



Veritas™ Volume Replicator Option by Symantec

A Guide to Understanding Volume Replicator

A technical overview of replication capabilities included in Veritas Storage Foundation™ and the Volume Replicator Option.

Veritas™ Volume Replicator Option

by Symantec

A Guide to Understanding Volume Replicator

Contents

Introduction: Replication and disaster recovery planning	6
Understanding the need for replication	8
Replication modes	8
Technical considerations for replication	10
Remote mirroring and replication	10
Remote mirroring with Storage Foundation	11
Replication with the Veritas Volume Replicator Option by Symantec	11
Volume Replicator Option technical details	14
Operational modes and data flow: Synchronous	14
Operational modes and data flow: Asynchronous	15
Initializing secondary systems for replication	16
Recovery after problems	18
Secondary/network outage	18
Secondary failure	18
Primary failure	19
Volume Replicator role changes	19

Contents *(cont'd)*

Using the secondary system20
Using In Band Control messages to control FlashSnap™21
Volume Replicator in the customer environment21
Effects of Volume Replicator on host resources and application performance21
Understanding bandwidth needs using VRAdvisor21
Volume Replicator integration with other products23
Veritas™ Cluster Server by Symantec23
Veritas Storage Foundation for Databases24
Veritas NetBackup™24
Replication capabilities summary24
Storage architecture independence24
Maintaining write order fidelity25
Native replication over Fibre Channel and IP networks25
Scalable25
Initialization options25
Summary26

Introduction: Replication and disaster recovery planning

There are many threats that organizations face today when it comes to the reliability and viability of their data. Logical corruption, or loss of data due to threats such as viruses and software bugs, can be protected by ensuring that there is a viable copy of data available at all times. Performing regularly scheduled backups of the organizations' data typically protects against logical types of data loss. Another threat that may result in data loss is component or hardware failure. While most devices have begun to build in redundancy, there are other technologies such as application clustering technologies that can protect against a failure of a component while continuing to enable applications to be available.

Just as the levels of protection for logical and component failures have grown, so has the reliance on the information systems being protected. Many companies now realize that logical and local protection is no longer enough to guarantee that the organization will continue to be accessible. This loss can stem from planned downtime such as complete site maintenance to unplanned downtime such as power or cooling loss to natural disasters such as fire and flooding to acts of terrorism or war. The loss of a complete data center facility would so greatly affect an organization's capability to function that protection must be established at the data center level.

Many companies have implemented significant disaster recovery (DR) plans to protect against the complete loss of a facility. Plans likely include steps for recovering communication lines, staffing critical functions, recovering data and restoring business applications at a remote location. One issue that is common among many DR plans is the information processing recovery plan. For many years companies have been taking regular data backups at the primary data center, then duplicating these tapes on a regular basis for shipment offsite to the DR facility. While a tape-based backup solution is the needed safety net for disaster recovery planning, there is still a need in the IT environment to provide higher levels of protection for critical data. While a tape backup approach may meet the needs of much of the data within an organization, there are many data types that cannot afford the levels of data loss inherent to a tape backup approach.

The first step to understanding which technologies are necessary for particular data types is to understand when and if appropriate technologies are needed. The key measure of disaster recovery technologies is based on recovery point objectives and recovery time objectives.

Veritas Volume Replicator Option by Symantec

A Guide to Understanding Volume Replicator

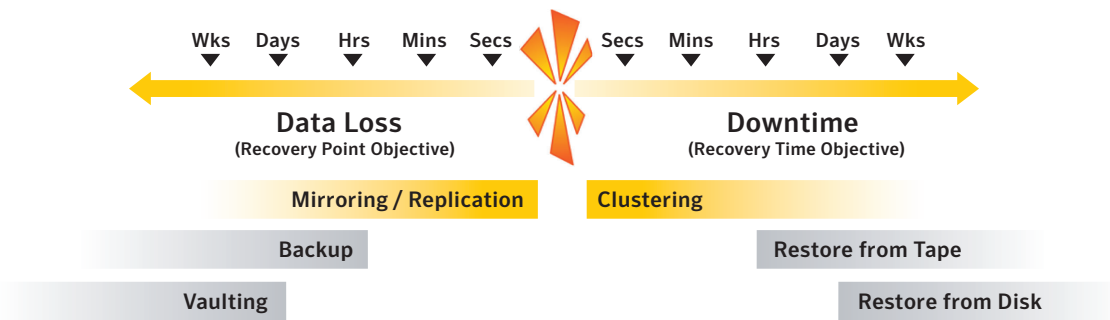


Figure 1. Identifying application recovery point objectives (permissible data loss) and recovery time objectives (permissible downtime) should be the first step in choosing disaster recovery technology.

A complete disaster recovery plan is not delivered by any one technology or service, but rather by an accumulation of steps that are implemented in order to provide the needed recovery point objective (RPO) and recovery time objective (RTO) for each application and business function. A recovery point objective is the point in time to which an application's data must be recovered to resume business transactions. A recovery time objective is the maximum elapsed time before lack of business function severely impacts an organization. When analyzing a disaster recovery solution, many components must be implemented in order to guarantee application availability.

Figure 1 outlines software technologies that map to a customer's RPO and RTO requirements. The burst in the middle represents a disaster. To the left of the burst is the recovery point objective with the appropriate software technology based on business needs. For example, if a particular application can afford a day or more of data loss, then a tape backup approach is all that that is needed for that application. However, if a day or more of data loss will cause substantial business impact, then replication technologies must be implemented into the IT environment to protect against substantial data loss. To the right of burst is the recovery time objective. If a business can afford to take a day or more to resume normal business activity, then manual tape restore will satisfy their business needs. Organizations can improve on this RTO by using bare metal restore technologies that dramatically reduce the amount of time it takes to get a server up and running. However, if those technologies do not meet the application's RTO, then clustering technologies must be implemented in the IT environment to protect against substantial downtime.

Understanding the need for replication

Replication is a technology designed to maintain a duplicate data set on a completely independent storage system at a different geographical location. Replication differs from tape backup/restore methods because replication is completely automatic and far less labor-intensive. In addition, replication technologies can be used to reduce the recovery point objective of critical applications.

Whether motivated by disaster, site failure, or a planned site migration, Symantec's replication technologies make it possible to distribute data for seamless data availability across sites. Veritas Storage Foundation by Symantec provides remote mirroring capabilities natively over Fibre Channel protocols. For organizations that wish to replicate their data natively over a standard IP network, an optional capability of Veritas Storage Foundation, called Veritas Volume Replicator, can reliably, efficiently, and consistently replicate data to remote locations. Symantec's replication technologies provide a robust, storage-independent disaster recovery solution when data loss and prolonged downtime cannot be tolerated.

Replication modes

The two main types of replication are synchronous and asynchronous. Both have their advantages and disadvantages and should be available options for the IT administrator. Each uses a different process to arrive at the same goal, and each deals somewhat differently with network conditions. The performance and effectiveness of both depend ultimately on business requirements such as how soon updates must be reflected at the target location. Performance is strongly determined by the available bandwidth, network latency, the number of participating servers, the amount of data to be replicated, and the geographical distance between the hosts.

Synchronous replication

Synchronous replication ensures that a write update has been posted to the secondary location(s) and the primary location before the write operation is acknowledged to be complete at the application level. This way, in the event of a disaster at the primary location, the data recovered at the secondary location will be an exact copy of the data at the primary location. Synchronous replication produces the exact same data at both the primary and secondary location, which means the RPO of applications using synchronous replication would be zero. However, since the application transaction must travel to the secondary location(s) and back to the primary location before the application can continue with the next transaction, there will be some application performance impact. Synchronous replication is most effective in metropolitan area networks with application environments that require zero data loss and can afford some application performance impact. For all other applications, asynchronous replication should be a viable alternative.

Veritas Volume Replicator Option by Symantec A Guide to Understanding Volume Replicator

There are many scenarios that could affect the performance of replication in synchronous mode, including the amount of write activity on the system, the network pipe connecting the primary and secondary sites, and the distance between the two sites. A good rule of thumb to use is 3 ms of latency for every 100 miles of distance between the primary and secondary systems. Most configurations that use synchronous replication are set to change to asynchronous mode if the network link is lost between the primary and secondary site. This prevents the primary application from being affected by a network outage.

Asynchronous replication

Asynchronous replication eliminates the potential performance problems of synchronous methods. The secondary site may lag behind the primary site, typically only by less than one minute, offering essentially real-time replication without the application performance impact. During asynchronous replication, application updates are written at the primary and queued for forwarding to each secondary location as network bandwidth allows. Unlike synchronous replication, the writing application does not suffer from the application performance impact of replication and can function as if replication is not occurring. Asynchronous replication should be used in organizations that can afford minimal data loss but want to eliminate application performance impact or by organizations that would like to replicate data over a wide area network. This can also be the right choice if the network bandwidth between the two sites is large enough to handle the average amount of data, but insufficient to handle the peak write activity.

Data consistency

Whatever replication mode you select, you should ensure that the data at the secondary site is never corrupted or inconsistent. The last thing you need is a nonrecoverable replicated data set at the secondary location the very moment you need it most. The only way to ensure that the data is recoverable at the secondary location(s) is to ensure that the data arrives in the same order as it was written at the primary location. This is called write order fidelity. Without write order fidelity, no guarantee exists that a secondary will have consistent recoverable data. In a database environment, updates are made to both the log and data spaces of a database management system in a fixed sequence. The log and data space are usually in different volumes, and the data itself can be spread over several additional volumes. A well-designed replication solution needs to consistently safeguard write order fidelity. This may be accomplished by a logical grouping of data volumes so the order of updates in that group is preserved within and among all secondary copies of these volumes.

Veritas Volume Replicator Option by Symantec A Guide to Understanding Volume Replicator

Replication solutions running at the hardware level typically lack the ability to maintain write order fidelity when running in asynchronous mode. This is due to a lack of a persistent queue of writes that have not yet been sent to the secondary. If you are using hardware replication and wish to avoid application performance impact imposed by synchronous replication, you lose recoverability on the remote site, rendering the remote copy essentially useless. Therefore, maintaining write order fidelity when using replication technologies should be an absolute requirement to ensure the recoverability of data at a remote location.

Technical considerations for replication

Architecturally, a complete replication solution must provide a copy of all data at the primary and secondary locations, including database files as well as any other necessary binary and control files. The replication technology also must ensure that the data is accurate and recoverable.

The replication solution must be capable of being configured to support a secondary site over any distance. In today's environment, organizations must be allowed to use current infrastructure, including current data centers, regardless of distance. The replication solution must operate at any distance, whether the data centers are a few kilometers apart or thousands of kilometers apart, without adding undue cost or complexity. This means the replication technology must provide asynchronous support, over a long distance, without additional high cost items such as communication converters or additional disk space for staging data. The replication solution must be flexible enough to allow you to change the configuration of the data sets that are being replicated. This could include having volumes that are not replicated, because they are for temp files, or files that wouldn't be needed in a disaster. There should also be the ability to grow and shrink the volumes with no application downtime. Additionally, the solution should allow testing at the remote site in order to validate that the data at the secondary is recoverable.

Remote mirroring and replication

Symantec replication and remote mirroring technologies can dramatically speed recovery time and eliminate data loss by making current data available immediately at an alternate location. Organizations can replicate or mirror data via a storage area network or over any IP network in order to meet their disaster recovery needs. Unlike proprietary, inflexible hardware approaches, Symantec's replication and remote mirroring technologies are not dependent on any specific storage hardware platform. For example, replication can occur between storage arrays from the same vendor, regardless of array model or size, or replication can occur between different storage vendors' arrays. The only requirement is that the volume sizes match at each side. Symantec's

Veritas Volume Replicator Option by Symantec A Guide to Understanding Volume Replicator

software-based replication provides a reliable, efficient, and cost-effective solution for geographically mirroring data sets. It also has full database management system support, including IBM® DB2®, Microsoft® Exchange, Oracle®, SQL Server, and Sybase.

Remote mirroring with Storage Foundation

Veritas Storage Foundation is the industry leader in storage virtualization. It provides an easy-to-use, online storage management tool for heterogeneous enterprise environments. Organizations can extend their storage management functionality with the Veritas Storage Foundation remote mirroring capability in order to deliver a metropolitan area disaster recovery solution. These remote mirroring capabilities are available with the basic Veritas Volume Manager features included with Storage Foundation. Organizations can use Veritas™ Volume Manager to synchronously mirror data natively over storage protocols such as Fibre Channel, which makes it an ideal solution for disaster recovery within a metropolitan area network. Using Veritas Storage Foundation, customers wishing to implement a disaster recovery solution over Fibre Channel can create “just another mirror” of their data over an extended distance, to be made available should a complete site outage occur. This solution allows for using different storage arrays at the two sites, and is seamless to the storage administrators and users of the data.

Replication with the Veritas Volume Replicator Option by Symantec

For organizations who wish to replicate data natively over an IP network, Veritas Storage Foundation has an optional capability called Volume Replicator. The Volume Replicator Option reliably, efficiently, and consistently replicates data to remote locations over an IP network for maximum business continuity, removing the need for expensive proprietary network hardware as well as the need to have the exact same storage hardware at every site.

Since Volume Replicator is an extension of the Volume Manager included in Storage Foundation, Volume Replicator allows logical volumes on one system to be exactly replicated to identically sized volumes on another system. For example, an Oracle database may use several different volumes for various tablespaces, redo logs, archived redo logs, indexes, and other storage. Each component is typically stored in a logical volume or multiple volumes. Volume Replicator can provide an exact duplicate of these volumes to another system, at another site, and since the replication occurs at the volume level, it can replicate data between any storage hardware arrays. Volume Replicator can scale to support up to 32 secondary data storage sites and elegantly handles one-to-one, one-to-many, and many-to-one data replication configurations.

There are four main components that are added to the volume manager code base of Veritas Storage Foundation to provide Volume Replicator. These four components are replicated volume groups (RVG), storage replicator log (SRL), RLINKS, and data change maps (DCMs).

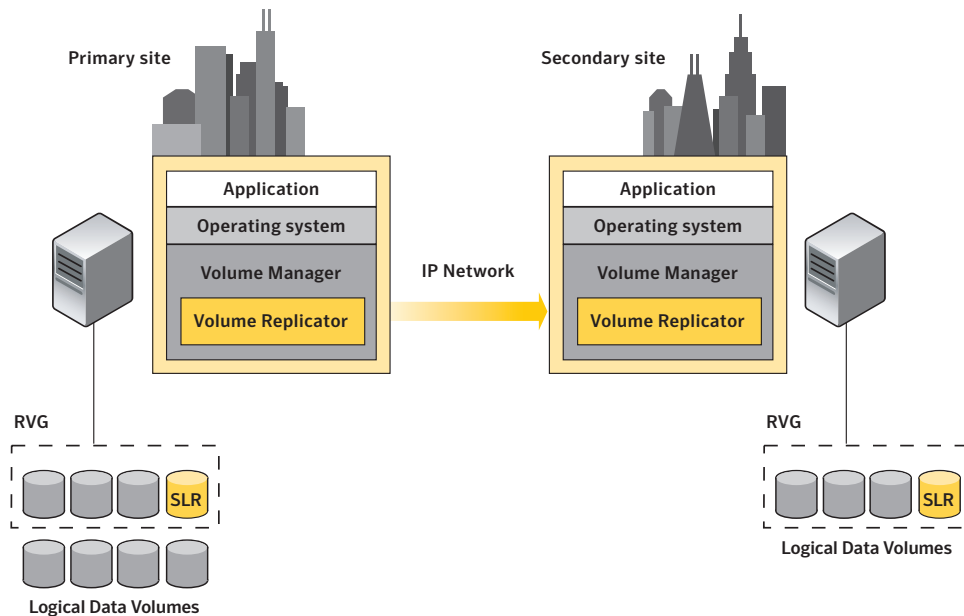


Figure 2. Volume Replicator extends the capabilities of Veritas Storage Foundation by enabling long distance replication over IP networks.

Replicated Volume Groups

With Volume Replicator, the concept of a disk group (found in Veritas Volume Manager) is extended to provide the concept of a Replicated Volume Group (RVG). As shown in figure 2 above, an RVG is a subset of volumes within a given Veritas Volume Manager disk group configured for replication to one or more secondary systems. An RVG can contain one or more data volumes in the disk group, up to and including all data volumes, but cannot span multiple disk groups. Multiple RVGs can be configured inside one disk group, but not the other way around. Volumes that are associated with an RVG and that contain application data are called replicated data volumes.

The concept of an RVG enables the user to pick and choose what data is replicated to the secondary site. Therefore, organizations need only to replicate their mission-critical data to a secondary location, which allows them to save money on replication implementations because they require a lot less storage at their secondary site. In addition, organizations that pay for bandwidth usage can save on bandwidth costs because they are only replicating data that has stringent recovery point objectives. All other data can be protected by simply using a tape backup approach.

Veritas Volume Replicator Option by Symantec A Guide to Understanding Volume Replicator

The data volumes in the RVG are under the control of an application, such as a database management system, that requires write-order fidelity among the updates to the volumes. Write ordering is strictly maintained within an RVG during replication to ensure that each remote volume is always consistent, both internally and with all other volumes of the group. At the simplest level, Volume Replicator exists within the Veritas Storage Foundation code base and has the capability to intercept any write destined for a logical volume within an RVG and replicate the write, in the correct order, to designated secondaries before the write is passed on to the actual data volumes.

Data is replicated from a primary RVG to a secondary RVG. The primary RVG is in use by an application, while the secondary RVG receives replication and writes to local disk. The concept of primary and secondary is per RVG, not per system. A system can simultaneously be a primary RVG for some RVGs and secondary RVG for others. The RVG also contains the Storage Replicator Log (SRL) and Replication Link (RLINK) explained in the following sections.

Storage Replicator Log

All data writes destined for volumes configured for replication are first persistently queued in a log called the Storage Replicator Log. Volume Replicator implements the SRL at the primary side to store all changes for transmission to the secondary(s). The SRL is a logical volume configured as part of an RVG. The SRL enables Volume Replicator to associate writes to specific volumes within the replicated configuration in a specific order, maintaining write order fidelity at the secondary. All writes sent to the volume layer, whether from an application such as a database writing directly to storage or an application accessing storage via a file system, are faithfully replicated in application write order to the secondary.

When implementing asynchronous replication, data integrity must be guaranteed at the remote site, or it must preserve write order fidelity. This essentially means that writes applied to the secondary storage must occur in the exact order they were applied at the primary. Asynchronous replication without this capability compromises data consistency at the disaster recovery site and may jeopardize the recoverability of the data. The SRL within Volume Replicator tracks writes in the correct order and guarantees that the data will arrive at the secondary site in that same order, whether operating in synchronous or asynchronous mode.

The SRL can also be used with synchronous or asynchronous replication to protect against network outages. Should a network outage occur when replicating in synchronous mode, Volume Replicator can automatically switch to asynchronous mode, and the writes can be stored in the SRL for later transmission to secondaries when the network is restored. This functionality protects the primary location from being impacted in the event of a network outage.

Veritas Volume Replicator Option by Symantec

A Guide to Understanding Volume Replicator

RLINKs

An RLINK is a replication link to a secondary RVG. Each RLINK on a primary RVG represents the communication link from the primary RVG to a corresponding secondary RVG, via an IP connection. RLINKs are configured to communicate between specific host names/IP addresses and can support both TCP and UDP communication protocols between systems.

Data Change Maps

Data Change Maps (DCMs) are used to mark sections of volumes on the primary server that have changed during extended network outages in order to minimize the amount of data that must be synchronized to the secondary site during the outage.

Volume Replicator Option technical details

Operational modes and data flow: Synchronous

In synchronous mode, all data writes are first posted to the SRL and then sent to the secondary location(s). The posting of the data to the actual data volumes happens in this time period also. When the secondary location(s) receives the data write and acknowledgement is sent back to the primary location, then the application acknowledges the data write to be complete. Therefore, synchronous replication should be used in environments that cannot afford any data loss and can afford some application performance impact. Overall performance in synchronous mode is governed by the amount of time that it takes to write to the SRL plus the round-trip time to send data to the secondary and receive acknowledgement.

The secondary acknowledges receipt as soon as the full write transaction (as sent by the application on the primary) is received into Volume Replicator kernel memory space. This removes actual writing to the secondary data volumes from the application latency. The primary tracks these writes in its SRL until a second acknowledgement is received from the secondary, signaling that the data has been written to physical storage. Both acknowledgements have an associated timeout, so if a packet is not acknowledged, it will be re-sent.

To maximize performance, Volume Replicator only waits for data to be received. Not waiting for data to be written at the secondary improves application performance. But it tracks all acknowledged, uncommitted transactions and can replay any necessary transactions if the secondary were to crash prior to actually writing its data to the physical storage. Synchronous mode has two possible RLINK settings: "fail" and "override." These settings deal with the behavior of Volume Replicator when the network connection is lost to a secondary site.

Veritas Volume Replicator Option by Symantec A Guide to Understanding Volume Replicator

With synchronous=fail, writes will be returned as failed to the calling application if contact is lost with the secondary. This is used in rare situations where the primary and secondary storage must never differ by even one write. This is typically not used, as it will cause an application failure at the primary side if anything happens to the secondary or the interconnecting network. The “synchronous=override” is the most common setting. This will keep replication in synchronous mode unless contact with the secondary is lost, then will shift to asynchronous mode and begin tracking the writes in the SRL. This allows the primary application to continue to run and provide service to customers while the backup capability is restored.

Operational modes and data flow: Asynchronous

In asynchronous mode, an application data write is first placed in the SRL, then immediately acknowledged to the calling application at the primary site. Data is then sent as soon as possible to the secondary location, based on available bandwidth. Therefore, asynchronous replication will not have any impact on the performance of the application, but the secondary location may be a few write requests behind the primary site should a disaster occur. Asynchronous replication should be used in environments where minimal data loss (typically measured in seconds for adequately sized networks) can be tolerated but where applications cannot afford the performance impact of synchronous replication.

Operational modes and data flow: Bunker Replication

Bunker Replication is a feature of Volume Replicator that enables bunker site configurations, wherein organizations replicate asynchronously over IP to a remote disaster recovery site, while also replicating the critical data synchronously (using Fibre Channel or IP) to a nearby bunker site. The key to this solution is that the bunker site only houses the data differential between the primary and the secondary site. This way, if an outage occurs at the primary site, the bunker site will immediately update the remote site with any data that would have been lost if only asynchronous replication had been used. An additional benefit of Bunker Replication is the possibility of increased bandwidth savings over traditional asynchronous solutions by permitting organizations to allocate bandwidth based on average bandwidth requirements and not peak bandwidth requirements.

Veritas Volume Replicator Option by Symantec A Guide to Understanding Volume Replicator

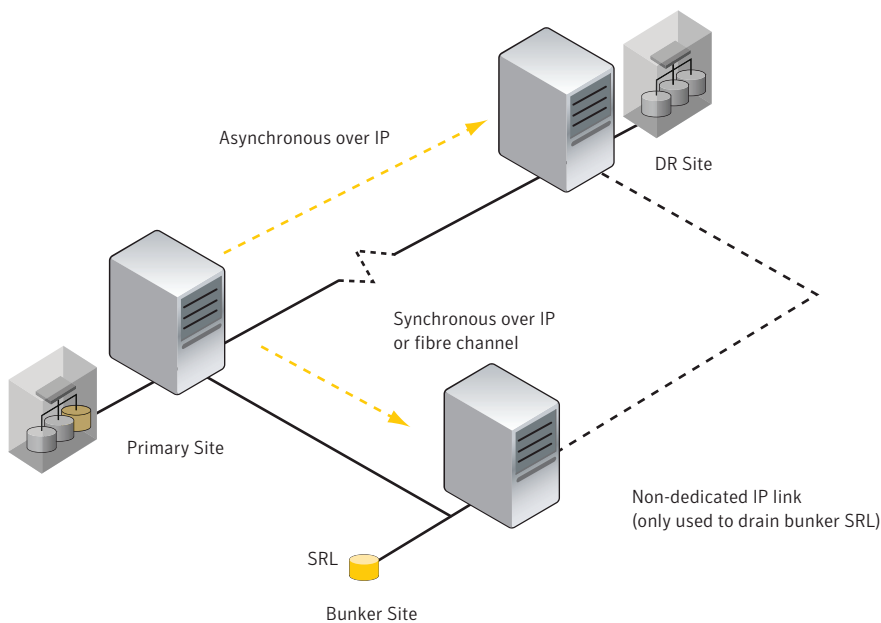


Figure 3. Bunker replication combines synchronous and asynchronous replication for zero data loss protection over any distance.

Using asynchronous replication to decouple application latency

One of the most compelling features of Volume Replicator in real-world environments is its ability to maintain full consistency at the secondary site while operating in asynchronous mode. Maintaining write order fidelity in asynchronous mode allows Volume Replicator to truly make use of the performance benefits available from asynchronous replication. By providing a high-bandwidth connection, customers can completely remove the latency penalty from replication and still maintain near-up-to-the-second data at the remote site. At the primary site, the application is acknowledged as soon as data is placed in the SRL. The application can continue to function as if replication is not occurring on the host. The data is then sent out almost instantaneously over the network to the secondary site. With adequate bandwidth, the SRL will not fill, so the actual data outstanding between primary and secondary is realistically whatever data is currently on the wire. This means a company can have near-up-to-the-second replication, at an arbitrary distance, with no application penalty.

Latency protection

Latency protection allows administrators to define how far a secondary is allowed to fall behind a primary in asynchronous mode. The latency protection feature allows automatic control of

Veritas Volume Replicator Option by Symantec

A Guide to Understanding Volume Replicator

excessive lag between primary and secondary nodes. Latency protection gives the network administrator the option to set the maximum number of updates that are allowed in the SRL, which is referred to as a `latency_high_mark`. When this number is reached, all update activity is delayed until the update backlog has reached a preset level, or the `latency_low_mark`. Latency protection ensures that the number of recent updates that could be lost in a disaster does not exceed a maximum determined amount. Latency protection is typically used to prevent the secondary from falling too far behind the primary in order to meet the recovery point objectives of the organization.

No distance limitations

Because Volume Replicator is truly unique in its ability to replicate data either synchronously or asynchronously over any standard IP network, there are no distance limitations. This allows organizations to utilize data centers that are already in place regardless of the distance between the locations.

Initializing secondary systems for replication

In order to begin replication of changed blocks, a replication solution must first begin with a known duplicate data set at each site. Volume Replicator offers several ways to get a secondary site up and running in order to begin the process of replication.

Empty

The simplest method to begin replication is to start with completely empty systems at each side. This can be done if Volume Replicator is installed while initially constructing a data center, prior to production. For Volume Replicator to use an empty data set, both sides must be identically empty.

Over the wire (Autosync)

Over-the-wire initialization is essentially using Volume Replicator to move all data from the primary to the secondary over the network connection. Overall this is a very simple process; however, with larger data sets it can take a prohibitively long time, especially if the primary is active while attempting to initialize the secondary. In addition, in situations where organizations are leasing bandwidth lines, this process can become fairly expensive. The items that must be taken into consideration are:

- The network bandwidth
- The amount of write activity on the system
- The size of the SRL

Veritas Volume Replicator Option by Symantec A Guide to Understanding Volume Replicator

This is the optimum way to perform the initial synchronization, because it can be repeated at any given time, if the solution is designed to allow for this process.

Local mirroring

Local mirroring is an option for very large data sets. In this method, the data storage array for the secondary site is initially placed at the primary site and Volume Manager is used to mirror the data between the two storage devices. Once the mirror is complete, the Volume Manager plex is split off and the array is shipped to the secondary site. This mode will allow large amounts of data to be initialized at SAN speeds, as well as allowing subsequent data written during the shipping period to be spooled to the primary SRL. However, this can be fairly expensive if the array must be shipped over a long distance.

Tape backup and restore initialization

The final option for initialization is through the use of a tape backup and restore. This is a unique feature of Volume Replicator and is not an available option for other replication technologies on the market today. It allows huge data sets to be synchronized using tape technology and immediately begin replication.

Checkpoint initialization is a hot backup of the primary side, with the SRL providing the map of what was changed during the backup. When a checkpoint initialization is started, a “check-start” pointer is placed in the SRL. A full block-level backup is then taken of the primary volumes. When complete, a check-end pointer is placed into the SRL. The data written between the check-start and check-end pointers represents data that was modified while the backup was taking place. This constitutes a hot backup.

The tapes are then transported to the secondary site, and the data is restored to the secondary systems using the tapes. When the tape load is complete, the secondary site is connected to the primary site with a “checkpoint attach” of the RLink. The primary will then forward any data that had been written during the backup (that data between the check-start and check-end). Once this data is written to the secondary, the secondary is an exact duplicate of the primary, at the time the backup completed. At this point the secondary is consistent, and simply out-of-date. The SRL is then replayed to bring the secondary up-to-date. Therefore, tape backup and restore initialization is ideal for organizations that wish to perform an initialization of a large data set or for environments where their secondary site is located over longer distances.

Recovery after problems

Volume Replicator is very robust in terms of tolerating outages of the network as well as of secondary systems. The SRL provides the key to recovering from outages, while maintaining a consistent secondary.

Secondary/network outage

Volume Replicator can be configured to handle network outages. An outage of a secondary system, or outage of the network to the secondary, is identical as far as Volume Replicator is concerned. When the secondary is no longer available, as evidenced by loss of Volume Replicator heartbeat on the RLink, the primary will simply track the data write changes in the SRL. When the secondary is repaired or network problems are resolved, the SRL will then send all the changes to the secondary location(s). In addition, Volume Replicator can be configured to stop operations at the primary site should a network fail. In this case, the primary will not allow writes until the network connectivity is re-established or the secondary is available for write activity.

Secondary failure

A failure of the secondary would be better defined as a failure of the secondary storage, resulting in a data loss on the secondary side.

There are several methods to recover from a secondary loss. The first is to rebuild the secondary storage and re-initialize using one of the methods discussed above. The second method is to take regular backups of the secondary environment using a Volume Replicator feature called "Secondary Checkpoints." Secondary Checkpoints allow a pointer to be placed in the primary SRL to designate a location where a backup was last taken on the secondary. Assuming the primary has a large enough SRL and secondary backups are routinely taken, a failure at the secondary can be repaired by reloading the last backup and rolling the SRL forward from the last secondary checkpoint.

Primary failure

A failure of the primary can be broken into several possible problems. A complete failure of the primary site is handled by promotion of a secondary to a primary, affecting a disaster recovery takeover. This is exactly the scenario for which Volume Replicator was built.

For primary outages, such as server failure or server panic, the customer can choose to either wait for the primary to recover or shift operations to a secondary server or location.

For situations involving actual data loss at the primary, the customer can shift operations to a secondary or restore data on the primary.

Volume Replicator role changes

Role changes are actions to promote a system that was previously a secondary to a primary. This can be due to a complete site outage where the primary site is not available or simply a role reversal to allow a secondary site to take over operations. Single server or application failures can often be handled locally using Veritas Cluster Server by Symantec by simply restarting applications on a locally-clustered node. In the event of wide-spread failure at the primary site, Veritas Cluster Server can also be used to move the primary or secondary to a new server, for a complete role change.

For example, imagine a two-node cluster at Site A, acting as the Volume Replicator primary, and a two-node cluster at Site B, acting as the Volume Replicator secondary. If a single node in Site A dies, the Volume Replicator primary will simply move to the second node at Site A under Veritas Cluster Server control. The same is true of a single system failure at the secondary Site B. Veritas Cluster Server would restart the Volume Replicator secondary on the opposite system. For situations such as a complete failure of Site A, the Volume Replicator secondary at Site B can be promoted to a primary and applications started to access the underlying data in read-write mode. This is an example of using Volume Replicator to facilitate disaster tolerance for a data center. To automate this entire procedure, Symantec offers global clustering capabilities through Veritas Cluster Server to monitor and control applications at separate sites, connected by replication.

Primary migration

A migration of primary to secondary systems is a controlled shift of primary responsibility for an RVG. Data is flushed from the existing primary SRL if necessary, and then control is handed to the existing secondary. The original primary is demoted to a secondary, and the original secondary is promoted to a primary. This is a very simple operation carried out with one or two commands and allows rapid shift of replication primary between sites. All data outstanding at the primary site is sent to the secondary prior to allowing the migration to take place.

Secondary takeover

A secondary takeover is a somewhat less graceful event, in that the secondary is promoted to a primary without a corresponding demotion of the original primary. These types of migrations typically happen in the event of a complete site outage without any prior notification. When a takeover is accomplished, the secondary is brought up in read-write mode in the exact state it was in at time of takeover. Any data written by the primary in asynchronous mode and not sent to the secondary is not available. After a secondary takeover, the original primary must be resynchronized to be an exact duplicate of the new primary.

Returning to the primary location

After a fail-over has occurred to a secondary location, returning to the original primary can be accomplished using easy failback, a feature in Volume Replicator that allows for rapid resynchronization of an old primary after a secondary takeover. This removes the need for a complete over-the-wire synchronization of all the data or differential-based synchronization.

When a secondary is promoted in a takeover operation, it immediately begins utilizing the DCM associated with each volume. This tracks where data has been written on the new primary. When the old primary comes back online and a failback operation is requested, the new primary communicates with the old primary and determines which data blocks need to be synchronized, meaning only the data blocks that changed after the fail-over occurred need to be resynchronized. This results in the old primary being made an exact duplicate of the new primary in a very short time. This also means that any data that was written on the old primary is permanently overwritten. Volume Replicator makes no attempt to “merge” differences between systems.

Using the secondary system

Enterprise versions of Storage Foundation include point-in-time copy capabilities, allowing a complete mirror break-off (snapshot) of a Volume Replicator volume to be taken and mounted for operation. The benefit of this is that it allows organizations to access the data at the secondary sites for off host processing, such as reporting and backups.

Using In Band Control messages to control FlashSnap™

Volume Replicator provides an advanced messaging capability to control specific events at a secondary from the primary. It does this with In Band Control (IBC) messages sent from the primary to the secondary. IBC messages are placed in the SRL like any other write traffic and are processed in SRL order. For example, consider a database running at the primary site, with Volume Replicator in asynchronous mode. The administrator places the database in hot backup mode to perform an onsite backup. An IBC can be placed in the SRL at this time to signal the secondary when to snap off a mirror, knowing that the IBC will not be received until all data ahead of it in the SRL (right to the time of shifting to hot backup) has been received. As soon as the IBC is entered into the SRL, the database can be taken out of hot backup mode at the primary site. This allows operations at the primary and secondary sites to be coordinated to occur at the exact same time in terms of data consistency.

Volume Replicator in the customer environment

Effects of Volume Replicator on host resources and application performance

Volume Replicator typically has very little effect on host CPU and memory resources and has been measured in the 2 to 5 percent range. This type of CPU usage can be similar to the CPU impact one may notice doing a simple find command in UNIX. Volume Replicator also converts all of the write activity to sequential writes. This is the fastest configuration for writes in most cases, and some customers have observed an increase in write performance using Volume Replicator when compared with not using replication within their environment.

Understanding bandwidth needs using VRAdvisor

In order to replicate data to another location, bandwidth must be available. For environments utilizing Fibre Channel connectivity, Veritas Storage Foundation can be used to mirror the data between the two locations. For environments with IP connectivity, Volume Replicator should be used. Volume Replicator does not specifically require a network dedicated to itself, is resilient to temporary network outages, and includes error-handling capabilities to alert the administrator of critical events.

A very common question is, “How much bandwidth do I need?” The answer is that enough bandwidth must be provided to move all write traffic to each secondary site in any given time period. For example, if 10 gigabytes of data are written in a 24-hour period, then enough bandwidth must be provided to move 10 gigabytes of data in 24 hours. If the configuration is set to synchronous, then attention should be paid to the peak write activity. This activity can impact the performance of the application if the network bandwidth isn’t sufficient for the peak traffic.

The SRL can be used to spool data during time periods when write traffic exceeds replication bandwidth. In order to assist in the proper determination of the bandwidth and SRL size, the Volume Replicator Advisor (VRAdvisor) utility is available for use. The VRAdvisor tool will assist the user in determining the optimal size of the SRL by taking into account the rate of data writes over a given time period, network bandwidth, and different outage durations.

Veritas Volume Replicator Option by Symantec

A Guide to Understanding Volume Replicator

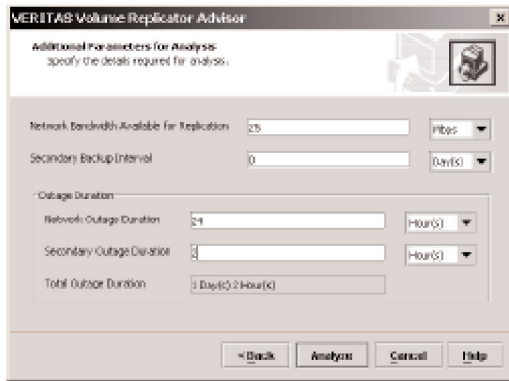


Figure 4. By adjusting input parameters, users can plan for numerous scenarios to help avoid unexpected replication results

Collection of data

The VRAdvisor can collect sample data write statistics based on various parameters. The data is collected over a period of time, in a file that you have specified. If Storage Foundation is installed, then the vxstat command will be used for collecting data. Otherwise, the iostat command will be used. After the data change rate has been collected, the data can then be analyzed to make determinations on the optimal size of the SRL based on the different parameters that you supplied. This result would provide you with the optimum size of the SRL for immediate requirements. You can also calculate the size of the SRL based on the future requirements, changes, and other factors which you are aware of and may affect the SRL.

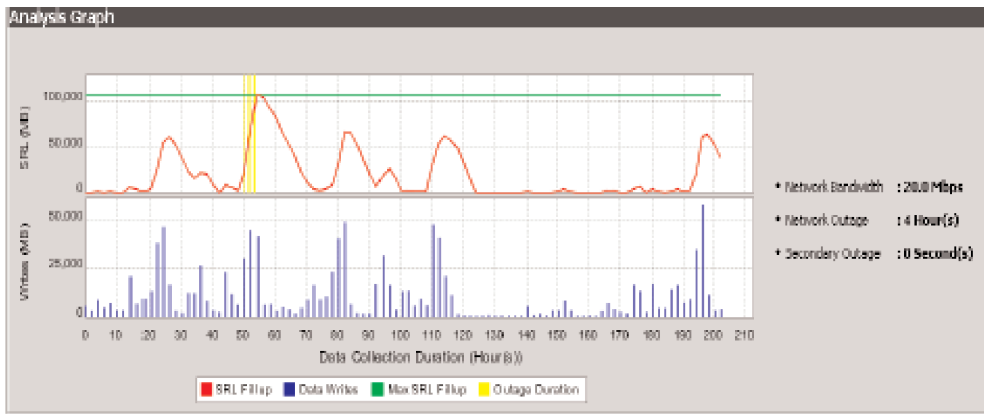


Figure 5. Using VRAdvisor to trend application activity over a period of time helps ensure that bandwidth settings and storage requirements are configured correctly.

Analysis results

Figure 5 displays the analysis results, which are generated based on the inputs that you have specified. The graphical display region displays the analysis results with the help of two graphs. The x-axis for both the graphs consists of the data write duration values based on the information collected on the system. The y-axis of the top graph highlights the SRL fill rate over the data collection period. In addition, the peak SRL fill-up size is indicated against a maximum outage window. This window is displayed in yellow and would indicate a worst-case scenario. The second graph highlights the different write rates at periods throughout the collection period.

Volume Replicator integration with other products

Veritas™ Cluster Server by Symantec

Volume Replicator is a component of an overall high availability and disaster recovery solution. It fits very well into an overall high availability infrastructure provide by Veritas Cluster Server HA/DR.

The full integration of the global cluster capabilities of Veritas Cluster Server HA/DR by Symantec and Volume Replicator provides a powerful disaster recovery solution. Veritas Cluster Server handles local availability issues. Veritas Volume Replicator replicates critical data to a remote site, and the global clustering capabilities of Veritas Cluster Server monitors and manages the clusters at each site. In the event of a site failure or complete failure of applications at the primary site, the global clustering feature of Veritas Cluster Server will control the shift of replication roles to the secondary site, bring up critical applications, and redirect client traffic with a single command or mouse click.

Veritas Storage Foundation for Databases

Veritas Storage Foundation for Databases offers raw device performance with the manageability of the a file system., online administration of storage, and the flexibility of storage hardware independence. Veritas Storage Foundation for Databases can be used within the local environment to maximize the performance of the database, while Volume Replicator can be used to replicate data to the secondary site for disaster recovery protection.

Veritas Volume Replicator Option by Symantec

A Guide to Understanding Volume Replicator

Veritas NetBackup™

In order to have complete disaster recovery solution, every environment should be backed up on a regular basis using NetBackup. By combining NetBackup and Volume Replicator, organizations can be assured that their data is protected.

Replication capabilities summary

The following section will summarize the capabilities of Storage Foundation and the Volume Replicator Option.

Storage architecture independence

Volume Replicator replicates between any major hardware platforms to eliminate vendor-specific storage limitations. For example, using Volume Replicator, customers can replicate between a single vendor's alike arrays, between a single vendor's dissimilar arrays, or between two different vendors' arrays. This means the only architecture restriction for Volume Replicator is duplicate volume sizes at each end. In order to replicate a 200-gigabyte volume, the customer must create a 200-gigabyte volume on the primary and secondary(s). This allows the customer to create a secondary site utilizing older or less expensive hardware and only requires customers to replicate critical data, chosen based on volume, to the secondary site, saving on bandwidth and storage costs. In addition, Volume Replicator provides the flexibility to change the storage configuration as data sizes grow, change, shrink, or are moved—without impacting replication.

Maintaining write order fidelity

Volume Replicator maintains write order fidelity, even in asynchronous mode to guarantee data consistency on the secondary. This is critical to providing a complete, consistent copy of data at a remote site, without requiring synchronous replication.

Native replication over Fibre Channel and IP networks

Veritas Storage Foundation technologies can replicate over Fibre Channel and IP networks natively. Volume Manager can replicate over Fibre Channel and Volume Replicator can replicate over IP networks without the need for any expensive specialized networking devices. In addition, native replication over IP allows organizations to replicate data over any distance.

Veritas Volume Replicator Option by Symantec

A Guide to Understanding Volume Replicator

Scalable

Volume Replicator can scale to up to 31 separate locations for many-to-one and one-to-many replication scenarios.

Initialization options

Volume Replicator can assist in getting your disaster recovery site up and running quickly. There are three initialization options available with Volume Replicator. The first option involves sending all of the data over the wire. The second option is by setting up local mirroring between arrays and shipping the array to the disaster recovery site. The third option that is unique to Volume Replicator combines replication and backup to get the disaster recovery site up and running quickly. The organization performs a normal backup at the primary site and inserts a checkpoint. Then the tapes are sent to the disaster recovery site and a tape restore is performed. Only the data that has changed since the time the checkpoint is inserted is sent over the wire. This allows organizations to get up and running without sending large datasets over the wire or having to pay expensive shipping costs to ship storage arrays. All three operations can be performed completely online.

Summary

Veritas Storage Foundation with the Volume Replicator Option can effectively and efficiently replicate data to another location in order to provide protection from disaster scenarios. This allows organizations to replicate their data between any storage devices, over a standard IP connection, and across any distance for the ultimate in disaster recovery protection.

About Symantec

Symantec is the world leader in providing solutions to help individuals and enterprises assure the security, availability, and integrity of their information.

Headquartered in Cupertino, Calif., Symantec has operations in more than 40 countries.

More information is available at www.symantec.com.

For specific country offices and contact numbers, please visit our Web site. For product information in the U.S., call toll-free 1 (800) 745 6054.

Symantec Corporation
World Headquarters
20330 Stevens Creek Boulevard
Cupertino, CA 95014 USA
+1 (408) 517 8000
1 (800) 721 3934
www.symantec.com

Copyright © 2006 Symantec Corporation. All rights reserved. Symantec, the Symantec logo, FlashSnap, NetBackup, Veritas, and Veritas Storage Foundation are trademarks or registered trademarks of Symantec Corporation or its affiliates in the U.S. and other countries. IBM and DB2 are trademarks of International Business Machines Corporation in the United States, other countries, or both. Microsoft is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries. Other names may be trademarks of their respective owners. Printed in the U.S.A. 09/06 10747367