PS Series Asynchronous Replication Best Practices and Sizing Guide

Dell EMC Engineering
November 2016

A Dell EMC Best Practices Guide

Revisions

Date	Description
August 2013	Initial release
November 2016	Added updates for delegated space in multiple pools as well as a sizing example

Acknowledgements

Updated by: Chuck Farah

Table of contents

1 Introduction

This white paper describes the Dell EMC PS Series asynchronous replication feature and presents lab validated best practices to help IT and SAN administrators understand and fully utilize the powerful set of volume data replication features delivered with every PS Series storage array.

The PS Series builds on a unique peer-storage architecture that is designed to provide the ability to spread the load across multiple array members to provide a SAN solution that scales with the customer's needs. This pay-as-you-grow model allows customers to add arrays as their business demands increase the need for more storage capacity or more I/O capacity.

Every PS Series array includes additional features such as snapshots, clones, replication, and all-inclusive software: Group Manager, SAN Headquarters (SAN HQ), and Host Integration Tools. The built-in snapshot feature enables quick recovery of files and clones for recovery of files or volumes, and the replication feature allows the implementation of disaster recovery initiatives.

The PS Series software includes storage system-based replication. This feature (known as asynchronous replication) provides the ability to replicate data volumes to peer PS Series storage arrays situated in remote locations without setting the volumes offline. Asynchronous replication provides a disaster-recovery option in case the original volume (or the entire PS Series group) is destroyed or otherwise becomes unavailable. Asynchronous replication is a point-in-time replication solution that offers extended-distance replication. PS Series asynchronous replication provides asynchronous, incremental data synchronization between primary and secondary replicas. Scheduled replication events update the remote copy of the data with all the changes that occurred on the primary copy since the last replication event occurred.

1.1 Audience

This white paper is intended for storage administrators who are involved in the planning or implementation of a data recovery solution using PS Series asynchronous replication. Readers should be familiar with general concepts of PS Series iSCSI storage as well as Ethernet LAN and WAN network concepts.

1.2 Key benefits of using PS Series asynchronous replication

Using asynchronous replication with a PS Series SAN as part of your storage infrastructure can provide multiple benefits:

SAN-based replication is included at no additional cost: PS Series asynchronous replication replicates volumes to remote sites over any distance by leveraging existing IP network infrastructure.

Ease of use: PS Series Group Manager and SAN Headquarters (SAN Headquarters) are included at no additional cost. These easy-to-use GUI based tools allow for managing and monitoring PS Series arrays, including all replication oriented tasks.

Manual Transfer Utility: The Manual Transfer Utility is also included at no additional cost. This host-based tool integrates with the native replication function of the PS Series firmware to provide secure transfer of large amounts of data between PS Series groups using removable media. The Manual Transfer Utility is beneficial in environments where data protection is critical but bandwidth is limited.

Asynchronous replication addresses varying needs using powerful features and configuration flexibility:

- Multiple recovery points are efficiently stored.
- Per-volume replication schedules permit varying service levels.
- You can fast failback to the primary site by synchronizing only the data that has changed while the secondary site was in use.
- One-way, reciprocal, one-to-many, or many-to-one replication paths are possible.
- Thin replicas provide space efficiency.

2 Asynchronous replication overview

PS Series asynchronous replication is used to replicate volumes between different groups as a way to protect against data loss. The two groups must be connected through a TCP/IP-based network. This means that the physical distance between the groups is not limited. The replication partner group can be located in the same data center, or it can be in a remote location. In practice, the actual bandwidth and latency characteristics of the network connection between replication groups must be able to support the amount of data that needs to be replicated and the time window in which replication needs to occur.

IT administrators can enable replication through the Group Manager GUI or the Group Manager CLI. Once enabled, volume replication functions can be managed using the GUI, the CLI, or Auto-Snapshot Manager (ASM) tools. ASM tools are included as part of the Host Integration Tools for Microsoft®. VMware® also has similar Host Integration Tools known as Virtual Storage Manager (VSM). A replica can be created from a single volume or a volume collection. A volume collection allows up to eight volumes to be grouped together so that the replication process for that collection can be managed as if it was a single volume. This is useful when replicating multiple volumes that belong to the same application.

Asynchronous replication initially creates a copy on a secondary storage system, and then synchronizes the changed data to the replica copy. A replica represents the contents of a volume at a point in time at which the replica was created. This type of replication is often referred to as *point-in-time replication* because the replica copies the state of the volume at time the replication is initiated. The frequency in which replication occurs determines how old the replica becomes relative to the current state of the source volume.

2.1 Terminology

This document uses the following PS Series terminology:

Table 1 Terminology

Term	Description
Asynchronous replication	The built-in replication feature included with every PS Series array.
Replica	A point-in-time synchronized copy of a PS Series volume stored in a secondary group.
Replica set	A collection of all point-in-time synchronized replicas for a specific source volume.
Pool	A pool is a storage space that each member (array) is assigned to after being added to a group. A group may have up to four pools, and a pool may have up to 8 members[1].
Group	A group consists of one or more PS-Series arrays connected to an IP network that work together to provide SAN resources to servers. A group may contain up to 16 members[2] and is managed as a single storage entity.
Primary group	A group containing the source volume(s) to be copied or replicated.
Source group	Same as primary group.

Term	Description
Secondary group	A group containing the replica or copy of the source volume(s).
Destination group	Same as secondary group.
Delegated space	The amount of space on the secondary group that is *delegated* to a replication partner, to be reserved for retaining replicas.
Replica reserve	The space allocated from delegated space in the secondary group to store the volume replica set for a specific volume.
Local reserve	The amount of space reserved on the local or primary group for holding temporary snapshots and failback snapshots of the source volume.
WAN emulator	A device used to simulate distance and impairments in a WAN.

[1]PS4000, PS4010, PS4100, PS4110, PS-M4110, and PS4210 arrays may only have two members in a pool, or 28 members if using VMware Virtual Volumes (VVols).

2.2 Configuration limits and storage requirements

PS Series arrays allow for up to 256 volumes for replication per group (PS4000 to PS4210 series arrays are limited to 32 volumes enabled for replication per group), but only 16 volumes per group can be simultaneously replicating. A group may also have up to 16 replication partners. A volume can have only one replication partner at a time.

Asynchronous replication requires reserved disk space on both the primary and secondary groups. The amount of space required depends on several factors:

- Volume size
- The amount of data that changes (on the source volume) between each replication period
- The number of replicas that need to be retained on the secondary site
- If a failback snapshot is retained on the primary group

The default values that appear in Group Manager are sufficient to ensure that enough space is reserved for at least one successful replication, even if the entire contents of a volume are altered between replicas. Initial replication of a volume will cause the complete contents of the volume to be copied to the secondary group. Subsequent replication events will copy only the changed data.

2.3 Local reserve, delegated space and replica reserve

Local reserve is a volume-level setting that defines how much space is allocated on the primary group to support replication processing. Delegated space is a group -level setting that defines the total space dedicated to receiving and storing inbound replica sets on the secondary group from a primary group replication partner. Replica reserve is a volume-level setting that defines how much space is allocated per volume from delegated space on the secondary group for storing all replicas of the volume (the replica set). The locations and relationships between local reserve, delegated space and replica reserve are shown in Figure 1.

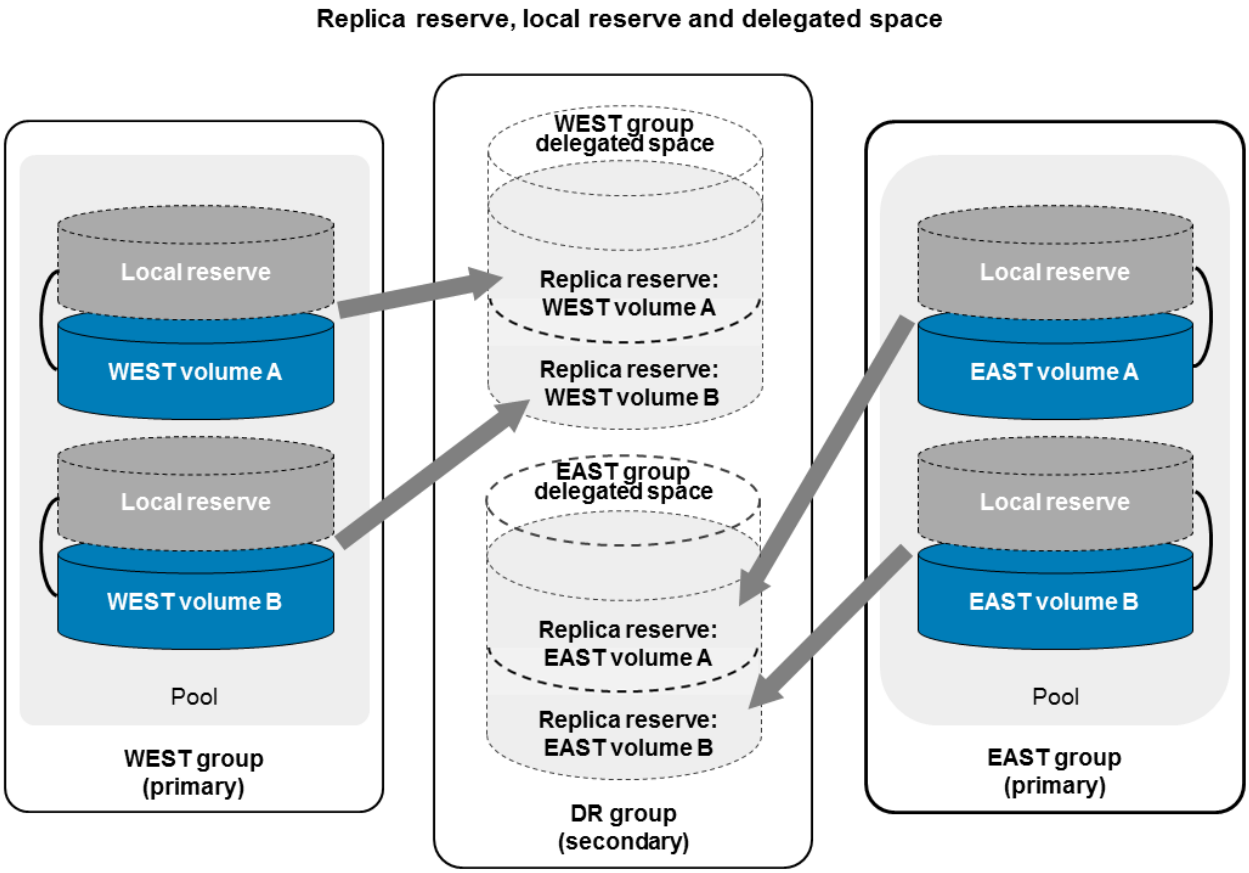

Figure 1 Local reserve, replica reserve, and delegated space

A storage group can also have multiple replication partners. One typical example is where multiple locations need to replicate to the same remote disaster recovery site. For example, in Figure 1, volumes within the EAST and WEST groups are both replicating to the DR group. It is not necessary to have separate pools to support each site. A single pool at the remote site can have partnerships with up to 16 other sites and receive replicas from them. In Figure 1, two primary sites (EAST and WEST) are replicating into the same pool on the secondary (DR) site. Of course the DR site in this example must be sized appropriately to support replica storage requirements for both of the replication partnerships shown.

2.4 Replication partnerships

With the exception of failback events, PS Series replication is always a one-way process, in which data moves from the primary group volume to the secondary group replica set. Replication also allows for reciprocal partners, which means that you can use two operational sites as recovery sites for each other. Figure 2 shows examples of different replication partnership scenarios. To simplify the diagram, the fast failback path is only shown in the basic partner path example within Figure 2 (fast failback works for all replication paths).

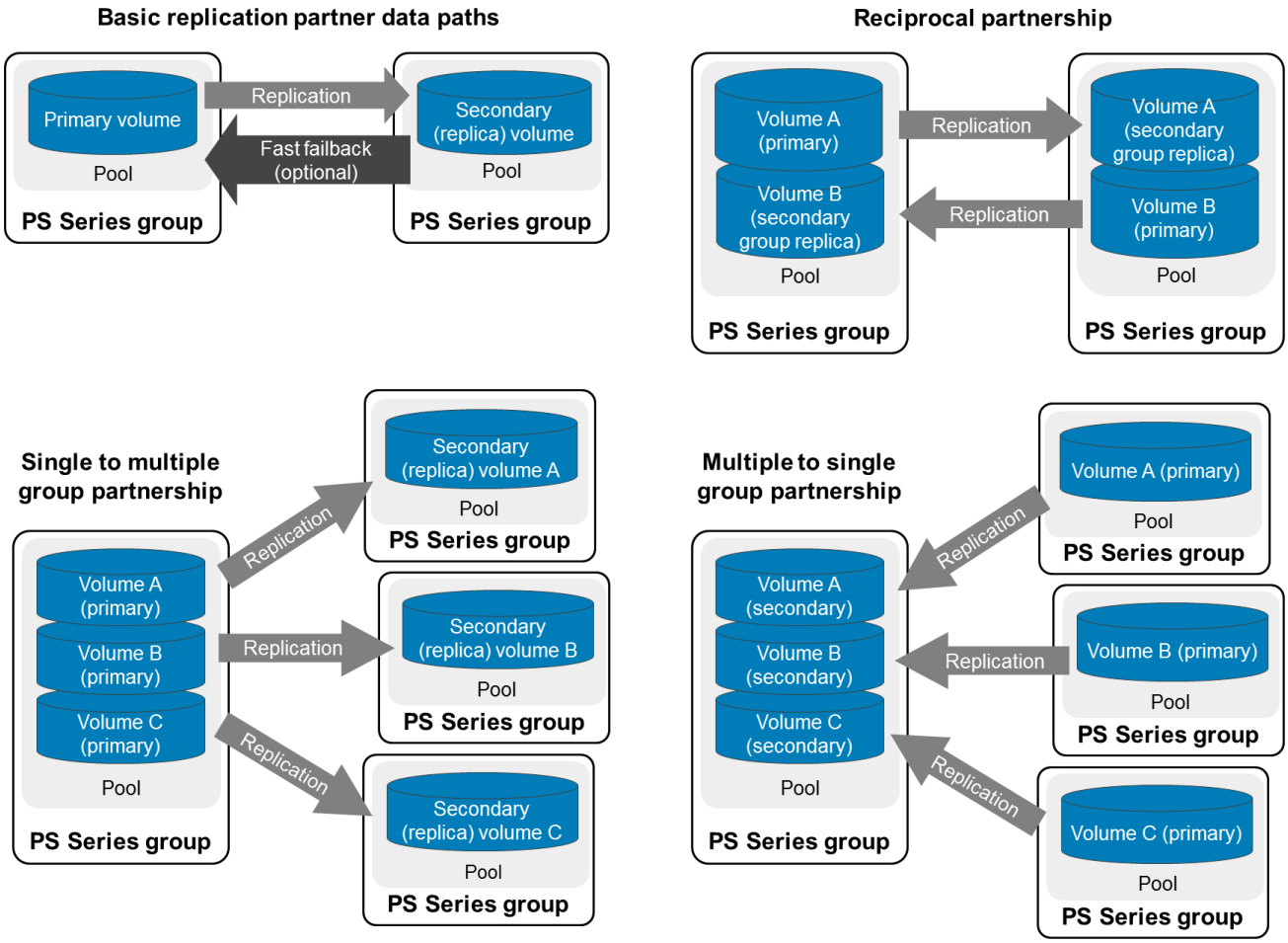

Figure 2 Replication partnership paths

A volume's replica set is identified by the volume name with a numbered extension. The number corresponds to the number of replication partners, in the order they were added. For example, an inbound volume from the first replication partner, with the name **repl-vol1**, will have a replica set with the name **repl-vol1.1** as shown in Figure 3. A volume from a second replication partner, also with a source volume named **repl-vol2** will have a replica set named **repl-vol2.1**. Each replica that makes up the replica set is identified with a date and time stamp that correlates to the time that the replica was created.

Navigation tip: **Group Manager GUI** (secondary) > **Replication** > **Inbound Replicas**

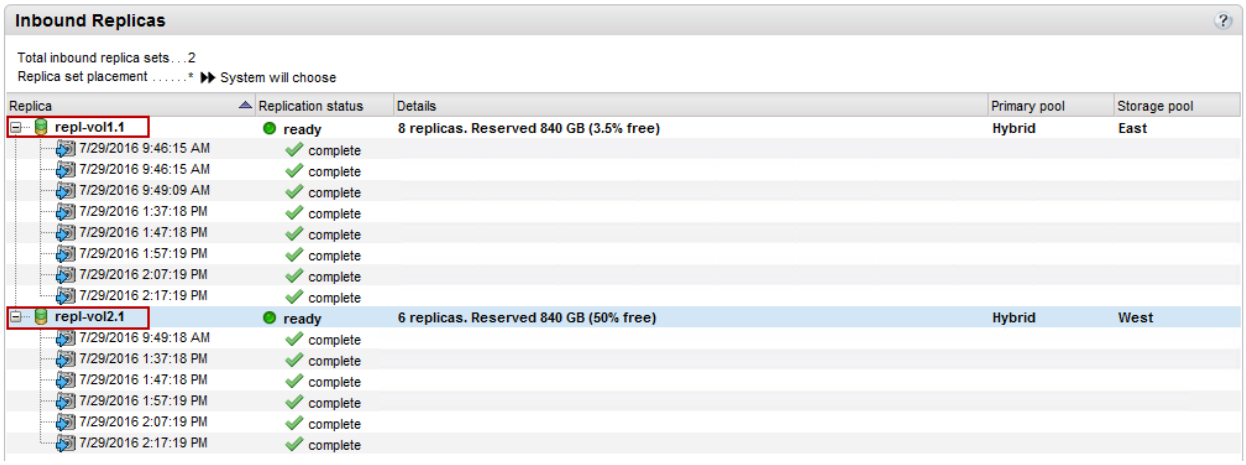

Figure 3 Inbound Replicas

Typically a volume will be replicated across a network (SAN or WAN). However, Dell EMC also provides a Manual Transfer Utility (MTU). This tool allows an administrator to create a local replica on a portable storage system (such as an external USB disk). After transporting the portable storage to the secondary site, you can transfer the replica data from the portable storage system to the secondary group. Once the volume transfers, then asynchronous replication will synchronize only new changes that occurred after the copy onto portable storage was made to the remote replica, conserving WAN bandwidth. If necessary, you can use the MTU for both initial and subsequent delta replication events. The Manual Transfer Utility can also be used to conserve WAN bandwidth if a large one-time update is applied to a volume.

For unusually large amounts of data, expedited shipping of external media (tape or disk) may be faster than electronic transfer across a network. Also, depending on how often you need to replicate, shipping may be more cost-effective than provisioning the connection path with the necessary bandwidth.

The design of the network link between primary and secondary groups will affect how fast changes replicate between arrays. This in turn can also dictate how many replicas can be created in a given time period. For example, if you have a constant volume data change rate and it takes four hours to replicate changes made (since the last replica) then in practice you should be able to complete up to six replication processes within a 24-hour period.

3 PS Series replication process

When a replica is created, the first replication process completes the transfer of all volume data. For subsequent replicas, only the data that changed between the start time of the previous replication cycle and the start time of the new replication cycle is transferred to the secondary group. Dedicated volume snapshots are created and deleted in the background as necessary to facilitate the replication process. Logically, you could describe a volume replica set as a combination of the following:

Volume replica set	=	A full copy of the primary volume, with data synchronized to the beginning of the most current completed replication.	+	A time-sequenced set of replicas, in which each replica corresponds to the state of the volume at the beginning of a prior replication.

The number of prior replicas that are stored on the secondary group is limited by the size of the replica reserve allocated for that volume and the amount of data that changes. See section 7.4 to properly size a PS Series replication solution. The replication process can be described as a series of phases. The flowchart in Figure 4 shows the process phases, focusing on how the process tracks and copies changes that occur between each replica cycle.

The primary group checks for availability of sufficient delegated and replica reserve space on the secondary group at the beginning of each replication processing phase. If adequate space is not available, the process will pause and generate an event message. Replication will continue once sufficient space is made available. These parts of the process are not shown in the chart.

Proper sizing and capacity planning should be considered before the replication solution is decided. Section 7,

Best practices for planning and design, provides direction on the proper capacity as well as performance considerations for asynchronous replication with PS Series storage.

The following subsections refer to the phases shown in Figure 4.

3.1 Replication setup (one-time)

This phase configures the replication partnership and volume replication settings.

3.2 Replication processing (repeating)

Primary-to-secondary volume data replication is completed in this phase. The process steps vary based on replication status (first or subsequent) and fast failback mode (enabled or disabled). During this process, the local reserve is consumed by a hidden snapshot (and the fast failback snapshot if enabled). Volume data changes that occur during the replication processing phase are stored by the hidden snapshot in the local reserve. The replica reserve allocated to the volume within delegated space on the secondary group receives all volume data changes. The replica reserve is consumed by the most recent complete volume replica plus all prior replicas stored in the replica set.

3.3 Between replication events (repeating)

Once first replication has occurred, the system continues to keep track of volume data changes that occur so that subsequent replication processes can copy those changes to the replica set. This tracking process does not consume additional space.

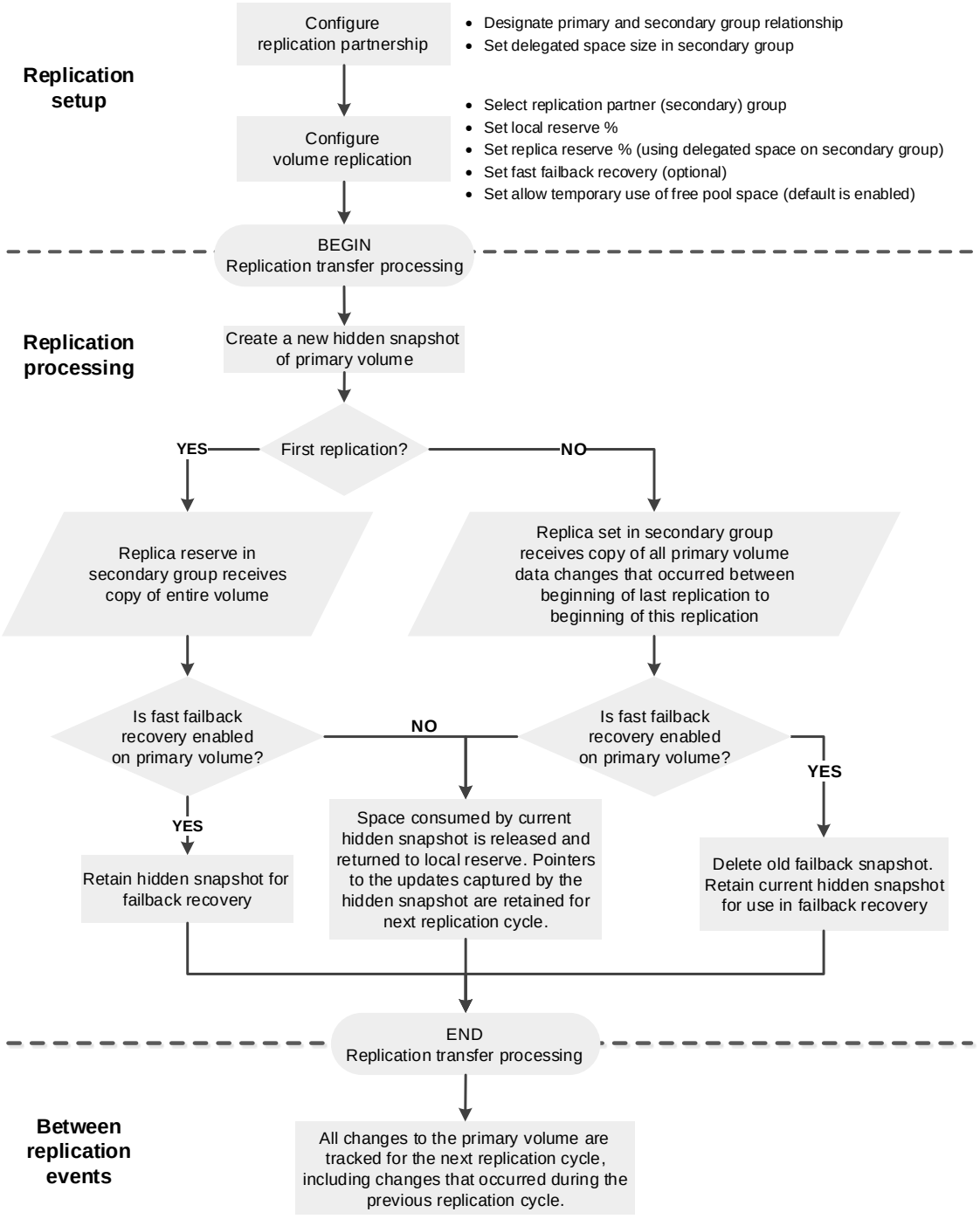

Figure 4 Replication process

With replication, it does not matter if the volume is thin provisioned or uses a traditional volume. In either case, only the data that has changed will be copied to the replica. On the secondary side, volumes are always thin provisioned to conserve available capacity used by the replica reserve for that volume.

3.4 Fast failback

With fast failback enabled, you can ensure that the volume data preserved by the failback snapshot on the primary group always matches the volume data in the most recent replica stored on the secondary group. If you have an event that causes failover to the secondary group and the workload subsequently writes changes to the replica volume, failback snapshot supports a quicker failback to the primary group by replicating only the changes made to the replica volume during the time it was active as a recovery volume on the secondary group. If the failback snapshot is not enabled, you must replicate the entire volume contents back to the primary group to complete the failback operation. Depending on the size of the volume, the failback scenario can take significantly longer to complete if fast failback is not enabled.

3.5 Space borrowing for replication

PS Series firmware version 8.0 provides the ability for snapshots, local and remote replicas, and deleted volumes in the Group Volume Recovery Bin to temporarily borrow space beyond the configured reserves. This feature, called space borrowing, simplifies configuring reserve space, improves space utilization, and enhances management of snapshots and replica reserves.

While it is possible to enable or disable space borrowing for snapshots, space borrowing for replication is automatic and cannot be disabled.

Remote replicas can borrow beyond their total replica reserve, but the total amount of configured reserve space must still fit within the delegated space. If there is insufficient delegated space on the secondary group, the system requires manual administrative intervention to increase the amount of delegated space.

Also, if the replica reserve for a volume is configured with a very low value, such as the minimum 105%, the system can potentially require manual administrative intervention to increase the reserve percentage so that an in-progress replica can continue. In-progress replicas are not eligible to borrow space.

Note: To use space borrowing for replicas, all members in the secondary group must be running PS Series firmware version 8.0 or later. Space borrowing for snapshots requires PS Series firmware v6.0 or later. For more information on space borrowing, refer to *Space Borrowing for Snapshots and Replicas*.

4 Test topology and architecture

To properly design a replication scenario, administrators must understand how the quality of the network connection between groups can affect replication. Also, as discussed previously, when data changes on the source volume, it will be replicated over the network link. The amount of changes occurring will directly affect how long it takes for each replica to complete. To help illustrate these points, asynchronous replication was set up in a lab and test results gathered.

The test configuration (see Figure 5) consisted of a designated primary site and secondary site, although all of the hardware was physically located in the same data center. Storage on the primary side used three PS Series PS6010XV arrays connected to a pair of Dell PC8024F switches. All three members were configured as a single pool.

The secondary side used another pair of Dell switches and a single PS Series PS6510E array configured in the default pool. For redundancy, each pair of PC8024F switches was connected by creating a LAG (Link Aggregation Group) using two 10 Gb ports on each switch, and the storage controller ports were distributed across the switches. The primary and secondary sites were connected through an Apposite® Technologies Netropy® 10G WAN emulator. This allowed throttling of bandwidth and adding impairments such as latency to the WAN connection to simulate the various speeds and conditions that one might encounter with a WAN.

The primary site connected a Dell EMC PowerEdge™ R610 server to the PC8024F iSCSI SAN switches. The PowerEdge R610 ran Microsoft Windows Server® 2008 R2, which allowed mounting of the volumes and generating disk I/O to overwrite or change the existing data on the disk. Several volumes of 100 GB and 10 GB were created to use for replication testing. This configuration was used throughout all testing unless noted in the individual test cases later in this document.

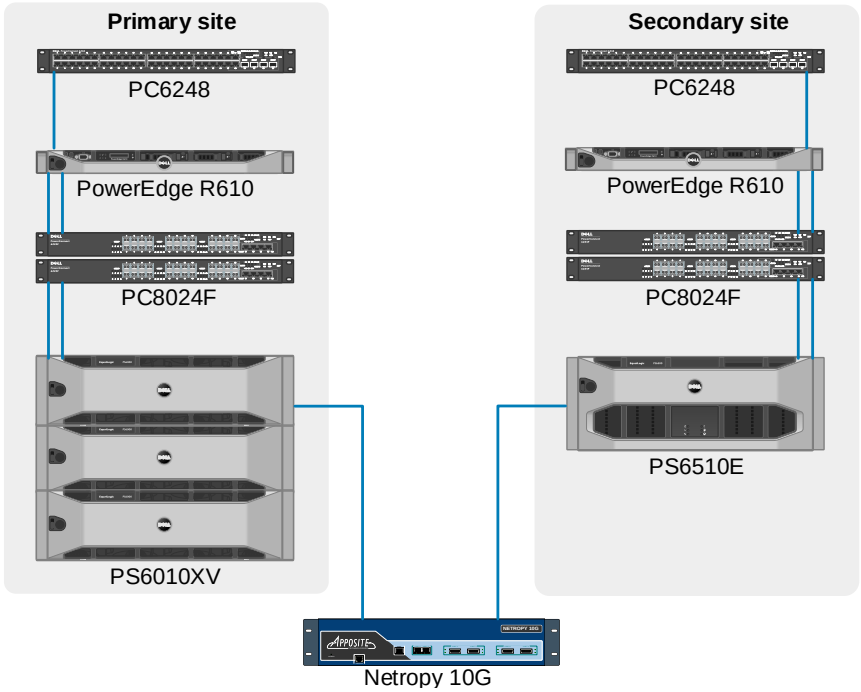

Figure 5 Test topology

5 Test methodology

Because actual WAN links can vary greatly, the Netropy 10G WAN emulator was used to simulate the WAN connection. The Netropy 10G uses dedicated packet processors to ensure precision and repeatability. Besides throttling bandwidth, it can also inject latency, jitter, and packet loss. This allowed simulating the behavior of several different kinds of SAN and WAN links between sites. The Netropy 10G was configured to participate in flow control and to set a queue size (analogous to a router's internal buffer) of 1024 KB (or 1 MB).

Performance was initially measured using a 10 Gb connection path with no WAN emulation. Once the baseline performance across a 10 Gb link was measured, the WAN emulator simulated speeds of 1 Gb, OC3 (155 Mbps), T3 (43.232 Mbps), and T1 (1.544 Mbps) networks.

Table 2 WAN speeds

WAN connection	Speed
10 Gb Ethernet	10 Gbps
1 Gb Ethernet	1 Gbps
OC3	155 Mbps
T3	43.232 Mbps
T1	1.544 Mbps

Actual WAN links can vary in speed or guaranteed bandwidth from provider to provider. Each test case used one of the values shown in Table 2 with the WAN emulator to simulate the bandwidth of a WAN or SAN connection between replication partners. Next, each step of the bandwidth throttling exercise was combined with a random packet loss of 0.1 percent or 1.0 percent and latency (each direction) of 10, 20, and 50ms. The results of each combination were recorded for comparison to the previous runs in which there was no additional packet loss or latency.

Before each replication, I/O was run to the volume to simulate changed data. In some cases, 100% of the data was changed, and in other cases, only a portion of the volume. A replica was manually created each time. When three volumes were replicated simultaneously, a replication group was created allowing replication to start on all three volumes at the same time. This study measured the effects of the following parameters on asynchronous replication:

- RAID level for the primary and secondary groups
- Single volumes compared to multiple volumes
- Thin provisioning
- Connection bandwidth
- Packet loss
- Latency and the TCP window
- Pool configuration

6 Test results and analysis

This section details the test results and analyzes the data to explain the effects of each parameter on PS Series asynchronous replication.

6.1 Effect of RAID level for the primary and secondary groups

This test measured the time it took to replicate a 100 GB volume between primary and secondary sites using the full 10 Gb bandwidth. Then, the RAID level was changed on both primary and secondary groups to compare the effect. First, all of the arrays were set as RAID 10 with all three arrays in the primary site as a single pool, while the single array at the secondary site was in its own pool. Then, they were reconfigured as RAID 50 and the test was repeated.

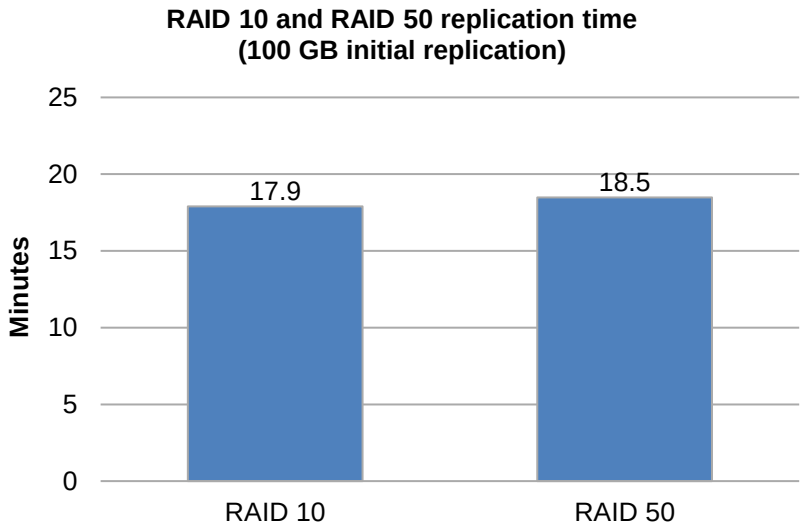

Figure 6 Replication time for RAID10 and RAID50 volumes

The results (see Figure 6) showed that there was a very slight difference (less than 5 percent) for the time it took to replicate a 100 GB volume that resided in a RAID 10 pool compared to a volume residing in a RAID 50 pool, with the RAID 10 volume being slightly faster. Because there was little difference, RAID 50 volumes was chosen for the remainder of the test cases because it offered greater overall storage capacity in the pools.

6.2 Effect of single or multiple volumes

When a single volume is replicated, it will consume less network (WAN) bandwidth than if multiple volumes are being replicated simultaneously. When multiple volumes are configured for replication at the same time, the asynchronous replication process will create an iSCSI session for each volume and transmit the replication data over the available controller ports. When there is adequate bandwidth available between replication partners, this can allow more data to be replicated at once compared to replication of a single volume.

This test compared the time to complete replication for a single 100 GB volume (100% changed data) and the three 100 GB volumes (100% changed data) across different network link speeds. Figure 7 shows that at the

full 10 Gb speed, three times the amount of data is replicated (300 GB compared to 100 GB), but this replication only takes about twice as long for the three-volume configuration. As the available bandwidth across the WAN link is decreased, the amount of time to replicate all three volumes increases to where it is about three times as long. This is because the WAN connection now becomes a limiting factor on the amount of data the network supports.

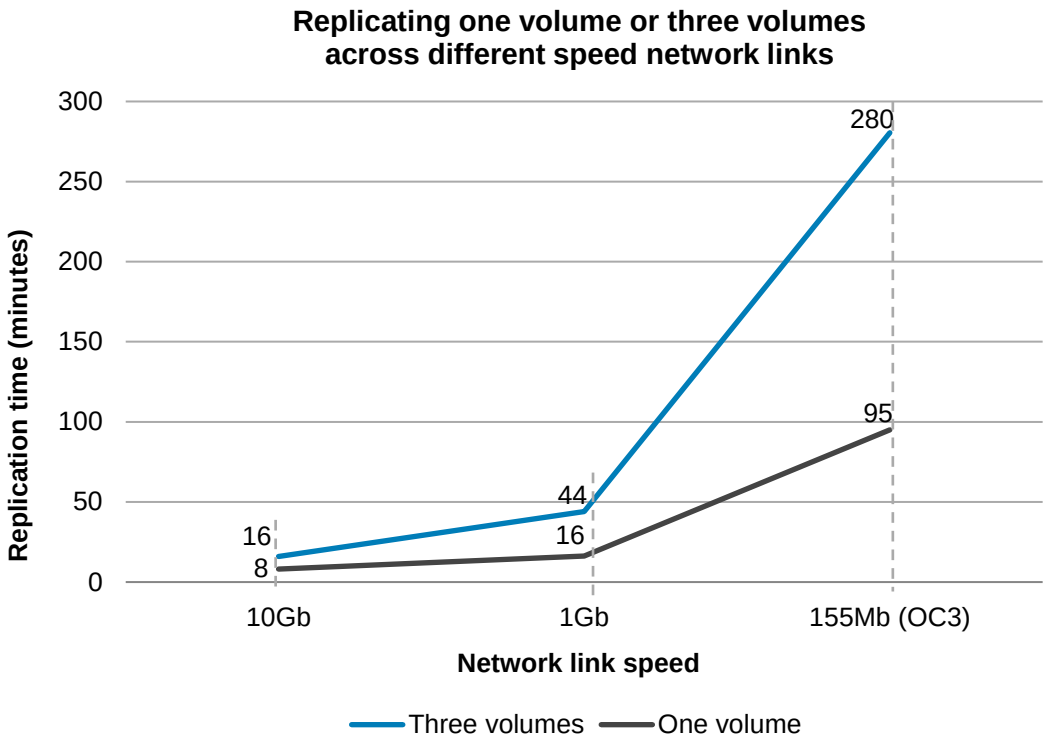

Figure 7 Replicating one volume or three volumes

6.3 Effects of thin provisioning

When a thin-provisioned or a regular volume is replicated, only the actual data in use will be copied to the replication partner. Whether there is a 100 GB volume that is thin provisioned at 10 percent, or a standard 100 GB volume, if only 5 percent of the either volume is filled with data, then only about 5 GB will be copied to the replication partner in either case. As soon as a portion of the volume is actually written for the first time, the array will physically allocate disk space, and therefore these changes will be also copied to the replication partner during the next replication cycle.

For example, when a new Windows NTFS volume is partitioned and formatted, if the **quick format** option is chosen, then only a small portion of the disk is actually modified — the master boot record (MBR) — and very little data will be replicated during the initial synchronization (assuming no additional data has been copied to the volume yet). However, if the full format option is chosen, Windows will attempt to *zero* every part of the new volume, which in turn forces the array to allocate pages from the pool. Even a thin-provisioned volume will now report that 100 percent of the pages have been allocated (in Group Manager). In this case the

volume behaves as if it were a standard volume and the contents of the entire volume will be copied during initial replication.

6.4 Theoretical bandwidth of links and replication time

Because the TCP/IP protocol carries some overhead, the speed of the network alone provides insufficient information to estimate the time it will take for replication. In a typical SAN environment, Dell EMC recommends using Jumbo Frames (9000 bytes) to get maximum performance. However, because most WAN routers or other long-distance connectivity does not support Jumbo Frames, we would typically consider the standard MTU of 1500 bytes in our calculations. Use the following calculations to calculate an estimate for maximum link throughput:

Maximum throughput (MB/sec) = [WAN link speed (Mb/sec)/8 bits per byte] x 93% protocol efficiency

Use the estimate for throughput to estimate replication time for a given amount of data:

Replication time = volume size (MB) / throughput

Considering an OC3 link, which is rated at 155 Mbps, use the following:

Maximum throughput in MB/sec = 18.02 MB/sec

Time to transmit 100GB volume = 102400 / 18.02 = 5683 sec (~95 minutes)

This result closely correlates with the actual time measured in the lab test, as shown for the OC3 (155Mb/sec) link speed in Figure 8. These sample calculations assume that all of the link bandwidth is available.

6.5 Bandwidth effects

Asynchronous replication is affected by the speed of the link between replication partners. When replicating within the same data center or between buildings, the available bandwidth may be equal to the full speed of the SAN. In other cases, a slower WAN link may be utilized. Figure 8 shows the effect that network link speed can have on the time it takes to complete replication.

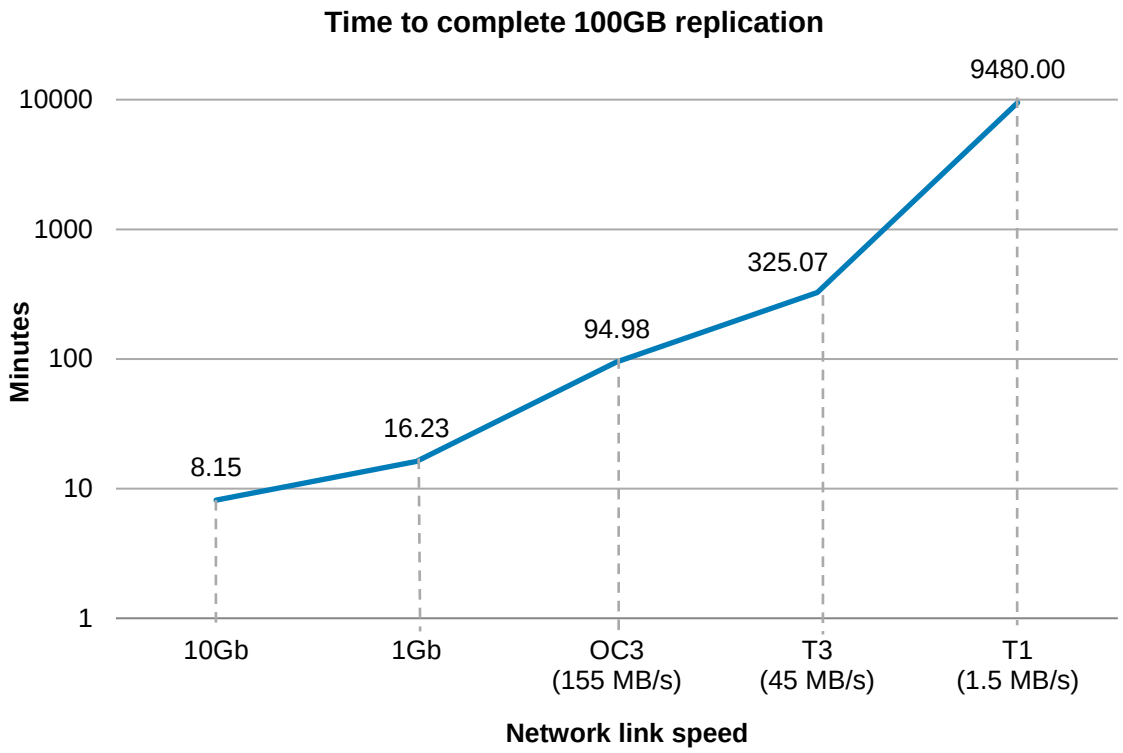

Figure 8 Effects of WAN speed on time to complete 100GB replication

This shows that the time it takes to replicate 100 GB of data doubles when the network is reduced from 10 Gb to 1 Gb. Reducing the network link to OC3 causes the replication time to increase to more than 11 times the speed at 10 Gb. When the network speed reduces to T1 rates, replication time increases to 9480 minutes, or more than 6.5 days. Clearly, a single T1 network link may not be adequate if large amounts of data need to be replicated.

6.6 Packet loss effects

Asynchronous replication is also affected by the quality of the link between replication partners. If a link is dropping packets, the asynchronous replication processes will have to resend those segments. When packets are unacknowledged (lost), TCP/IP protocol invokes an algorithm known as *slow start* (see RFC 5681, TCP Congestion Control). Slow start is part of a normal congestion control strategy to avoid sending more data than the network or other devices are capable of handling. However, if too many packets are dropped and slow start is invoked too often, it will affect the throughput of the network and slow down replication.

Figure 9 demonstrates that the addition of a small amount of random packet loss (1/10th of a percentage) caused only a slight variation in the time it took to replicate 10 GB of modified data. However, one percent of random packet loss was added, replication time doubled. Because packet loss can have a significant effect on replication time, it is important to monitor the quality of the link between sites.

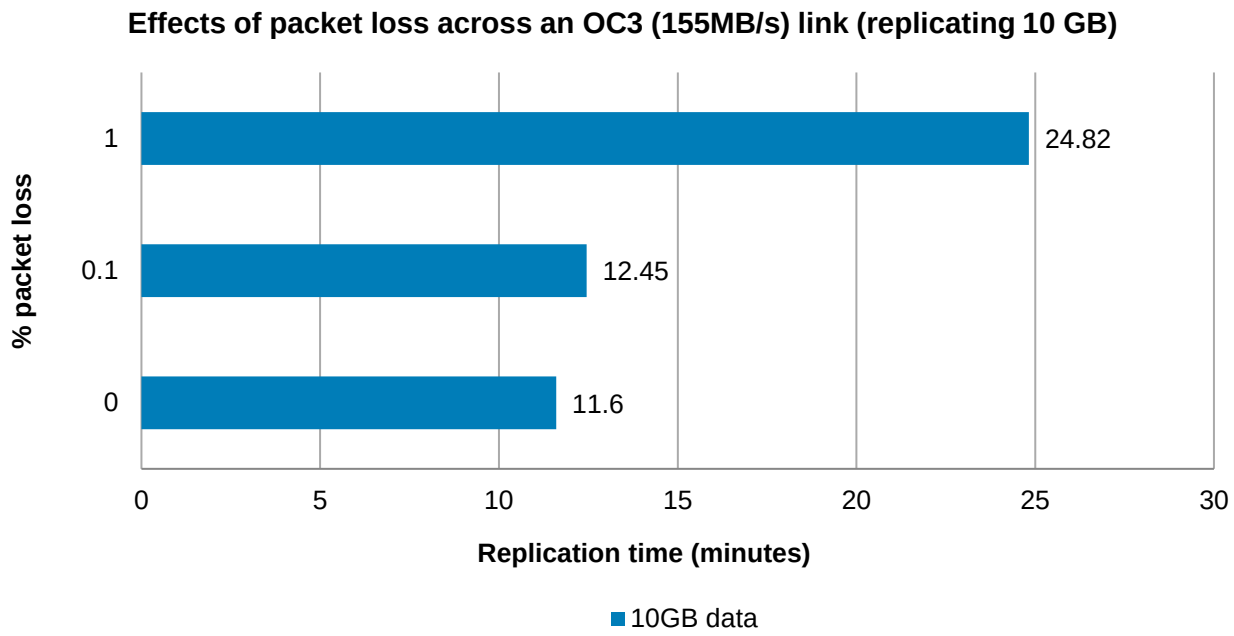

Figure 9 Effects of packet loss

SAN Headquarters can be used to monitor retransmission rates and set alerts when high levels of retransmits are detected. In a normal and healthy iSCSI SAN, the percentage of retransmits should remain below 0.5 percent. While there are other contributors to the cause of retransmits, a *noisy* WAN link or a misconfigured device somewhere in the path could cause packet loss and lead to slow replication performance.

6.7 Latency and TCP window size effects

For the purpose of this discussion, latency is how long it takes a packet of data to travel from one point to another across the network. Latency is inherent in any network, including an iSCSI-based SAN. Typically, iSCSI SAN latencies are quite small, and usually measured in microseconds or milliseconds. Several factors affect latency in a storage system, including the time it takes to retrieve data off a hard disk, cache hit ratios, and the speed of the connection to a host system. In our tests, we wanted to understand how the latency caused by a WAN link might affect asynchronous replication.

Distance will add latency — the round trip (send and acknowledgement) for a data packet exchange will take longer as distance separates the devices. The maximum possible speed that data can travel is equivalent to the speed of light in a vacuum: 186,282 miles per second. In practice, data speed is attenuated by the cable medium. In the case of fiber optic cables, the glass fiber slows the light signal down to approximately 124,000 miles per second or about 199,560 km/second. Table 3 shows the approximate distance a data packet can travel in a given time across fiber optic cables (values in the table should be doubled for round-trip calculations):

Table 3 Latency induced by distance

Time	**Approximate distance traveled through fiber optic cables**	
Milliseconds	Miles	Kilometers
1	124	200
10	1,240	1,996
20	2,480	4,007
50	6,200	9,978
100	12,400	19,956

Of course, there are other factors that affect a network packet's round-trip time. For example, the amount and type of network switching and routing devices it encounters along the way will all contribute incrementally to the overall link latency. In many cases, the device latencies will add more to overall link latency than distance alone.

In the lab test environment, a WAN emulator was used to inject various amounts of latency into the replication network links. A volume was then replicated across the link, and the time it took to complete the replication was measured. These tests used a 10 GB volume and it was ensured that the entire volume was overwritten each time (100 percent of the data had changed) before creating a new replica.

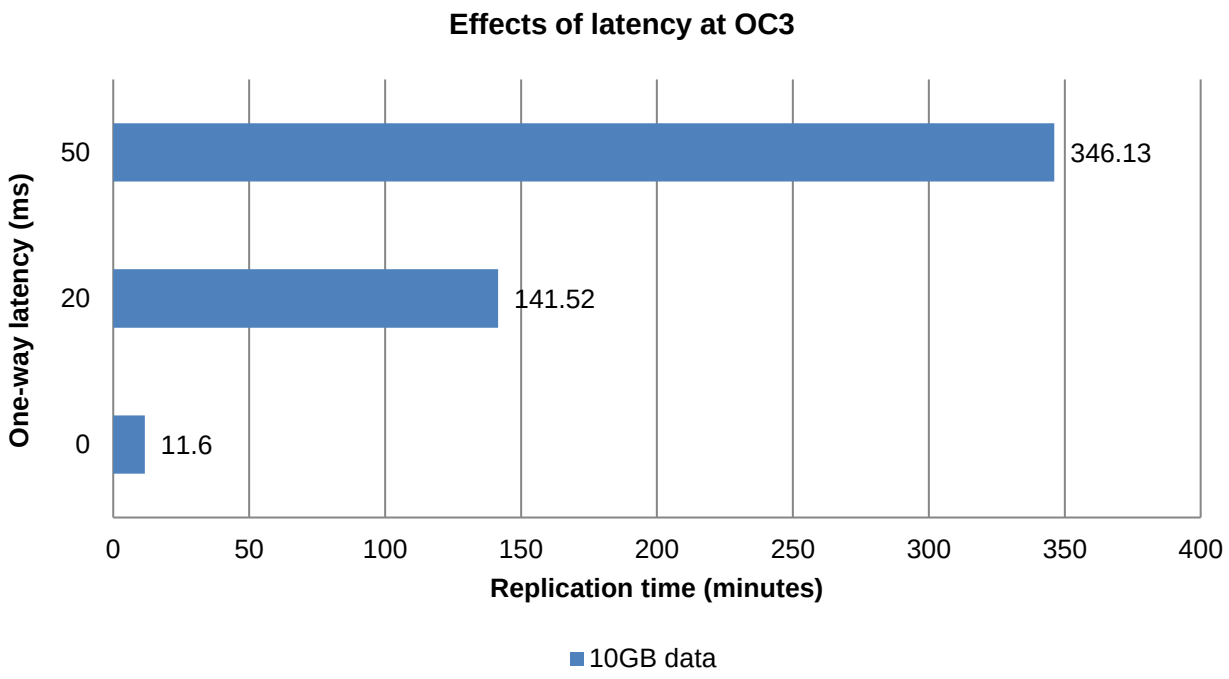

Figure 10 Effect of latency on replication time

Figure 10 shows the time it took to replicate 10 GB of data across an OC3 (155 Mbps) WAN link for three different simulated link latencies. The results clearly show a significant impact on the performance of replication across the WAN link. When 20 ms of latency was added in each direction (40 ms round trip), the replication time increased by a factor of 12. When 50 ms of latency was added (100ms round trip), the replication time increased by a factor of over 30.

When a device sends TCP data packets over a network, it will attempt to send as many as possible within its *TCP window* size. Once it reaches its TCP window size limit, it will stop transmitting packets and wait until it receives an acknowledgement back from the receiving device. After it receives acknowledgement, it can then send more packets if necessary. The maximum receive window size for standard TCP is 65,535 bytes. TCP also allows the devices to scale (increase) the size of this window as high as 1 GB (see RFC 1323, TCP Extensions for High Performance). With larger window sizes, it is possible to have more packets *in flight* at a given point in time. This window size scaling feature allows TCP to optimize data transmission over a variety of link conditions and possibly improve its throughput.

As latency grows, it becomes more likely that devices will have to wait while data is in flight. Since they are waiting, and not actually sending or receiving data (the packets are still traveling somewhere in the middle), their overall throughput efficiency can decrease. Increasing the size of the TCP window allows more data to be placed on the wire at time, thus keeping the packet exchange going closer to 100 percent of what is theoretically possible. Of course, this also represents a greater risk—if more data is outstanding and an acknowledgement is never received (a packet is lost somewhere or times out), then the sending device will have to retransmit all of the data that was not acknowledged. If this scenario occurs too often, then performance may be degraded.

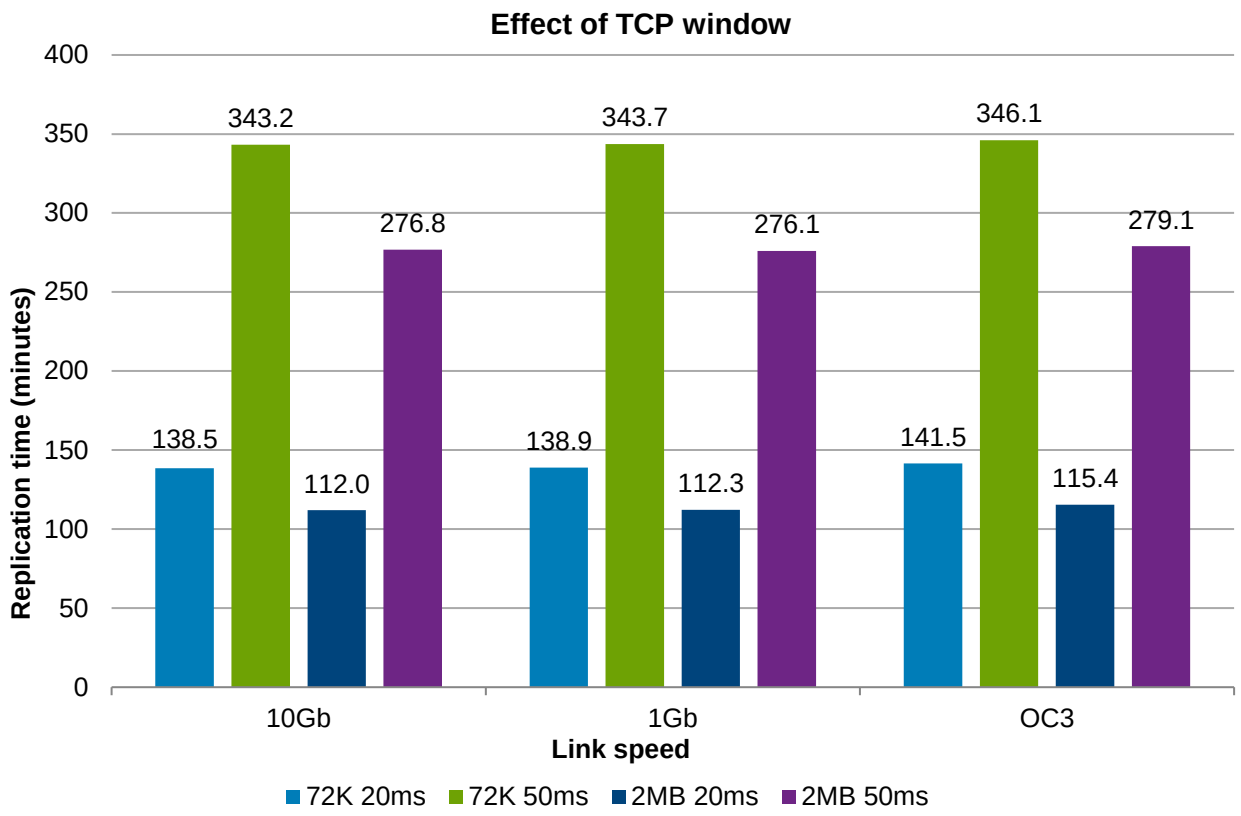

Figure 11 Effects of TCP window

Figure 11 shows the effect of increasing the size of the TCP window from 72K to 2MB when replicating across different WAN links speeds with 20ms or 50ms link latency. Based on these results, the following conclusions can be made:

- **Higher connection latency decreases maximum replication throughput.** As expected, when increasing the latency from 20ms to 50ms, the time to complete the replication data transfer increased significantly for all test cases.
- **Larger TCP window sizes can increase replication throughput for high speed, high latency networks.** When increasing the size of the TCP window from 72K to 2MB, an 18.9 percent average decrease was measured in the replication time for 20ms link latency, and a 19.5 percent average decrease in replication time for 50ms link latency.
- **For very low latency connections, the beneficial effect of increasing TCP window size will be minimal.** The same tests were run with zero added latency across the replication link (results not shown in Figure 11). For zero added latency, there was no measurable difference in throughput performance between 72K and 2MB TCP window sizes.

Note: Changing the TCP window size setting on PS Series controllers to non-optimal values can have negative side effects. The ability to change this setting is currently not supported using the PS Series Group Manager. Customers interested in changing TCP window size settings should contact PS Series Technical Support for assistance.

6.8 Pool configuration effects

The configuration of groups and pools may also have an effect on how quickly replication occurs. For example, the number of array members in a pool affects how many array members a volume may be distributed across. Because of the scale-out architecture of PS Series arrays, a group that contains multiple members has the potential to move data faster than a group with only a single member. However, if a slow link (slower than the SAN) exists between replication partners, then the performance impact of the link bandwidth will override any benefit that having multiple pool members provides. The test results in Figure 12 illustrate this behavior. When replicating across 1 Gb/s or 10 Gb/s links, there was a significant decrease in time to complete replication when the volume is hosted on a three-member pool when compared to a single-member pool. As the network bandwidth between replication partners decreases, the difference becomes negligible.

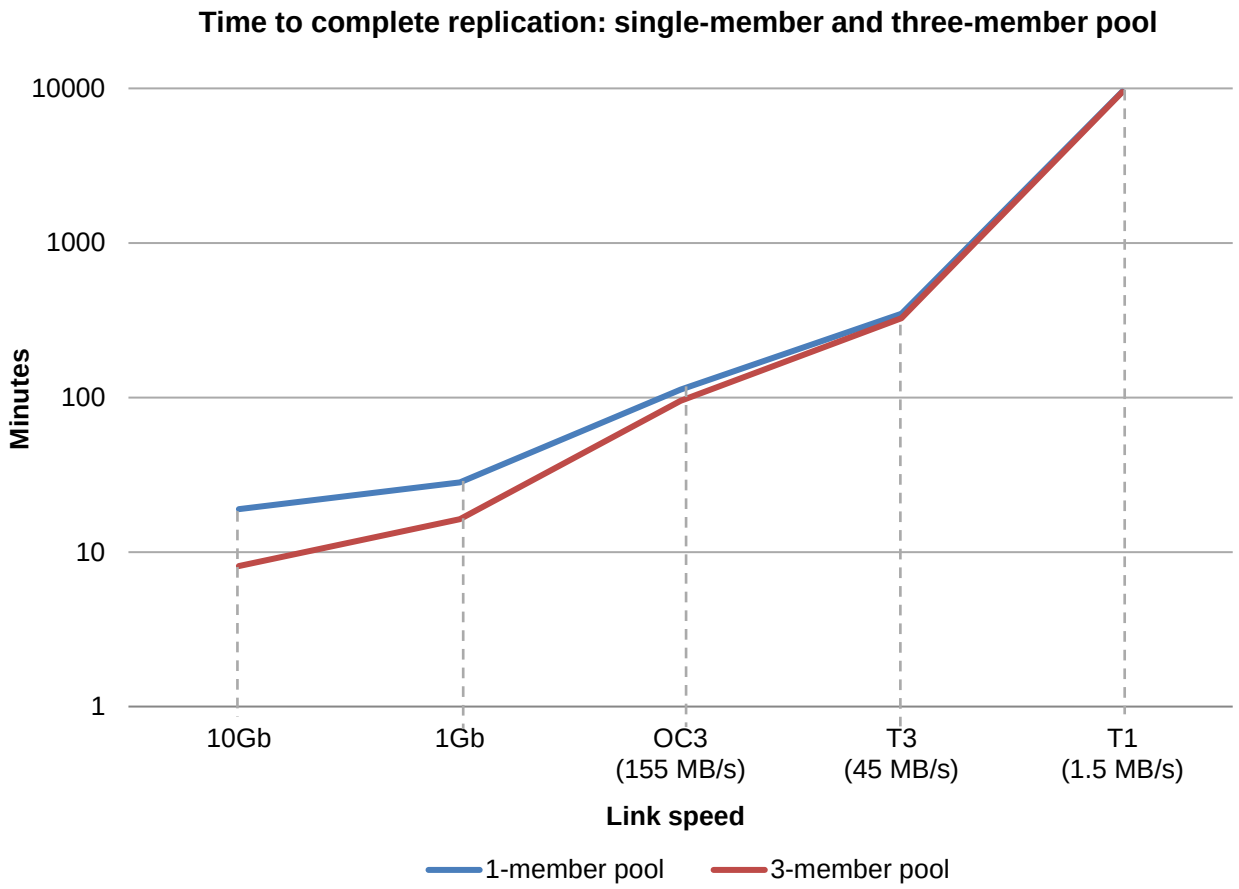

Figure 12 Effect of number of pool members on time to complete replication

6.9 Server I/O effects

Asynchronous replication is a background process, and as such is designed to have little impact on the performance of servers (hosts) connected to the SAN. On the other hand, that means that server I/O can affect the performance of replication. If there is a heavier workload from the attached hosts, the arrays may devote fewer resources to replication, which could cause replication times to be longer than when there is a lighter workload from the attached hosts.

Server I/O is another factor that you should take into account when considering recovery time objective (RTO) or recovery point objective (RPO) times. By using SAN Headquarters to monitor the I/O load and replication times, this can ensure that there is enough *headroom* built into the SAN to accommodate a difference in workload from day to day or during different parts of the day if multiple replicas are created.

7 Best practices for planning and design

7.1 Recovery time objective (RTO) and recovery point objective (RPO)

In simple terms, as it applies to replication and disaster recovery, RTO is how long a business can get by without a particular system or application in the event of a disaster. An RTO of 24 hours implies that after a disaster, the system or data needs to be online and available again within 24 hours.

The term RPO is generally used to describe the *acceptable loss* or the time gap of data to lose in the event of a disaster. A business must decide if it is acceptable to lose any data in the event of a disaster, and if so, how much can be lost without a significant impact to the business. For some applications or systems, this may be as long as 24 hours, while for others this may be as little as 15 minutes, or even zero.

The RPO and RTO requirements of a business must be met when implementing a disaster recovery plan. The asynchronous replication feature included with PS Series arrays allow data to be replicated across a SAN or WAN link to another PS Series array. How often a replica is synchronized must align with the RPO. In other words, if a volume is replicated and the changes are synchronized every four hours, then the most amount of data that would be lost if a disaster occurred at the primary site is less than four hours of data, because that is the interval between replicas. The time it would take to recover the data and make it available again should align with the RTO.

It is not uncommon to have different RPOs for the same data or system. One RPO requirement may apply to disaster recovery in which an entire site is offline, while there may be a different RPO requirement for local recovery. In such cases, it is common to use a combination of replication and array-based volume snapshots or volume clones to meet those requirements. In other cases, a combination of asynchronous replication and a host-based snapshot or replication management software product may be used. In either case, asynchronous replication serves as part of the total solution design required to meet the recovery objectives of the business.

7.2 The network

Replication uses ICMP and TCP port 3260 (standard iSCSI). These ports must remain open across the WAN link for replication to perform properly. Any switches, routers, and firewalls between the two sites must be configured as such to allow the arrays to communicate. The network should also be secured by using firewalls, VPN, encryption, or other means. However, firewalls must not use NAT or PAT between the sites for replication traffic.

A slow, underperforming network can greatly affect the speed of replication. If multiple replicas are scheduled simultaneously, then the arrays will attempt multiple streams through that slow link, which could increase congestion and cause abnormally slow replication performance. To prevent overloading the WAN link, the storage administrator must understand how much data will need to be transmitted through the WAN link as well as what the conditions of the link look like (such as latency or packet loss).

7.3 Tuning the WAN link

When replicating across a WAN, the data packets will probably be traveling through a router. A router generally has a memory buffer that stores incoming packets so that they can be processed and forwarded. If the WAN link is congested (or too small) and it becomes a bottleneck, this can cause incoming packets to fill up the memory buffer on the router. Eventually the router may be forced to discard (drop) incoming packets until it frees up space in the memory buffer. This in turn causes the sending side to timeout because the receiving side will never acknowledge receipt of the frames that were discarded.

If Group Manager or SAN Headquarters reports a high occurrence of retransmits during replication, this could be due to an overloaded WAN link that is dropping packets. One course of action would be to monitor and adjust the buffers in the router. Or, it may be necessary to implement Quality of Service (QoS) or Class of Service (CoS) on the router or any upstream switches to reduce the amount of the data flowing into the router. Adjusting router buffer settings is beyond the scope of this paper. Most manufacturers of these devices include the ability to adjust parameters that can affect traffic flow. You should be aware of these capabilities and use them if needed, particularly when replicating over slower speed link paths.

If the WAN link is not dedicated to the storage arrays, then QoS or CoS may be configured on a specific VLAN, the IP addresses of the arrays, or even by port (3260, the default iSCSI port). Of course, any other traffic that shares the link will also need to be managed and may affect the performance of the replication traffic. If possible, you should use a dedicated link for the storage replication traffic.

Although it is also beyond the scope of this paper to discuss in detail, some customers have also utilized WAN optimization products in their WAN links. These products may implement features such as compression, deduplication, or other packet optimization that can help improve the efficiency of a slower WAN link, and therefore decrease the time it takes to replicate changes across these links.

7.4 Planning for storage needs or volume sizes

The default, space-efficient guidelines for sizing replication reserves and delegated space are presented in Table 4. The default values indicated are recommended for most situations unless the actual change rate is well understood.

Table 4 Replication space, default value, and space-efficient value

Replication space	Default value	Space-efficient value
Local reserve (primary group)	No failback snapshot: 100 percent Keep failback snapshot: 200 percent	5% + %Change_Rate 10% + %Change_Rate
Replica reserve (secondary group)	200 percent (to ensure there is adequate space for the last replica and any replica in progress)	105% + %Change_Rate x (# of Replicas – 1)
Delegated space (secondary group for all replicas coming from a single group)	Must be large enough to hold the sum of all replica reserve sizes for all volumes replicating to that group	Monitor change rate, adjust to lower than default value, and continue monitoring

The system defaults are set conservatively so that replication will continue in all cases, even if the volume contents change completely from one replication to the next.

When using thin provisioning, be aware that percentages are based on the internally allocated size of the volume, rather than based on the size of the volume as reported to the server.

The local reserve default allows for the system to continue replicating even if 100 percent of the previously-written data was changed before the new replica can be completed. If fast failback is enabled for the volume, then the default setting provides an additional 100 percent for keeping a snapshot of the data that was previously replicated, even if the data was changed 100 percent between each point in time.

Similarly, the default replica reserve is set to 200 percent so that a complete copy can be stored on the remote site even as another completely changed copy is replicated. In this case, even if the volume changes 100 percent, there would never be a case where replication cannot proceed, and regardless of how the system is configured there will always be at least one replica copy stored on the remote site.

Dell EMC recommends monitoring the use of these reserves through Group Manager and SAN Headquarters to determine the optimal settings for your environment. After monitoring, make adjustments as needed to ensure the number of retained replicas meets RPO and RTO requirements for the business, and the replication reserve disk space is being used most efficiently.

Dell EMC also strongly recommends that overall free pool space does not fall below the following limit (whichever is smaller for the pool):

- 5 percent of the total pool space
- 100 GB multiplied by the number of pool members

If the utilization exceeds these levels, steps should be taken promptly to make more space available. This will ensure multi-member page balancing, performance balancing, vacate, snapshot, replication, and thin provisioning operations perform optimally. If choosing to take advantage of the **Borrow from Free Space**[1] replication option, this requires at least 10 percent free pool space on the primary side or the borrow option will be disabled. If this option is selected, free pool space will be temporarily used if there is not enough free local replica reserve space to complete a replication.

Although some applications perform a consistent number of volume writes, others have a workloads that change daily. Therefore, one replication operation might transfer little data and complete quickly, while another replication might transfer a large amount of data and take a long time.

In some cases, a volume might appear to have few changes, but the transferred data is relatively large. Random writes to a volume can result in a large amount of transferred data, even if the actual data changes are small. Some disk operations, such as defragmenting a disk or reorganizing a database can also increase the amount of transferred data because these operations modify the source volume.

[1] This setting allows temporary borrowing of free pool space if the local reserve size is inadequate. At least 10% free pool space must be available for this setting to take effect.

Figure 13 shows the remote replicas that are available for a volume as displayed in Group Manager from the primary group. In this case, there are five replicas of a volume on the remote storage system. The number of replicas retained on the remote partner system is determined by the size of the volumes being replicated and the amount of changed data that is replicated, the size of the replica reserve, and the delegated space. The total size of all replicas cannot exceed the total replica reserve space available (in Figure 13, 840 GB is allocated).

Navigation tip: **Group Manager GUI** (primary) > **Replication** > **Outbound Replicas** > select volume

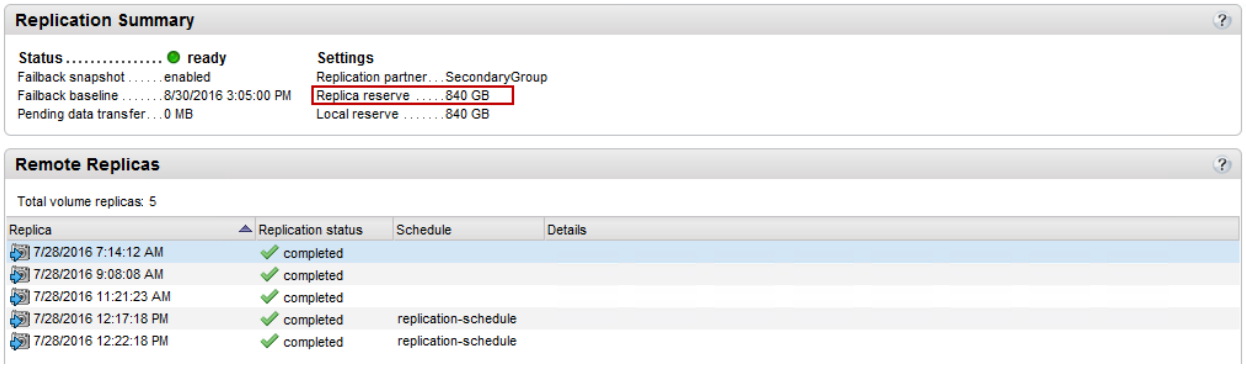

Figure 13 Outbound replicas for repl-vol1 on the PrimaryGroup

7.5 Initial capacity planning and sizing example

For the initial sizing of a solution for PS Series storage, the concepts described in this document may be more easily understood with a real-world example. Best practice sizing techniques are by design conservative enough to account for the needed space on the primary and secondary. The local replication reserve should also be large enough to handle the in-flight replications, as well as the initial volume capacity. The secondary will need similar space considerations.

So for a simple and conservative rule, 200 percent of the space per volume should be allocated on the secondary delegated space as well as the primary local replication reserve. This is not a hard and fast rule and the space efficient approach may be taken as indicated in Table 4. However, to arrive at the correct future change rate, detailed analysis and understanding of the applications, growth, and business needs should be considered carefully. The approach presented here will use the more conservative approach to avoid under sizing the solution.

Since replication is at the volume level, we will demonstrate how to size the volumes which will replicate from the primary to a selected secondary group and pool. The pool will need enough space to hold the total of all replicas of the volumes replicating to this secondary group. If replication schedules are used then the number of replicas to retain will also need to be considered.

Navigation tip: **Group Manager GUI** (primary) > **Volumes** > select volume > **Schedule** (tab)

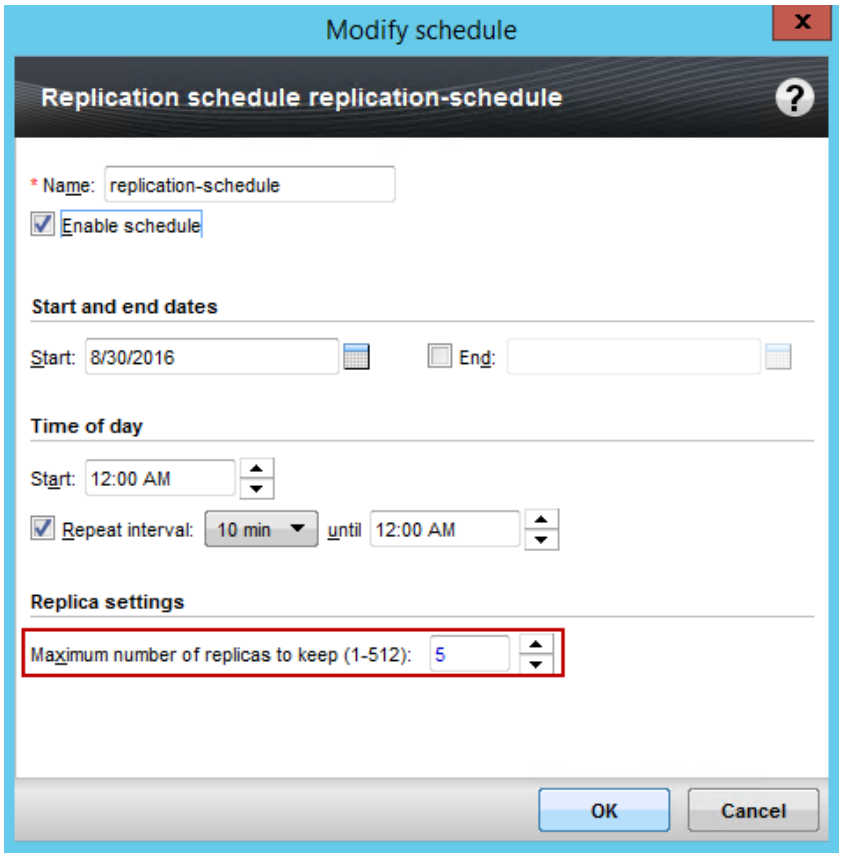

Figure 14 Replication cycle showing the maximum number of replicas to keep

Note: Each replica will contain only the changes between replication cycles.

For this example the *PrimaryGroup* will replicate two volumes to the SecondaryGroup. Both Groups have two pools, this is intentional to demonstrate the improvements with firmware v8 and higher which allows for multiple destination pools from a primary group.

The two volumes we will replicate have 420 GB as the reported size.

Navigation tip: **Group Manager GUI** (primary) > **Volumes**

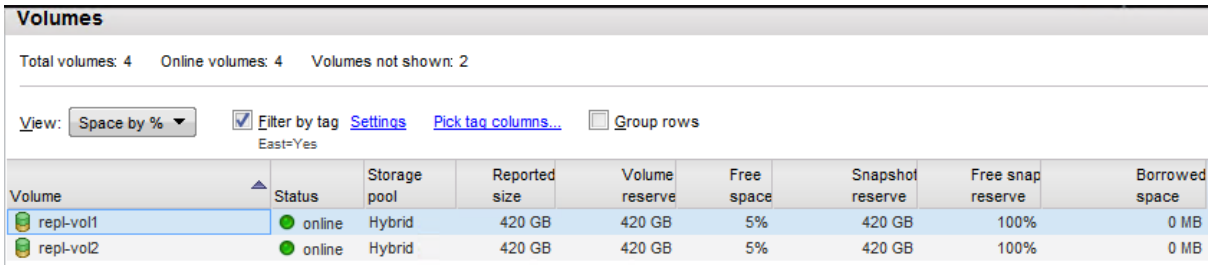

Figure 15 Volumes on the primary group that will be sized for replication.

7.5.1 Primary group space considerations

The PrimaryGroup is a group with two PS6210 members and two pools (in this example only the volumes in the hybrid pool will be used).

From a best practice perspective, sizing to the volume total reported size is a way to ensure that secondary group is sized to accommodate the volume potential growth. The particular needs of the environment may override this general rule, however the overall growth potential should still be considered.

Navigation tip: **Group Manager GUI** (primary) > **Volumes** > (select volume) > **Status** (tab)

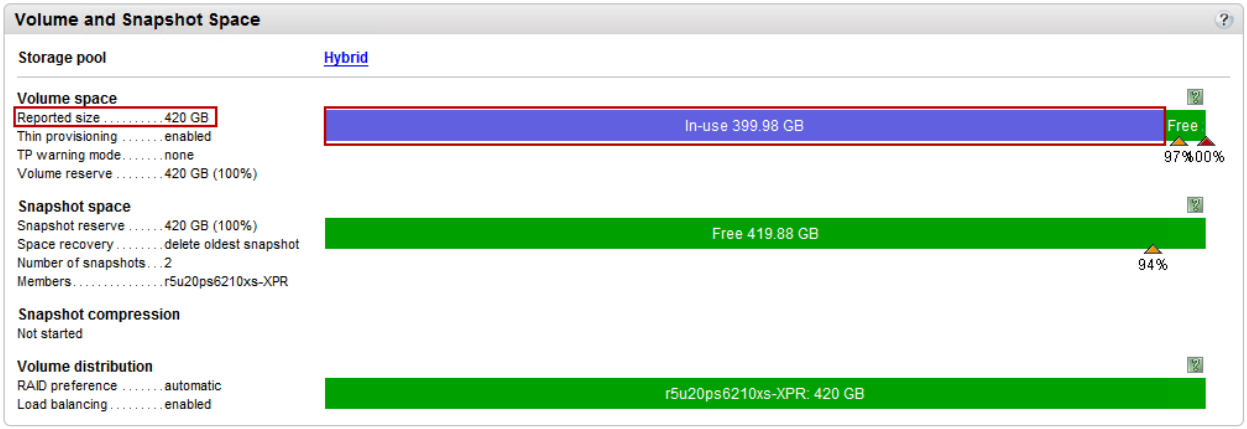

Figure 16 Example of using the reported size instead of the in-use space.

The total reported size is 840 GB, which includes two volumes each with 420 GB: repl-vol1 (420 GB) and repl-vol2 (420 GB).

The total space in use is also important to understand, since this will be the total amount of data initially replicated to the secondary group. **In use** for this example includes two volumes each around 400 GB. The first replication will synchronize 800 GB (repl-vol1 in use is ~400 GB and repl-vol2 in use is ~400 GB).

Navigation tip: **SAN Headquarters** (primary) > **Capacity** > **Pools** (menu select pool)

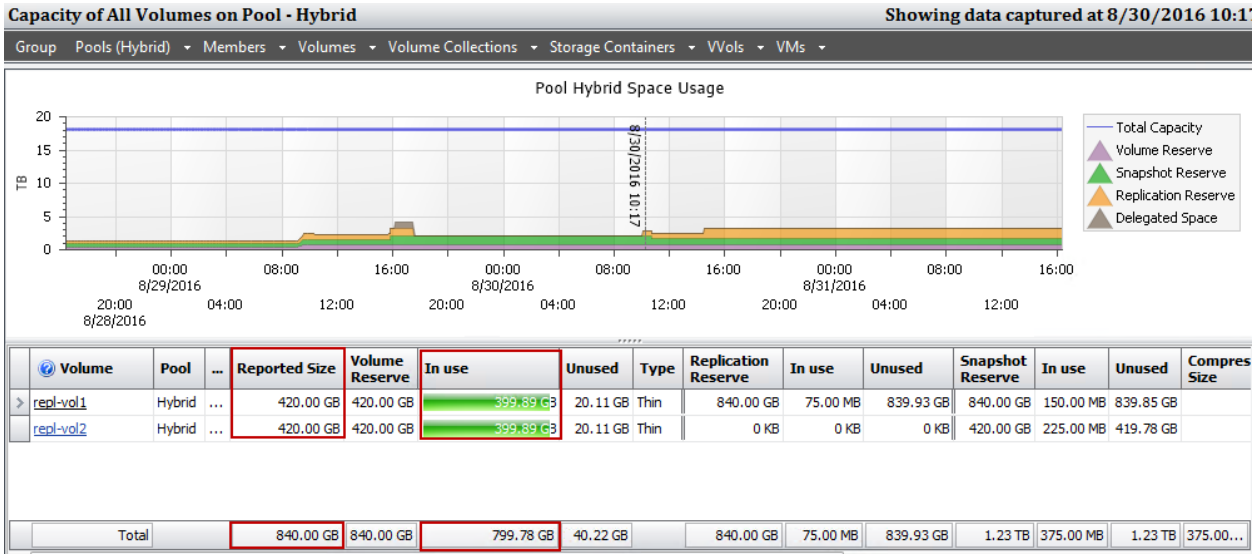

Figure 17 Reported Size and In use capacities for the two volumes

The local replication reserve should accommodate both the original usable space in the volume and the maximum change rate. In addition, typically the fast-failback snapshot is kept, which will keep the most complete replica to allow the partners to synchronize back only the changes that occurred during disaster recovery (promote to volume on the secondary). For these reasons, 200 percent of the reported size is recommended for each volume.

Applying this to our example, 200 percent of 420 GB equals 840 GB x 2 volumes, or 1680 GB total for local reserves.

This space will need to be accounted for locally. Keep in mind this is in addition to any local snapshots that may be needed on the primary group. Since there is already 840 GB of reported size on the primary group, it needs to be verified that an additional 840 GB is available on the PrimaryGroup's hybrid pool.

In summary, the PrimaryGroup will need the following:

Replication local reserve for all volumes: Total of 1.64 TB or 1680 GB (200 percent of volume reported size). The following details show the replication configuration for each volume.

- repl-vol1:
  - Total replica reserve of 200%: 840 GB (this will be reserved on the SecondaryGroup as part of the delegated space)
  - Local replication reserve of 200%: 840 GB
- repl-vol2:
  - Total replica reserve of 200%: 840 GB (this will be reserved on the SecondaryGroup as part of the delegated space)
  - Local replication reserve of 200%: 840 GB

Navigation tip: **SAN Headquarters** (primary) > **Capacity** > **Pools** (menu select pool)

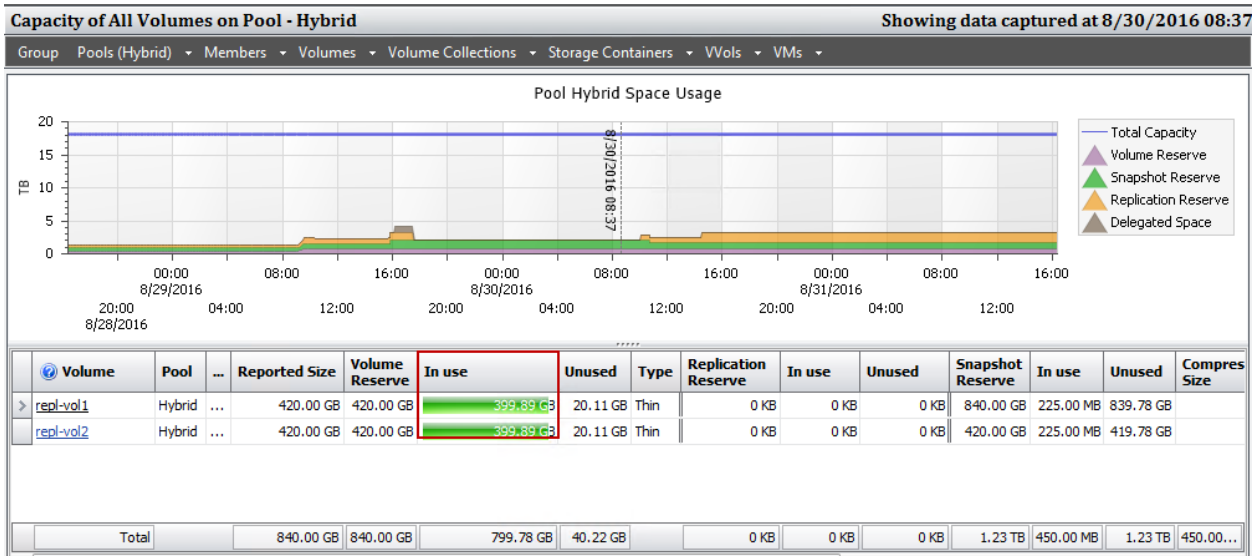

Figure 18 SAN Headquarters showing the In use space for each volume

As a demonstration, repl-vol1 will be configured for replication with the space reservation as indicated in the following screenshot. Both volumes will be configured this way.

Navigation tip: **Group Manager GUI** (primary) > **Volumes** > (select volume) > **Activities** > **Modify replication settings**

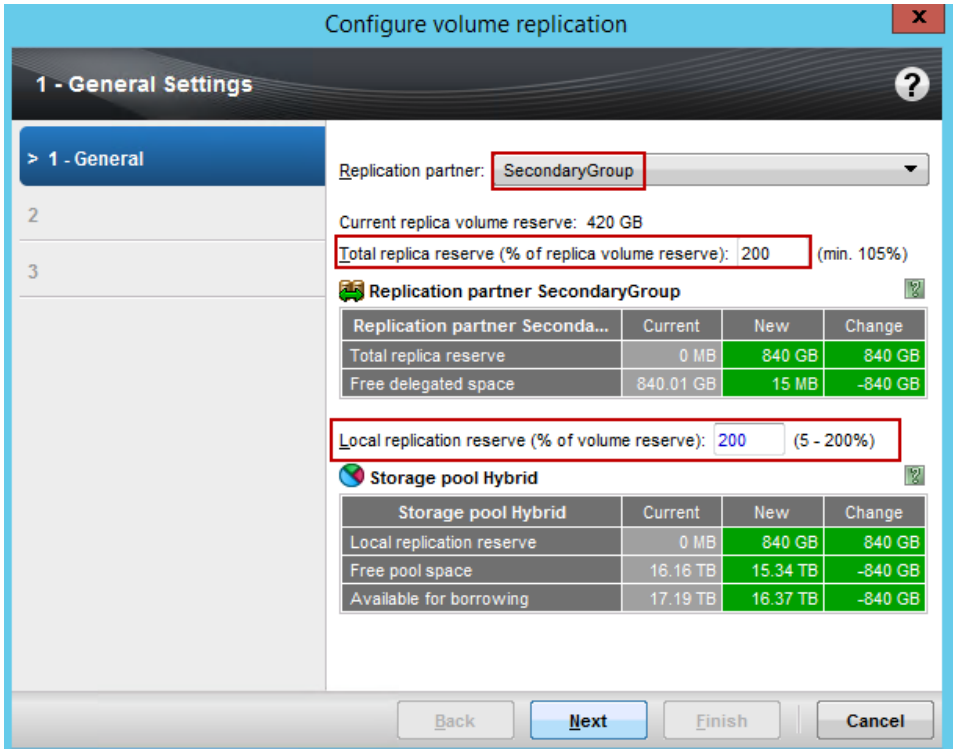

Figure 19 Repl-vol1 replication settings, total replica reserve, and local replication reserve are set to 200% (840GB)

After allocating the replication reserve, the PrimaryGroup will be left with 14.92 TB of free space in the hybrid pool where these volumes reside and will be sufficient for replicating the two volumes.

Navigation tip: **SAN Headquarters** (primary) > **Capacity** > **Pools** (menu select pool)

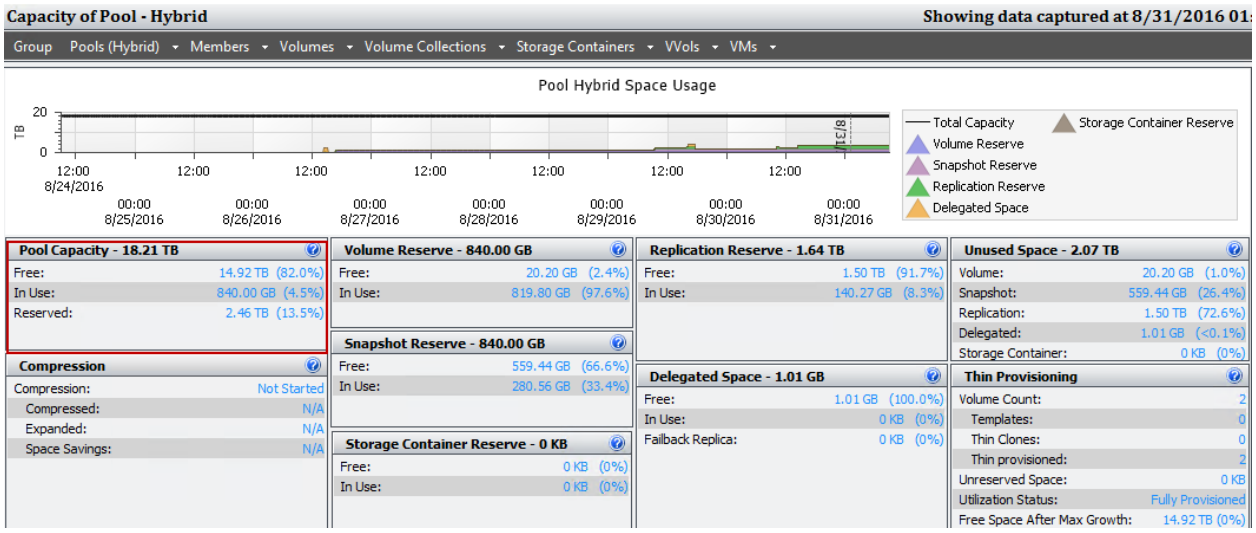

Figure 20 SAN Headquarters view: PrimaryGroup has 14.92 TB free after allocating replication reserve of 1.64 TB

Calculating additional space that may be needed from scheduled replications should take into account the number of replicas that are to be retained. The following example shows several replications occurring on one volume after a schedule is defined.

Navigation tip: **Group Manager GUI** (primary) > **Volumes** > (select volume) > **Replication** (tab)

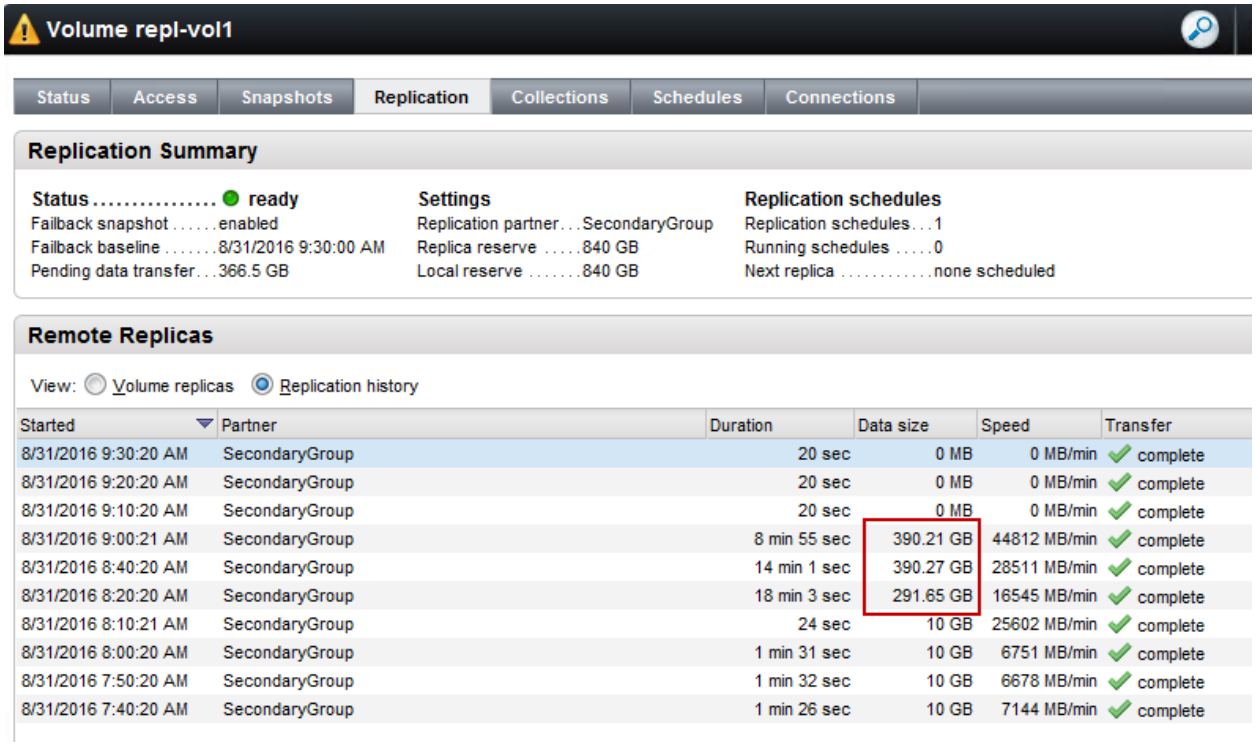

Figure 21 Replication history showing the amount of data transfer between scheduled replications.

On the occasion where large amounts of data are replicated, the local reserve may need to temporarily borrow space. The replica keep count as shown in Figure 14 may need to be lowered to keep the local volume replication reserve from being exceeded.

In addition, during the replication, the amount of space borrowed may be observed in the Group Manager GUI.

Navigation tip: **Group Manager GUI** (primary) > **Volumes** > (select volume) > **Status** (tab)

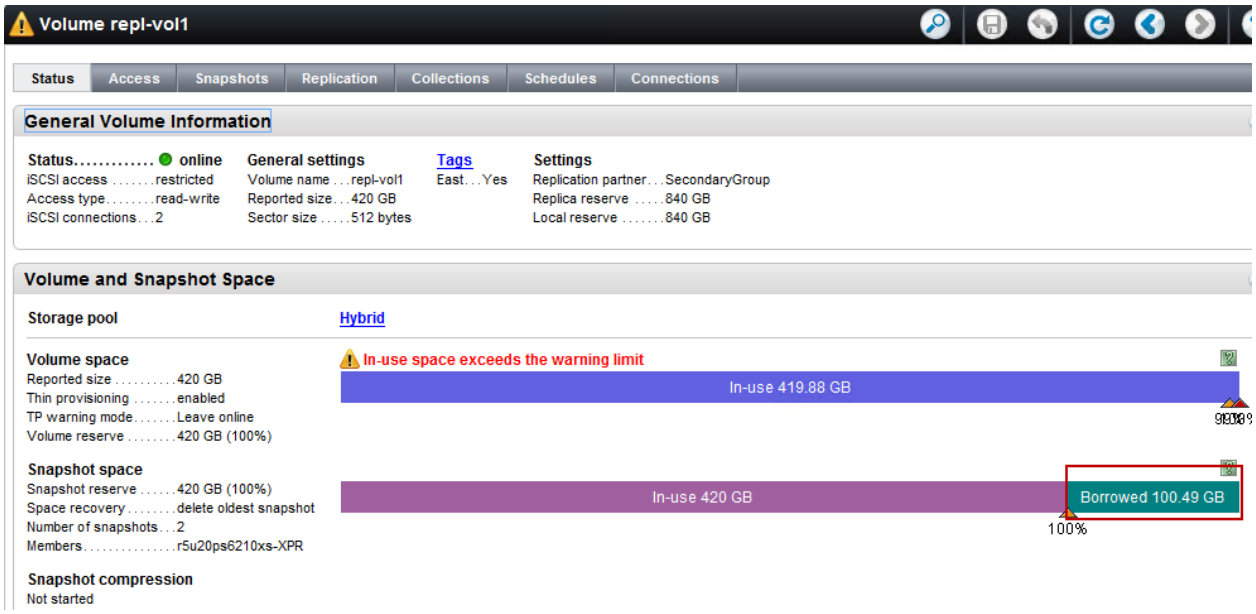

Figure 22 Temporary borrowing of data from free space during replication.

Replication borrowing was introduced in firmware v8 and allows for temporary use of available space. For instance, replication may need to use beyond the local reserve by borrowing the space from one of the previous replicas or local snapshots. After the replication is complete the borrowed space will be freed up if no longer needed.

Navigation tip: **Group Manager GUI** (primary) > **Volumes** > (select volume) > **Status** (tab)

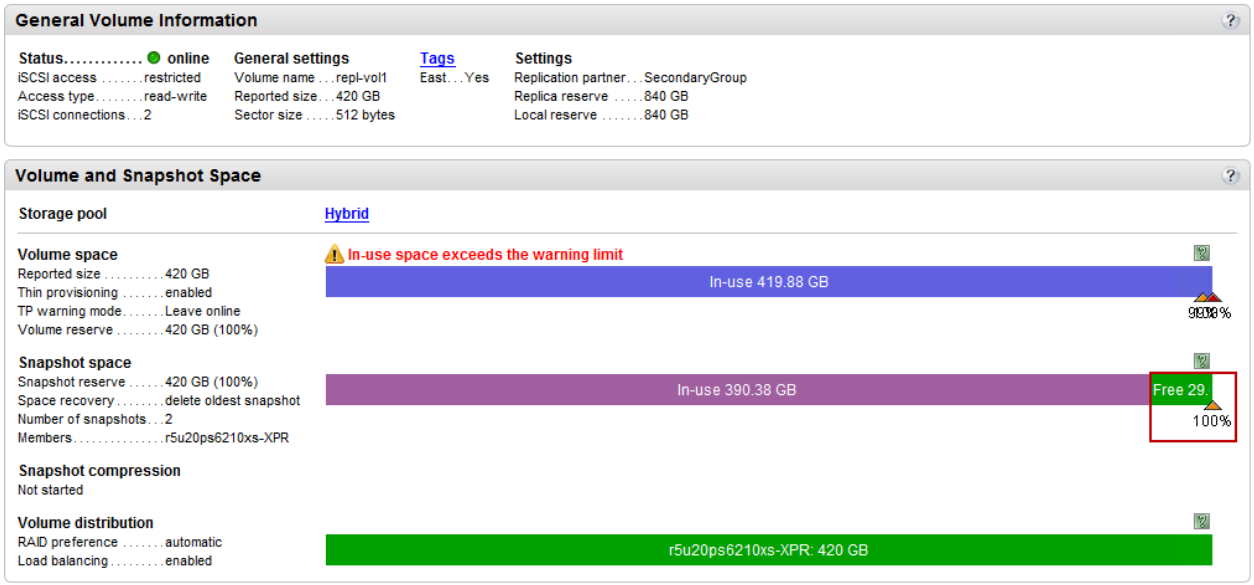

Figure 23 Borrowed space is free after the pending changes are replicated

7.5.2 Secondary group considerations

Secondary space considerations are typically the most important piece to sizing for replication, since an existing array may need to be used or a new PS Series array will need to be acquired. For this exercise, the assumption is that the secondary group will only be used as a replication target. The secondary group will need to be sized to accommodate all the volume replications as well as the number of replicas to keep.

The main consideration on the target will be the amount of delegated space that will be needed. Delegated space will hold all replicas and their changes between replication cycles. Typically, to size to this, you would need to know the total amount of primary data to replicate and some understanding of the change rate along with the number of replicas (snapshots) to keep. If you have a single 200 GB volume to replicate, then 400 GB should be allocated to delegated space to account for the possibility of 100 percent changes. In addition if multiple replicas are to be kept, then the total changes those snapshots contain should also be added into the calculation. For a 200 GB volume with two replicas (each with 25 percent changes kept), the delegated space would need to add another 100 GB for a total of 500 GB ((200% x 200 GB) + 100 GB of changes).

The delegated space is defined during the replication partner configuration step. For PS Series firmware v8 and higher, multiple pools may participate in the allocation of delegated space. For this example, we will allocate 200 percent of the reported space. For the two 420 GB volumes on the primary, this works out to 1.64 TB.

A PS Series array with at least 1.64 TB of usable space should be used as the secondary. The SecondaryGroup mentioned previously has plenty of usable capacity available.

Navigation tip: **Group Manager GUI** (primary) > **Group** > **Storage Pools**

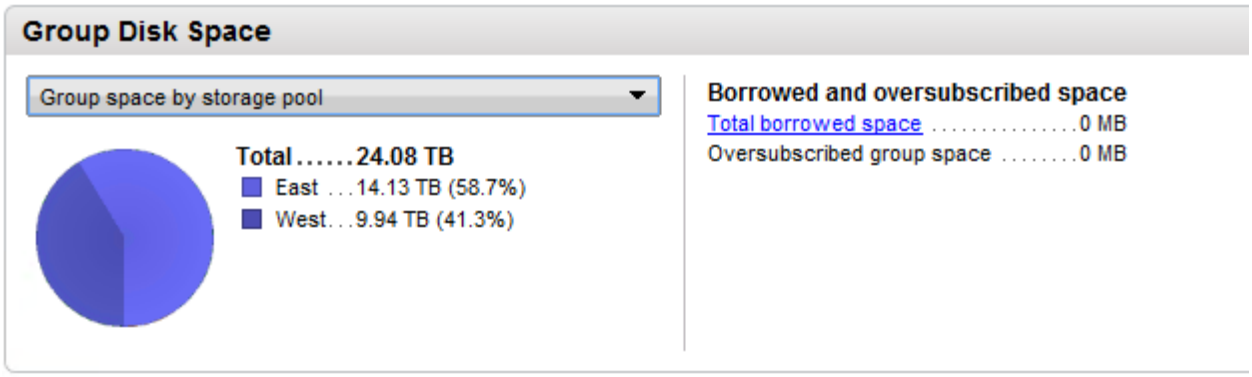

Figure 24 SecondaryGroup usable space by pool

The SecondaryGroup delegated space will be split between pools with 840 GB in the East pool and 840 GB in the West pool.

Total delegated space for the SecondaryGroup is 1.64 TB or 1680 GB (200 percent for each volume).

Navigation tip: **Group Manager GUI** (primary) > **Replication** > **Volume Replication**

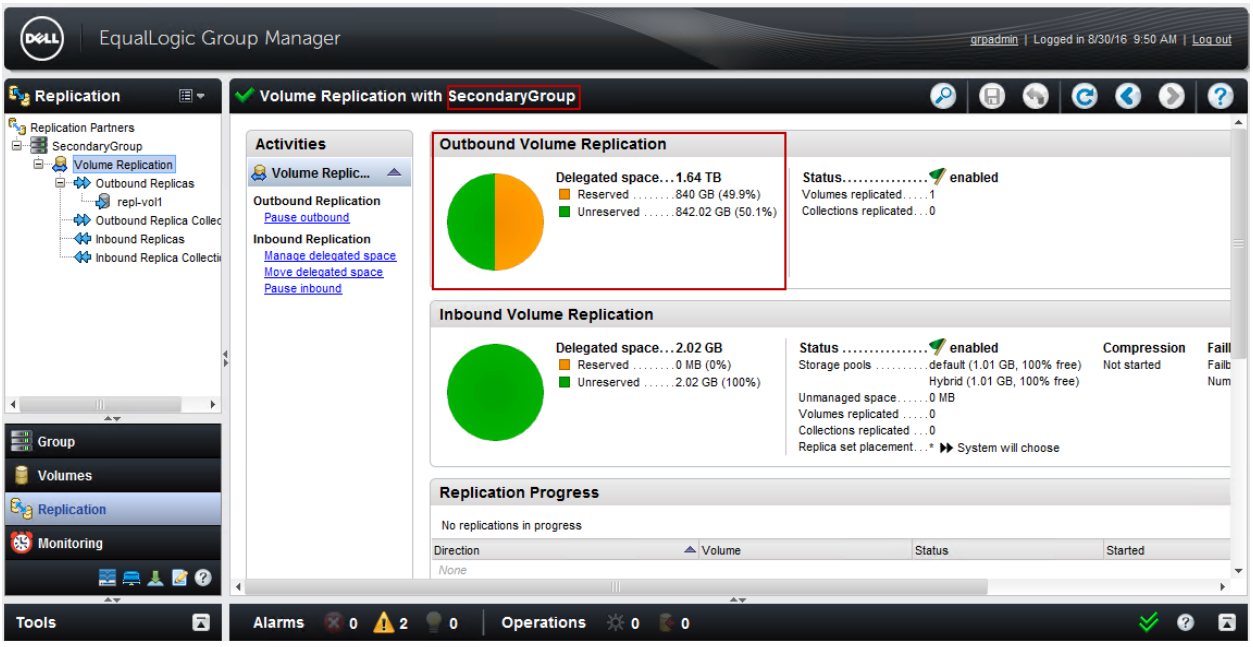

Figure 25 Total delegated space on the secondary (outbound from the primary perspective)

Summary of needed allocation for the secondary group:

- East pool delegated space: 841.01 GB
- West pool delegated space: 841.01 GB
- Total: 1682.02 GB or 1.64 TB

Navigation tip: **Group Manager GUI** (secondary) > **Group** > **Storage Pools** (select pool) > **Status** (tab)

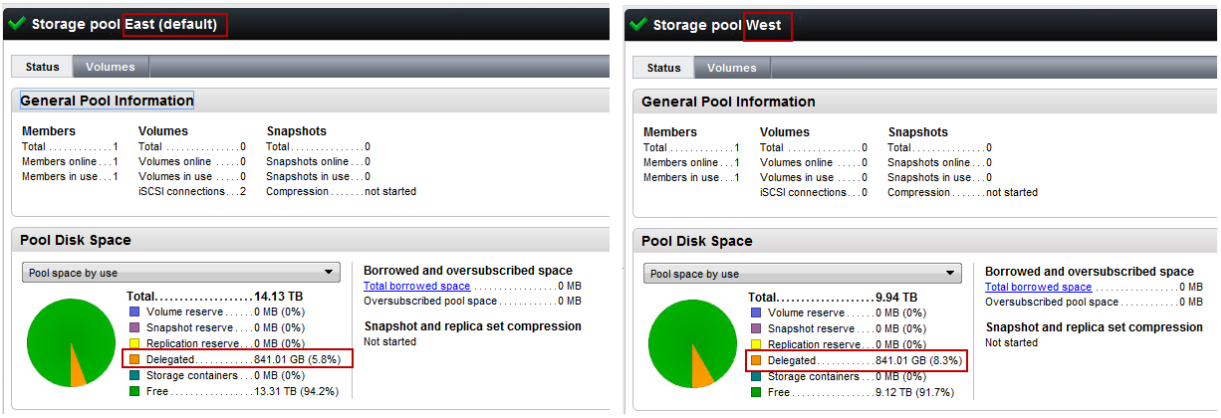

Figure 26 Side-by-side view of space allocation by East (default) and West pools

7.5.3 Initial replication and subsequent replication cycles

As a simple demonstration, first the repl-vol1 volume with 399.90 GB in-use space will be replicated to the *SecondaryGroup*.

Navigation tip: **SAN Headquarters** (primary) > **Capacity** > **Volumes** (select volume)

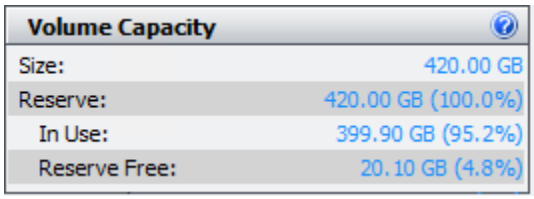

Figure 27 SAN Headquarters capacity pane showing 399.90 GB in use

The system will decide which pool to use for total replication reserve. The default policy is to let the system decide which pool the volume will replicate to on the secondary. For this case, the East pool was used for delegated space for the repl-vol1 volume.

Once the initial replication is complete, the delegated space on the SecondaryGroup will indicate near 50 percent reserved.

Navigation tip: **Group Manager GUI** (primary) > **Replication** > **Volume Replication**

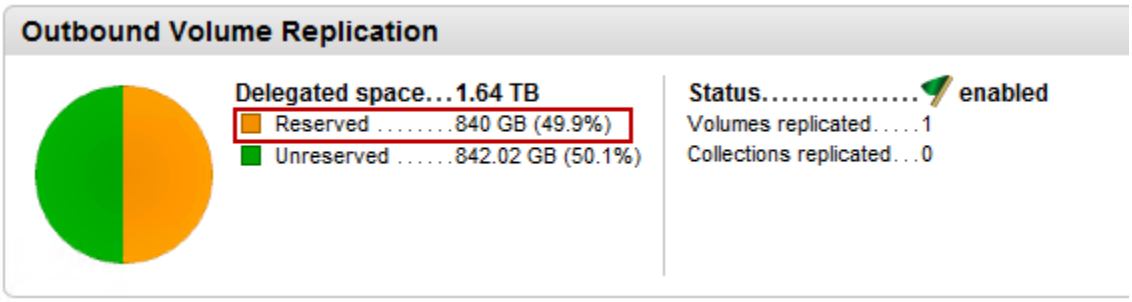

Figure 28 PrimaryGroup showing the used delegated space on the SecondaryGroup (outbound replication).

To demonstrate subsequent replication cycles, after the first replication, an additional 10 GB was added to repl-vol1. Those changes will be replicated on the next replication schedule or with manual replication.

Navigation tip: **Group Manager GUI** (primary) > **Replication** > **Outbound Replicas** > select volume

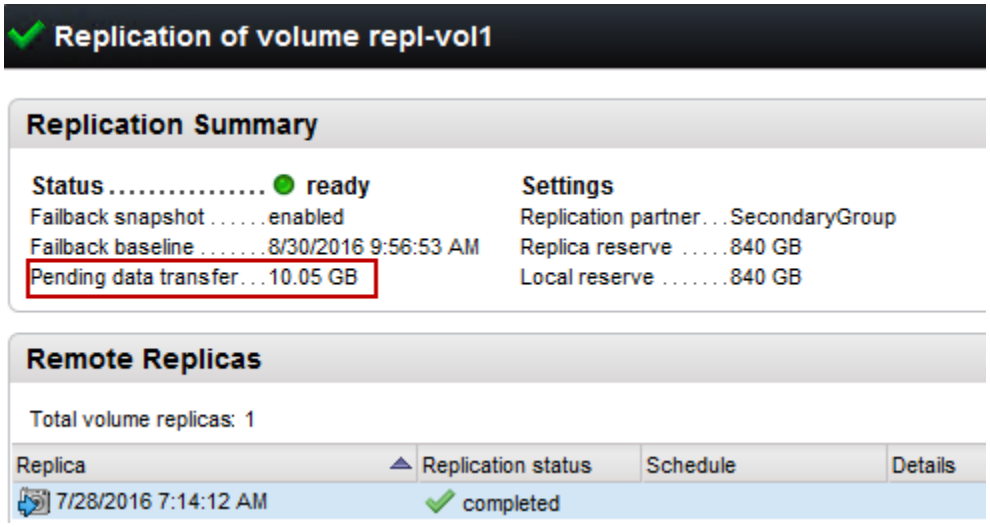

Figure 29 Group Manager GUI > Replication tab showing pending transfer of 10.05 GB

SAN Headquarters provides a good way to review the replication cycles and statistics.

Navigation tip: **SAN Headquarters** (primary) > **Outbound Replicas** > **Volumes** (select volume)

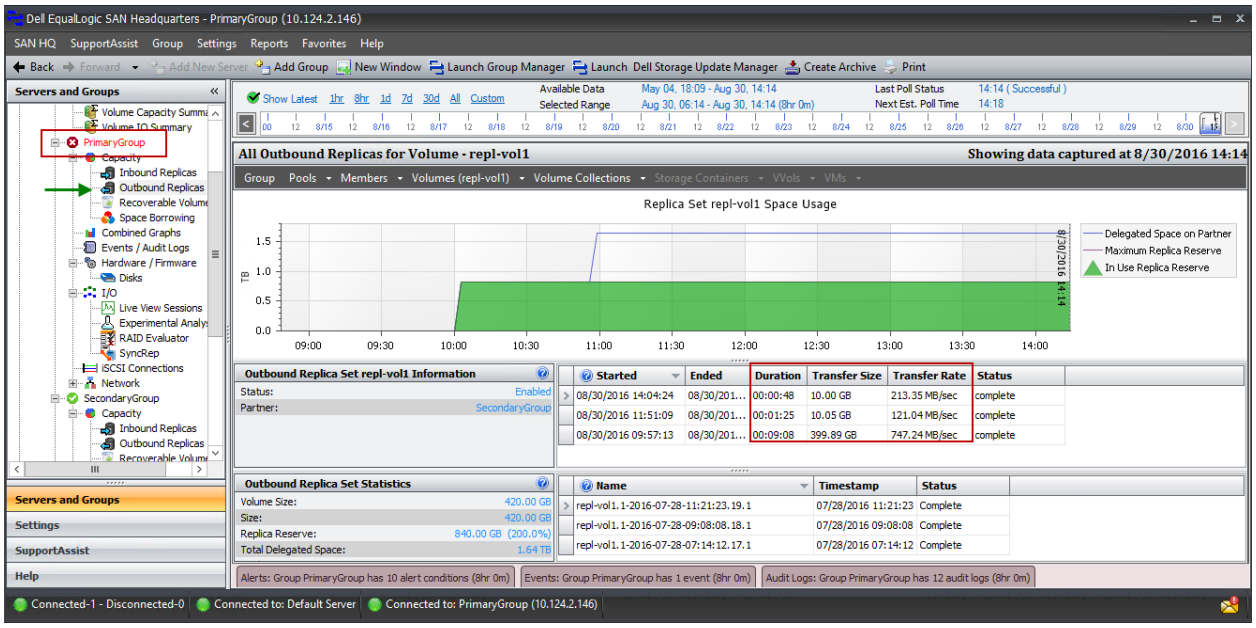

Figure 30 SAN Headquarters showing the duration, amount of data to transfer (transfer size), and the transfer rate

To continue, repl-vol2 will be configured for replication, and will perform the initial replication to the SecondaryGroup. The SecondaryGroup will determine the best location for the delegated space, which in this case is the West pool.

Navigation tip: **Group Manager GUI** (secondary) > **Replication** > **Inbound Replicas** (select volume)

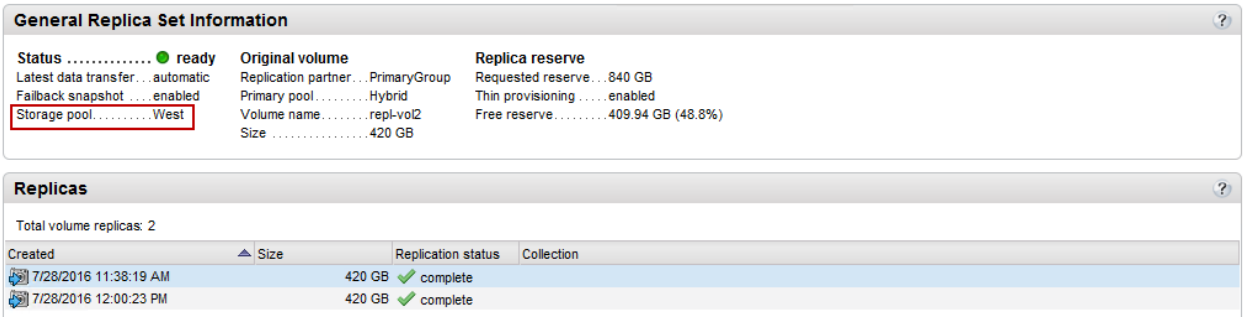

Figure 31 repl-vol2 is automatically placed in the West pool on the SecondaryGroup

Once complete, the total delegated space will be allocated. Monitoring the reserve and delegated space will be important for capacity planning. As an additional safeguard, SAN Headquarters may be configured to email alerts based on replication issues.

Navigation tip: **SAN Headquarters** > **Settings** (menu) > **E-mail Settings** > **Notifications** (tab) > **Warning** (tab)

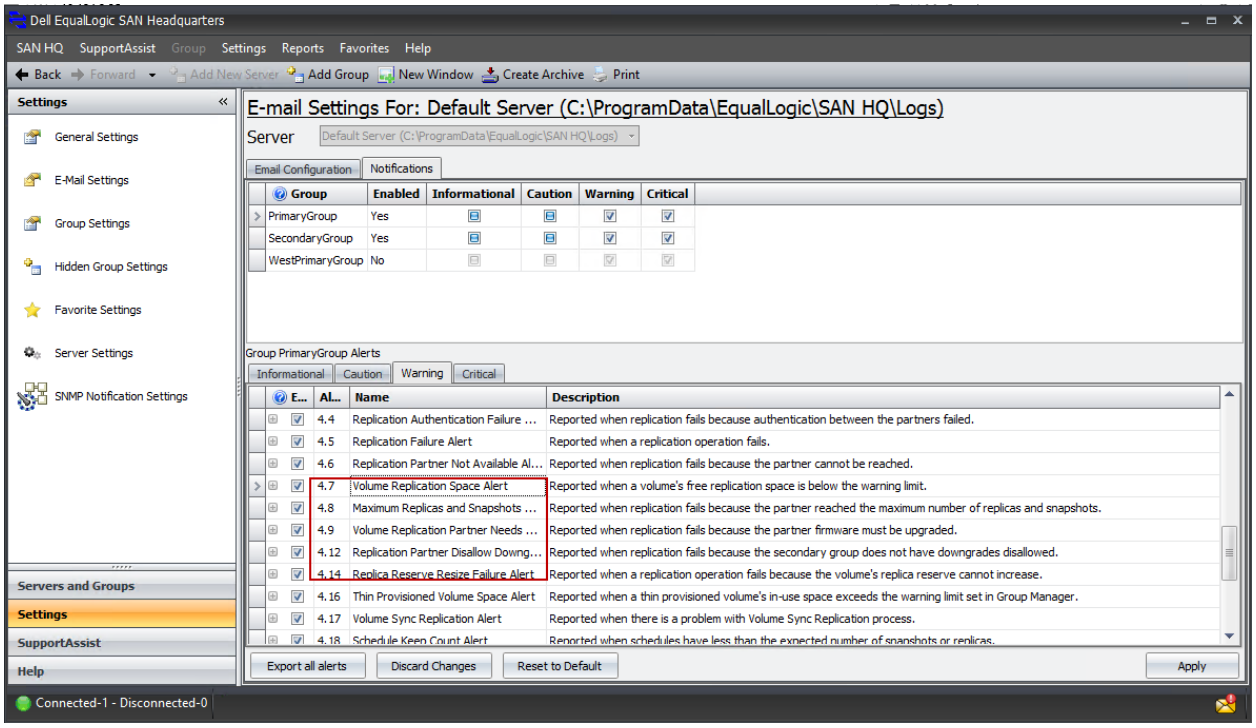

Figure 32 Email settings for replication alerts from SAN Headquarters

Note: To receive or change notification in SAN Headquarters, the Email Configuration tab must be filled out and tested.

Delegated space may also be monitored directly within SAN Headquarters as well, in addition borrowed space will be indicated.

Navigation tip: **SAN Headquarters** (primary) > **Capacity** > **Outbound Replicas**

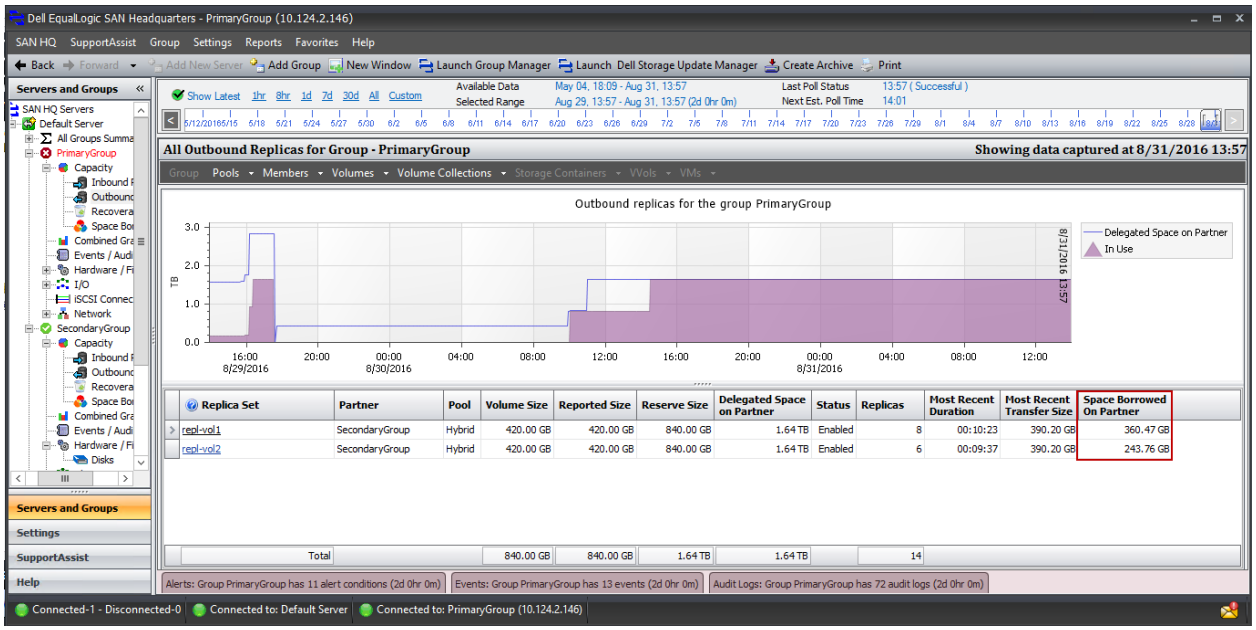

Figure 33 SAN Headquarters showing outbound replicas from the PrimaryGroup and the space borrowed on the partner (SecondaryGroup)

From the point of view of the secondary, the inbound replica total replica reserve usage may also help determine when to allocate more delegate space or acquire a new PS Series array. This is the case in the following example in which 100 percent of the reserve is in use.

Navigation tip: **SAN Headquarters** (secondary) > **Capacity** > **Inbound Replicas**

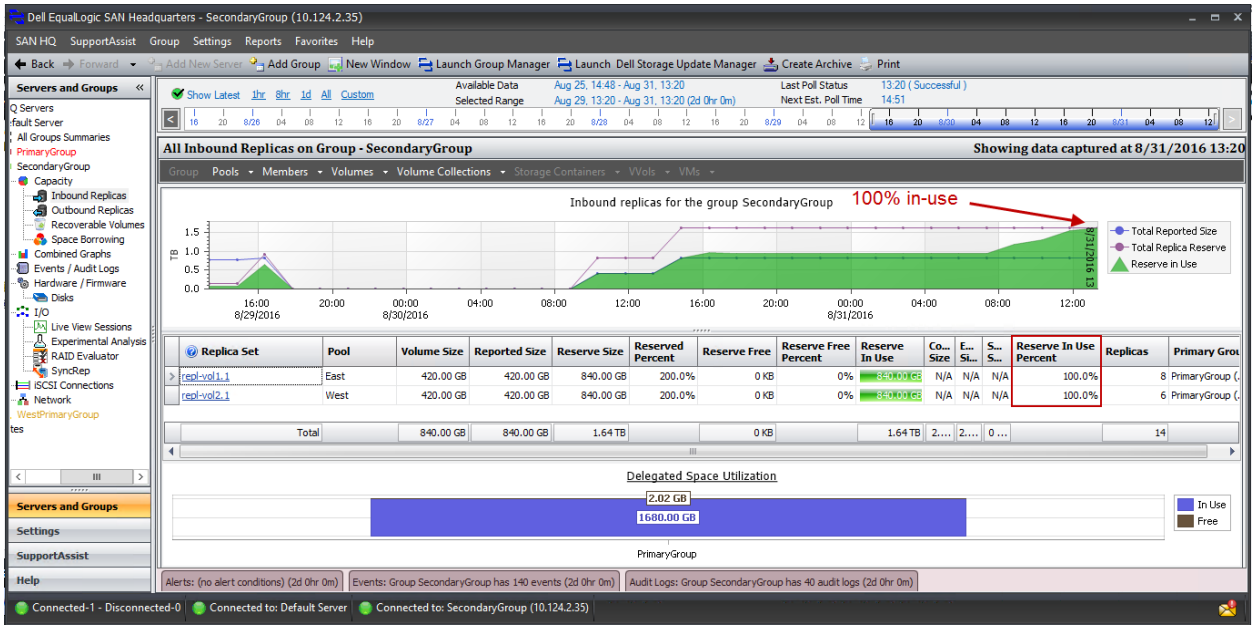

Figure 34 Total replica reserve on the secondary showing 100% in use.

For practical purposes, PS Series replication will borrow space as needed and efficiently use the space available. An example of when space borrowing is needed may be when additional space is required for the total replica reserve during a replication cycle. This may be observed in the Group Manager GUI.

Navigation tip: **Group Manager GUI** (primary) > **Group** > **Borrowed Space** > **Remote Delegated Space** (tab)

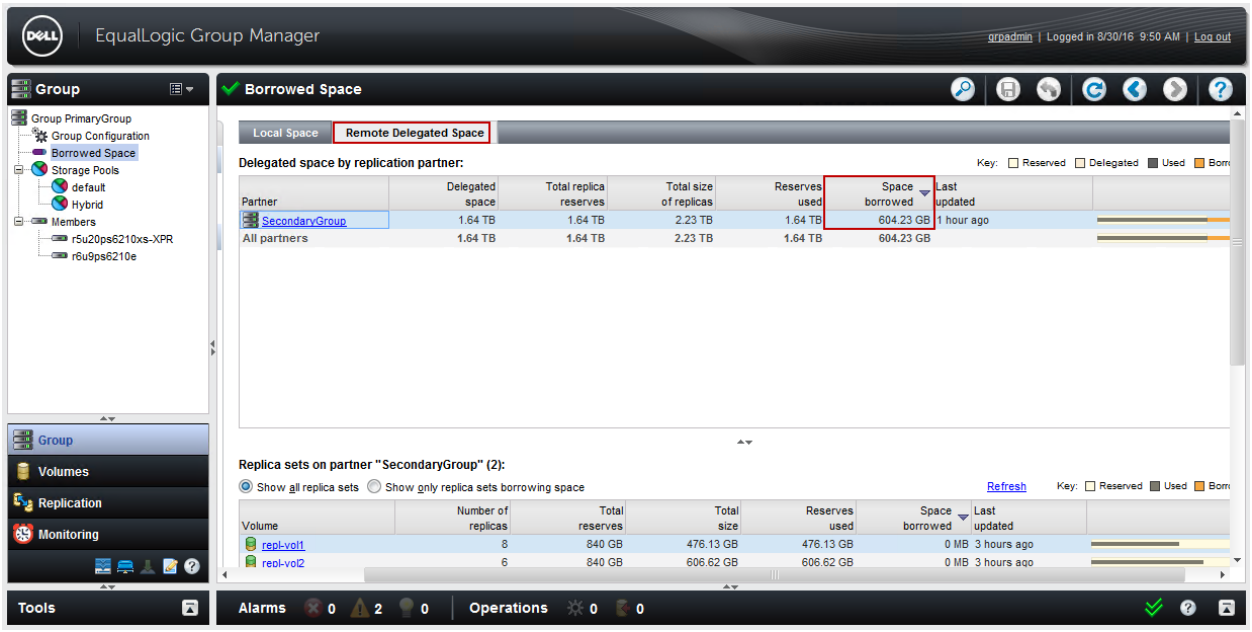

Figure 35 Amount of remote delegated borrowed space.

From a best practice perspective, the borrowed space should be temporary in nature. If replications are requiring borrowed space consistently over time, then the delegated space could be increased and that borrowed space will be freed up after the next replication cycle. See section 3.5 for more details.

Figure 36 shows the space borrowed returning to 0 MB after adding an additional 302 GB to both pools' delegated space.

Navigation tip: **Group Manager GUI** (primary) > **Group** > **Borrowed Space** > **Remote Delegated Space** (tab)

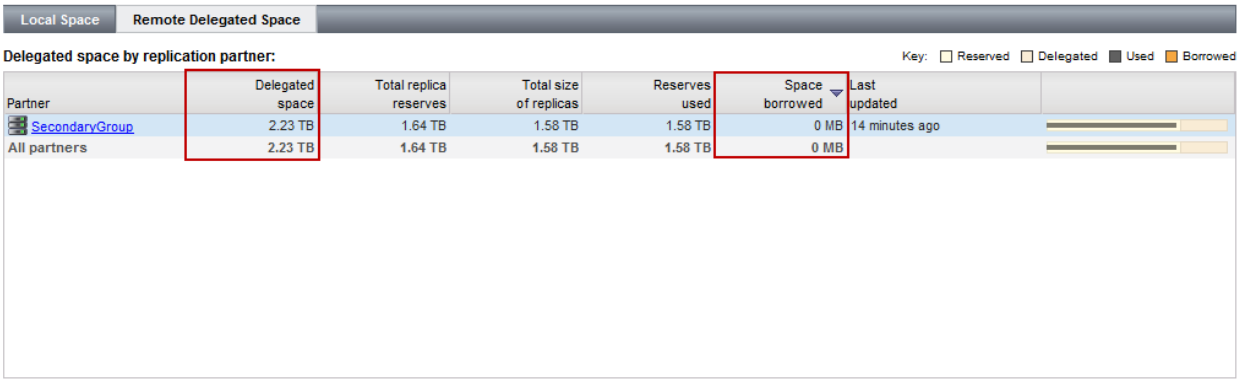

Figure 36 Space borrowed returns to 0 MB after increasing the delegated space

7.6 Monitoring Replication with SAN Headquarters

Figure 37 shows how SAN Headquarters allows you to monitor replication events over a relatively long period of time. SAN Headquarters will display the transfer rate and duration as well as the status of each replication cycle. System administrators should review these statistics to ensure that storage capacity is being used efficiently, that there is sufficient delegated space on the replication partner and that all replicas are successfully completing in the required time to meet RPO objectives.

Navigation tip: **SAN Headquarters** (primary) > **Capacity** > **Outbound Replicas** > **Volumes** (select volume)

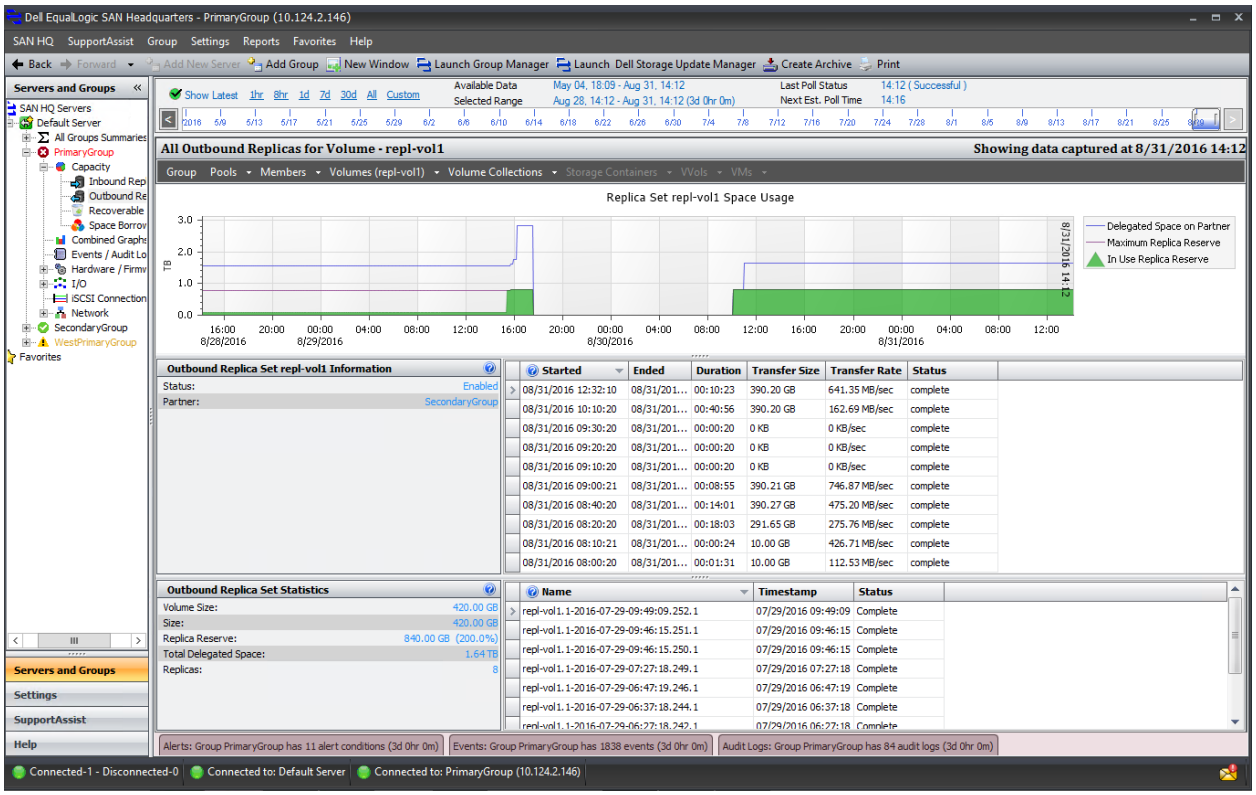

Figure 37 Monitoring Replication with SAN Headquarters

7.7 Replicating large amounts of data

Some of the testing used 100 GB volumes for replication. Some tests replicated multiple 100 GB volumes simultaneously, and others replicated only a single volume to show how this can affect overall replication times.

A 100 GB volume is not a huge amount of data in today's IT environments. By comparison, desktop and portable systems typically now have 500 GB or larger hard disks and file servers may contain multiple terabytes of user data. Replicating this amount of data can be extremely challenging over a slow link, and in some cases, may not work at all. For example, if a 10 TB file share volume experiences a 1 percent daily change rate, it would need to replicate approximately 100 GB per day. If the network is not fast enough to support replication of 100 GB in under 24 hours, than it is unlikely that it will meet the RPO for most businesses requirements.

All volumes will initially need to be fully synchronized. Depending on the initial space used in the volume and the speed of the network between sites, the required time to accomplish this task can vary greatly. The Manual Transfer Utility can be used to copy a volume to a portable disk drive and manually transfer the volume to be replicated to the remote site. However, if it is determined that the initial synchronization of replicas will take too long, this may just be the first sign that the network between replication partners is under-sized and may not be able to support a regular replication schedule.

Another option, which in some cases may be quicker and easier, is to connect the secondary storage systems in the local (primary) data center, establish the replication partnership, and do the initial full synchronization over the full-speed, local iSCSI SAN network. After the initial replication is complete, shut down the secondary storage and move it to the secondary or remote site. When the link is established, bring up the secondary storage and only the delta changes (the differences since the previous replication) will be replicated.

As mentioned previously, the use of WAN accelerators or optimizers is yet another method that may help to improve the efficiency of replicating large amounts of data over a WAN. These devices sit on either side of the WAN link and perform functions such as data deduplication, caching, and other packet optimization.

7.8 SAN-based replication or host-based replication

The asynchronous replication feature ships with every PS Series array to help meet the needs of most replication requirements. With host-based replication products, the host CPU will be responsible for reading data off the SAN, and then moving it from the host to the destination replica site. There may be situations requiring a host-based replication product instead of PS Series asynchronous replication. Here are some examples:

- The specific RPO or RTO requirements of an application cannot be satisfied by the capabilities of PS Series asynchronous replication alone.
- There is no network path available between the array controllers in the primary PS Series SAN group and the secondary group, and using the Manual Transfer Utility will not meet RPO/RTO requirements.

A Technical support and resources

Dell.com/support is focused on meeting customer needs with proven services and support.

Dell TechCenter is an online technical community where IT professionals have access to numerous resources for Dell software, hardware and services.

Storage Solutions Technical Documents on Dell TechCenter provide expertise that helps to ensure customer success on Dell EMC Storage platforms.

A.1 Additional resources

The following publications are referenced in this document or are recommended sources for additional information.

- *PS Series Configuration Guide*
- *Using Dell PS Series Asynchronous Replication*
- *Space Borrowing for Snapshots and Replicas*
- *Dell EqualLogic PS Series Arrays: Understanding Synchronous Replication*
- *Best Practices for Deploying Dell EqualLogic Synchronous Replication*
- *Protect Your Data And Your Business With SAN-Based Synchronous Replication*
- PS Series Technical Documents