready running
  |- 1:0:10:1 sdk 8:160  active ready running
  |- 0:0:9:1  sdl 8:176  active ready running
  |- 1:0:12:1 sdm 8:192  active ready running
  |- 0:0:5:1  sdn 8:208  active ready running
  |- 0:0:8:1  sdp 8:240  active ready running
  |- 1:0:4:1  sdt 65:48  active ready running
  `- 1:0:8:1  sdv 65:80  active ready running
vol_00 (36000d31000fba6000000000000000013) dm-3 COMPELNT,Compellent Vol
size=10G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=1 status=active
  |- 1:0:0:1  sdc 8:32   active ready running
  |- 0:0:0:1  sdb 8:16   active ready running
  |- 0:0:2:1  sdf 8:80   active ready running
```

```
|- 1:0:2:1  sdg 8:96   active ready running
|- 0:0:11:1 sdr 65:16  active ready running
|- 0:0:13:1 sds 65:32  active ready running
|- 1:0:11:1 sdx 65:112 active ready running
`- 1:0:13:1 sdy 65:128 active ready running
```

It should be noted that even though these additional paths are shown as active and will be actively used for I/O requests (Round Robin), any I/O requests sent to these paths will be proxied via the replication link to the primary Storage Center for commits. The use of these paths would thus introduce unwelcomed latency to any applications that may be latency-adverse; at this time, path priority definitions and grouping (for example, all paths are prio=1 by default and used in equal fashion) are not configurable between Linux hosts and Storage Center arrays. Therefore, it is not recommended to use this approach for any critical production use cases.

That being said, this scenario can still apply in use cases. In situations where a maintenance event requires Storage Center to be powered down, the volumes on the primary Storage Center array can have Live Volume for an alternate Storage Center array in a different building and/or lab. The roles of the Storage Center arrays can then be swapped, making the alternate Storage Center array the primary array (and the paths from them). The latent paths (from the formerly primary Storage Center array) can then be removed from multipath for the duration of these maintenance events while maintaining uptime and zero disruption to any and all functions and applications that may reside on the Linux hosts. Upon completion of these maintenance events, the volumes and roles of the Storage Center arrays can then be swapped back or left in place to the best discretion the engineering and business requirements.

10.3.2 Multi-site

In this multi-site scenario, the Linux hosts and Storage Center arrays are geographically disperse within a metropolitan (or multi state) region; sites may be connected via MAN or WAN technologies. Figure 53 depicts this scenario. It should be noted that this scenario can also be scaled down for single site deployments as well.

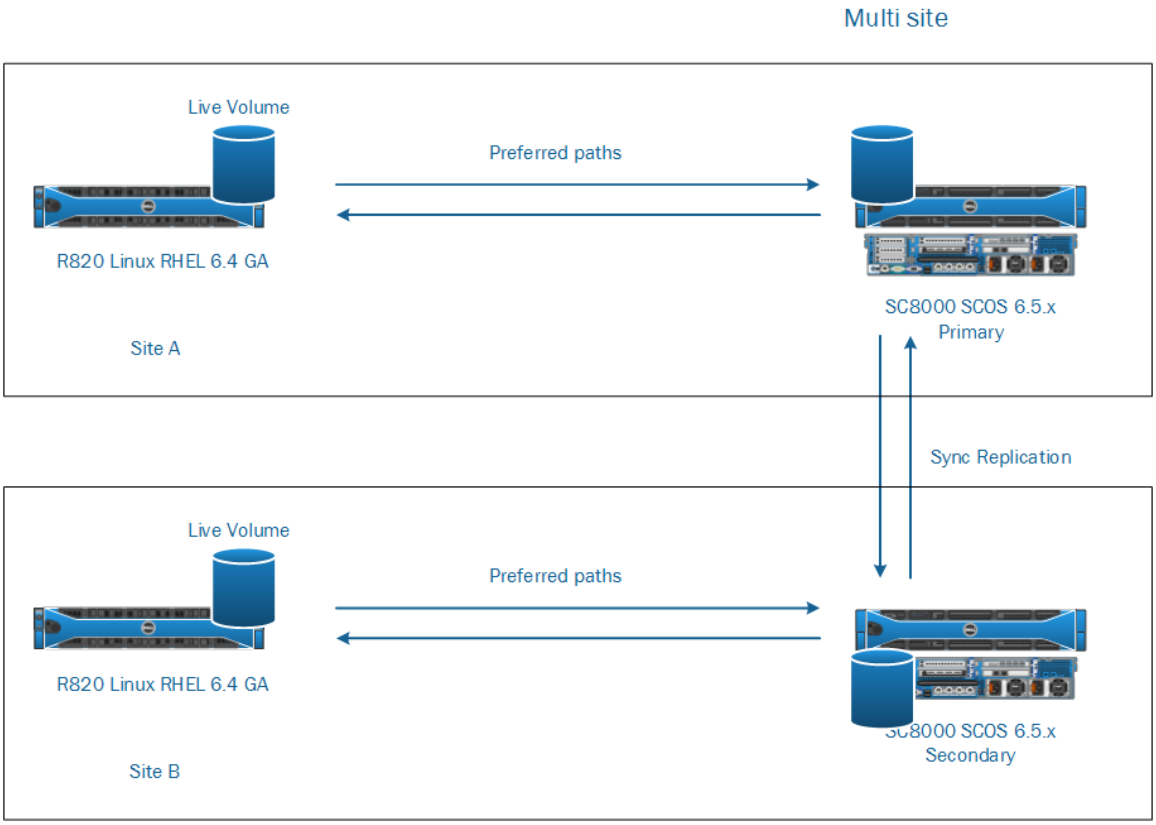

Figure 53 Live Volume/Sync Rep in Multi site use case

The Live Volumes are synchronously (in either HA or HC modes) replicated across Storage Centers. In this scenario, the alternate Storage Center array volumes are mapped to a secondary Linux host instead. It should be noted that the volumes mapped to the secondary Linux host are not shared volumes and do not possess shared I/O management and locking mechanisms. These secondary volumes would need to be remounted to reflect any data changes that were written to the primary volumes. For this reason, the integrity of data across both primary and secondary volumes is guaranteed (in either HA or HC modes) as long as the replication link is in a known good state.

The secondary Linux host can be used in various ways including, but not limited to, the use cases discussed below.

- These volumes can be used to present a consistent read-only copy of this data to a remotely located site; this can apply not only for data distribution reasons, but to also manage and minimize any data access latency concerns.
- The consistency of these volumes (and the integrity of the data that it guarantees) also lends itself towards database replication use. Databases (and applications) can be brought online at the remotely-located site, in either read-only mode or used in a disaster recovery after the roles of the Storage Centers are swapped (the secondary Storage Center becomes the primary Storage Center).
- These volumes can also be used to complement virtualization technologies (for example, RHEV) to allow for the replication of virtual machine workgroups from one site/hypervisor to another

site/hypervisor. RHEV hosts its virtual machines on a storage domain that in turn is correlated through a 1 to many relationship to one or multiple backing Storage Center volumes. These Storage Center volumes are enabled with Live Volume or are synchronously replicated to an alternate site or secondary Storage Center. At the alternate site, these Storage Volumes can then be used to reconstruct or import the storage domain into the alternate data center object and from which reconstruct the virtual machine workgroup.

A few technical and performance considerations should be given thought when exploring these use cases.

/etc/multipath.conf

Keep the contents of /etc/multipath.conf file consistent across all hosts sharing these Live Volumes; this file contains the definitions of volume WWIDs and their respective aliases and will ensure that volumes are identified across the different hosts as the same device, and contain the same data to maintain application integrity.

/etc/fstab

Use the /etc/fstab file in conjunction with /etc/multipath.conf to ensure that volumes are accurately mounted to their respective and proper mount points in the filesystem. Use the multipath device aliases (if defined) or default multipath device naming (mpathX) in /etc/fstab to maintain this consistency.

/etc/ntp.conf

Always use the /etc/ntp.conf file to maintain a certain degree of time-based integrity across all infrastructure. The use of /etc/ntp.conf file becomes even more critical when attempting to maintain data integrity across multiple cluster nodes/sites dispersed across larger geographic deployments (MAN, WAN).

Cluster configurations

In addition to the considerations mentioned above, take into account cluster specific configuration files as well and the best practices of their respective vendors. The discussion of cluster specific configuration remains outside the scope of this paper. The list of enterprise-class clustering technologies include but is not limited to Red Hat Cluster Suite, IBM PowerHA, Oracle Solaris Cluster, Oracle RAC, HP-UX ServiceGuard, and Symantec Veritas Cluster Server just to name a few.

Performance considerations

It should be noted that the dual commit nature of synchronous replication (I/O writes must be committed to the secondary Storage Center volume before it is committed to the primary Storage Center volume) may introduce latency to the applications generating these write requests. The use of synchronous replication guarantees the consistency and integrity of the data at both Storage Center sites when the acknowledgment is received by the requesting application. The use of synchronous replication should be made on the detailed and thorough analysis and understanding of the applications and its I/O needs.

The following demonstration has a small dataset (40MB) that is written to a primary Storage Center Live Volume and is synchronously replicated to an alternate Storage Center. The filesystem is formatted as ext4

and mounted (@ /vol_00) with the discard and sync (the sync option disables filesystem buffer caching) flags.

```
[tssrv216:/root]# cd /etc; time tar cvf - . | (cd /vol_00; tar xvf -)
[snip]
real    0m24.147s
user    0m0.122s
sys     0m1.421s
```

In this next demonstration, the synchronously replicated Live Volume at the alternate site is mounted to the secondary Linux host and writes are applied to the secondary Linux host. This demonstrates additional latency as all write I/O requests are proxied via the secondary Storage Center to the primary Storage Center for processing.

```
[tssrv217:/root]# cd /etc; time tar cvf - . | (cd /vol_00; tar xvf -)
[snip]
real    0m30.864s
user    0m0.111s
sys     0m1.422s
```

In this final demonstration, the replication link is quickly changed from synchronous replication to asynchronous replication and write I/O requests are applied to the primary Storage Center Live Volume. Note that the reduction in time required to commit these writes compared to the use synchronous replication above.

```
[tssrv216:/root]# cd /etc; time tar cvf - . | (cd /vol_00; tar xvf -)
[snip]

real    0m18.266s

user    0m0.104s

sys     0m1.328s
```

10.3.3 Multi-site with LVMR

In this alternate multi-site scenario, the Linux hosts and Storage Center arrays are geographically disperse within a metro (or multi state) region. Sites may be connected through MAN or WAN technologies. Additionally, a third site has been included using the Live Volume Managed Replication (LVMR) feature of Compellent Storage Centers. The figure below depicts this scenario. It should be noted that this scenario can also be scaled down for single site deployments as well.

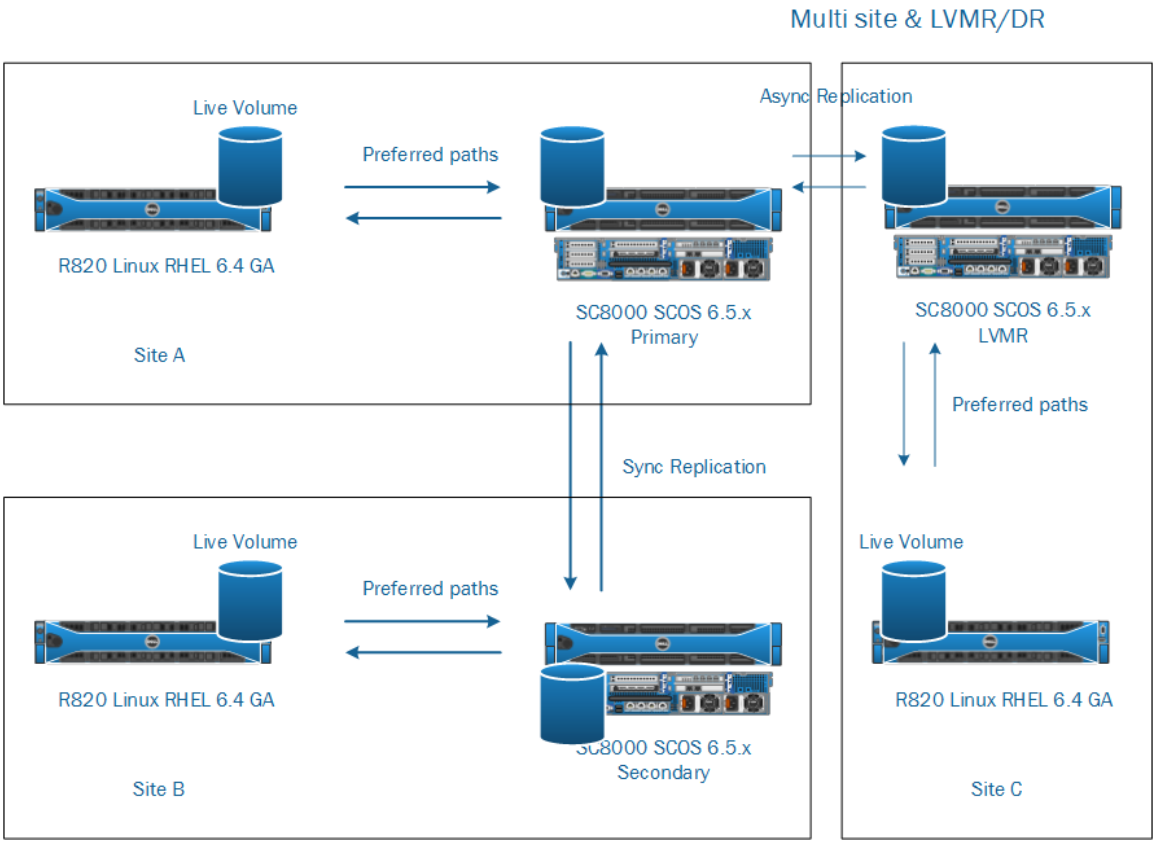

Figure 54 Live Volume or Synchronous Replication in multi-site use case with LVMR

The third site and data in this scenario is replicated asynchronously. One of its uses may include the remotely located disaster recovery copy of business critical data. It should be noted that the LVMR copy is always associated or paired with the primary Storage Center. If the primary and secondary Storage Center roles were swapped, the LVMR copy automatically transfers and associates or pairs itself with the Storage Center that assumes the primary role. Additionally, this asynchronously replicated copy (unlike a synchronously replicated) is current if the replication link has sufficient bandwidth and the **Replicate Active Replay** feature is enabled, or current as of the last captured Replay if the feature is disabled. Consequently, discussions should also be conducted around acceptable RPO and RTO thresholds and maintaining agreeable SLAs with all business entities involved.

The technical considerations discussed in the section 10.3.2 should also be taken into consideration when setting up this use case.

11 Use cases

Section 11 exhibits additional examples of how Live Volume can be used in a variety of environments. Live Volume is not limited to these use cases.

11.1 Zero downtime SAN maintenance and data migration

By utilizing Live Volume, maintenance activities can be performed without downtime on a Storage Center. This includes tasks such as taking a Storage Center offline to move its location, perform service-affecting enclosure or disk firmware updates, and migrate the volume to a new SAN.

11.1.1 Requirements

The requirements for this operation would be:

- MPIO installed and appropriately configured on the host computers.
- Server(s) properly zoned into both Dell Compellent Storage Centers.
- Server(s) configured on both Storage Centers.
- At least a 1Gb low latency replication link between Storage Centers.

Summary: In advance of a planned outage, Live Volume can non-disruptively migrate volumes from one Storage Center to another, enabling continuous operation for all applications – even after one Storage Center has completely powered down.

Operation: In an on-demand, operator-driven process, Live Volume can transparently move volumes from one Storage Center to another. The applications operate continuously. This enables several options for improved system operation:

- Redefine remote site as Primary for all volumes on local site
- Shut down local site
- Reverse process after planned outage is completed

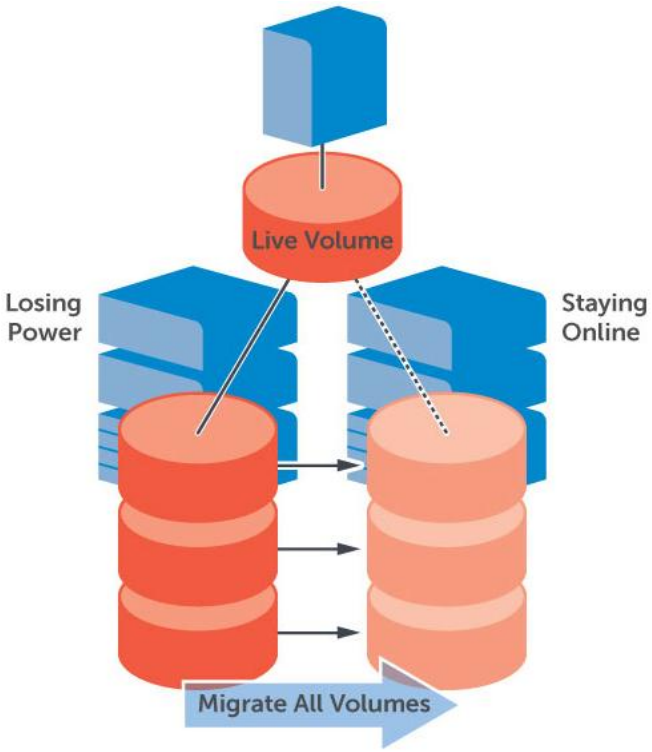

Figure 55

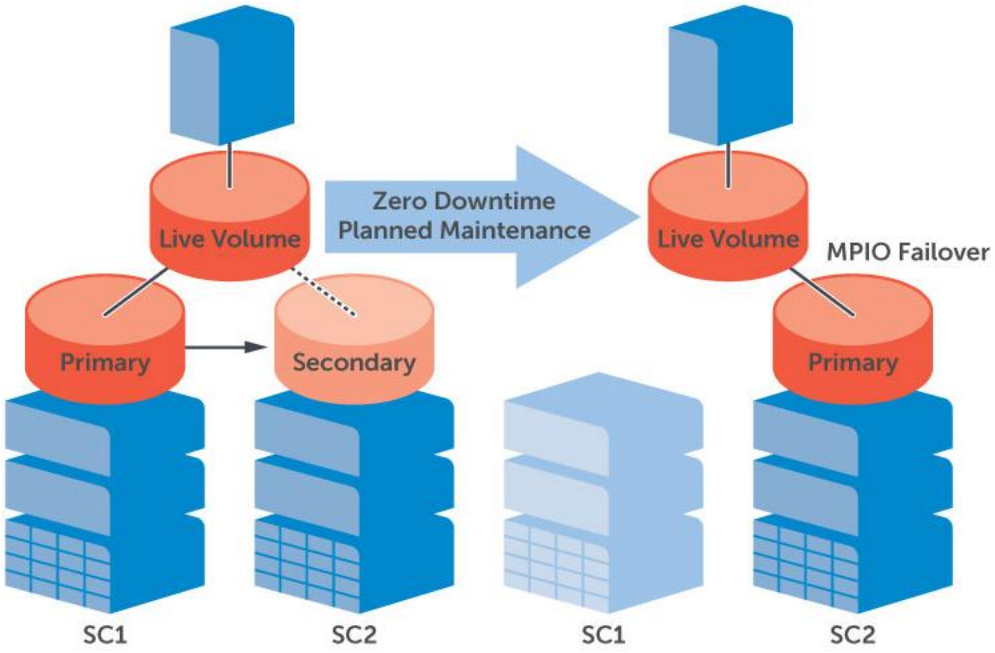

Figure 56

11.2 Storage migration for virtual machine migration

As VMware, Hyper-V, or XenServer virtual machines are migrated from data center to data center, Live Volume can automatically migrate the related volumes to optimize performance and minimize I/O network overhead.

Live Volume continuously monitors for changes in I/O traffic for each volume and non-disruptively moves the primary storage to the optimal Storage Center for optimum efficiency.

- Migrate the virtual machine using the server virtualization software
- Live Volume will track the changes in I/O traffic mapping and will perform a Primary/Secondary swap after a fixed amount of time and data have been transferred

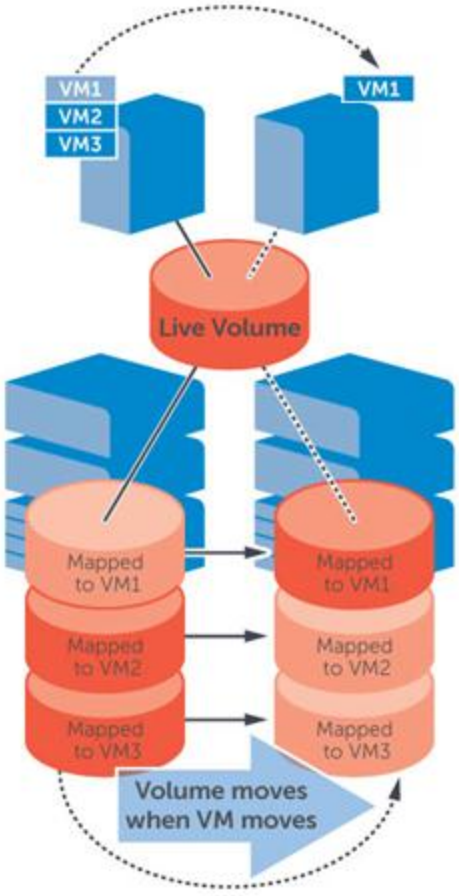

Figure 57 Storage follows the application (server virtualization)

11.2.1 Requirements

The requirements for this operation would be the following:

- Server(s) properly zoned into both Dell Compellent Storage Centers.
- Server(s) configured on both Storage Centers.
- Stretched Layer 2 networking between source and destination sites or hypervisors
- 1Gbps or better low latency iSCSI or Fiber Channel connectivity between the Storage Centers to support asynchronous or synchronous Live Volume replication.

11.3 Disaster avoidance

In anticipation of an unplanned outage (for instance, an approaching hurricane), Live Volume can migrate data to remote systems before the local system has an outage. Live Volume used in this manner will prevent data loss and will enable an extremely rapid restart at the remote site.

Operation: In an on-demand, operator-driven process, Live Volume can transparently move volumes from one Storage Center to another. The applications operate continuously. This enables several options for improved system operation:

- Redefine remote site as Primary for all volumes on local site
- Shut down applications on local site
- Restart applications on remote site
- Reverse process after risk of potential outage is gone

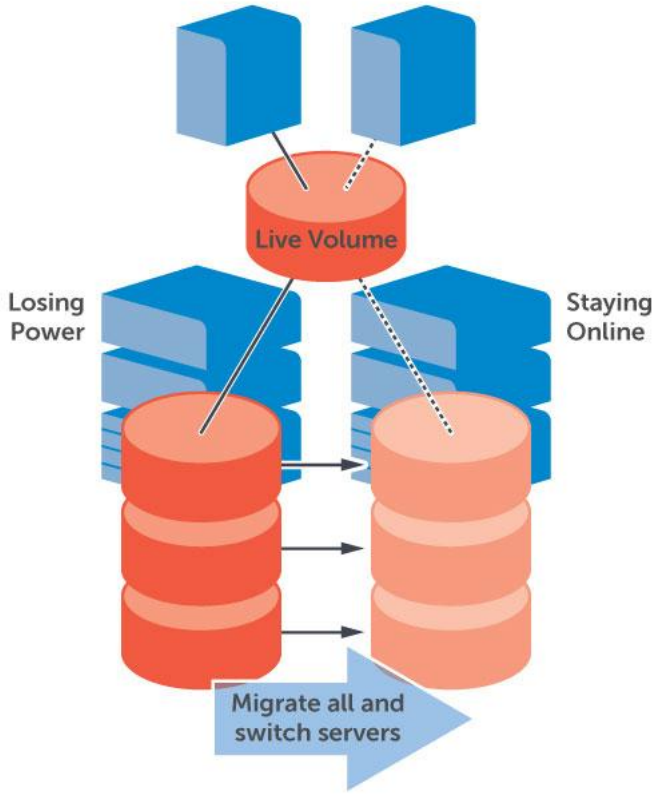

Figure 58 Disaster avoidance

11.4 On-demand load distribution

Transparently distribute workload, balance storage utilization, or balance I/O traffic between two Storage Centers.

Configuration: Storage Centers must be connected using high bandwidth and low latency connections, especially when synchronous replication is used with Live Volume.

Operation: In an on-demand, operator-driven process, Live Volume can transparently move volumes from one Storage Center to another. The applications operate continuously. This enables several options for improved system operation:

- Distribution of I/O workload
- Distribution of storage
- Distribution of front end load traffic
- Reallocation of workload to match capabilities of heterogeneous systems

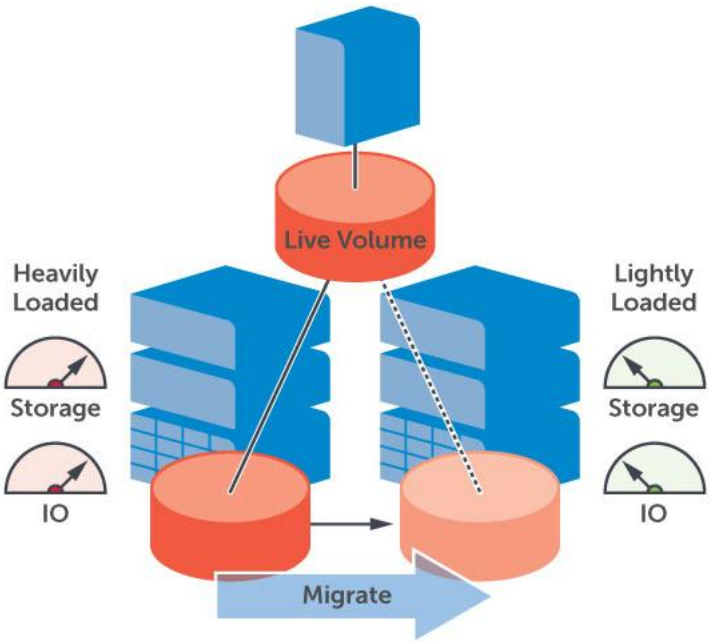

Figure 59 On demand local distribution

11.5 Cloud computing

Summary: Transparently distribute workload, balance storage utilization, or balance I/O traffic between multiple Storage Centers within a data center, enabling continuous flexibility to meet changes in workload and to provide a higher-level system up time.

Configuration: Live Volumes can be created between any two Storage Centers in a data center. Each Storage Center can have many Live Volumes, each potentially connecting to a different Storage Center.

Operation: In an on-demand, operator-driven process, Live Volume can transparently move volumes from one Storage Center to another. The applications operate continuously. This enables several options for improved system operation:

- Distribution of I/O workload
- Distribution of storage
- Distribution of front end load traffic
- Reallocation of workload to match capabilities of heterogeneous systems

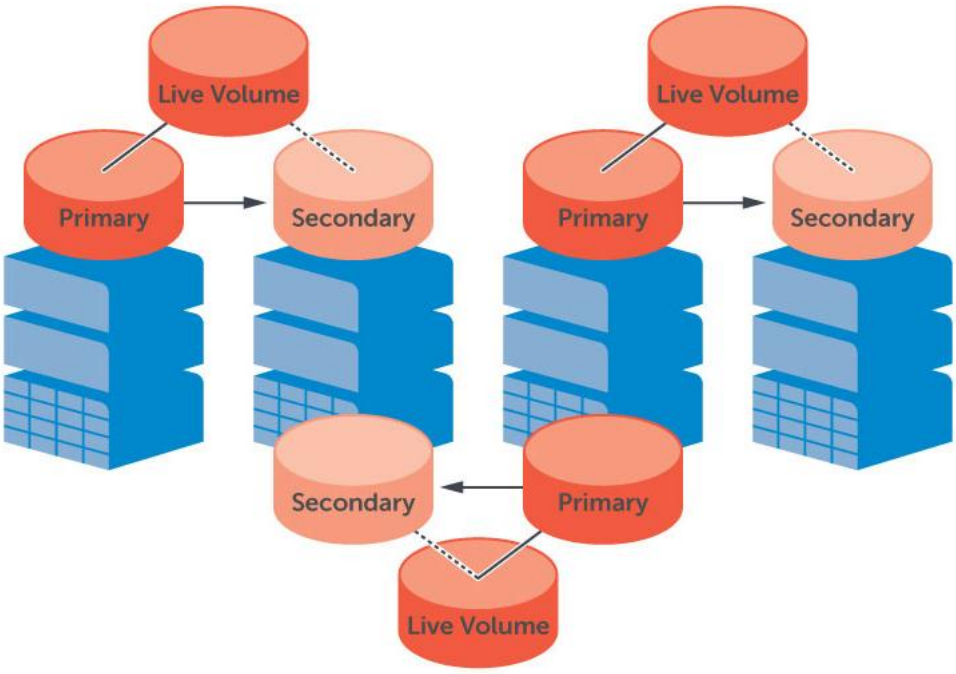

Figure 60 Cloud computing

11.6 Replay Manager and Live Volume

Replay Manager and Live Volume are mutually exclusive features and are not supported for use together. If storage hosts are split between data centers in a stretched cluster configuration using Live Volume, then full Replay Manager support is not currently available. While consistent Replays are replicated to the secondary Storage Center, they are not accessible as recovery points from Replay Manager.

A Resources

Dell Compellent Home Page: http://www.dellstorage.com/

Knowledge Center: http://kc.compellent.com/

Dell Compellent Best Practices with VMware vSphere 5.x:
http://en.community.dell.com/techcenter/extras/m/white_papers/20437942.aspx

Compellent Best Practices with Site Recovery Manager:
http://en.community.dell.com/techcenter/extras/m/white_papers/20438016.aspx

High Availability and Disaster Recovery with VMware vSphere Solutions Guide:
http://en.community.dell.com/techcenter/extras/m/white_papers/20438018.aspx

Hyper-V 2012 R2 Best Practices for Dell Compellent Storage Center:
http://en.community.dell.com/techcenter/extras/m/white_papers/20437923.aspx

Dell Compellent Storage Center Disaster Recovery for Microsoft Hyper-V Best Practices:
http://en.community.dell.com/techcenter/extras/m/white_papers/20437950.aspx

Dell Compellent Red Hat Enterprise Linux (RHEL) 6.x Best Practices
http://en.community.dell.com/techcenter/extras/m/white_papers/20437964.aspx

Red Hat 6 Device Mapper Multipathing
https://access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/DM_Multipath/MPIO_Overview.html

Red Hat 6 Virtualization Admin Guide
https://access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Virtualization_Administration_Guide/index.html

Red Hat 6 Installation Guide
https://access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Installation_Guide/index.html