ABSTRACT
As IT organizations are increasingly faced with more stringent SLAs, data replication has become a common component of many companies’ data protection plans. This white paper defines six types of replication and looks at the advantages and drawbacks of each. It also provides an overview of factors (such as protocol and total cost of ownership) that should be taken into account when developing a replication strategy. This white paper is intended to provide a technical overview as well as high level considerations for IT professionals who are considering augmenting their infrastructure with a replication strategy.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>3</td>
</tr>
<tr>
<td>Replication as part of data protection</td>
<td>3</td>
</tr>
<tr>
<td>Replication overview</td>
<td>3</td>
</tr>
<tr>
<td>Types of replication</td>
<td>4</td>
</tr>
<tr>
<td>Methods by category</td>
<td>4</td>
</tr>
<tr>
<td>Replication modes</td>
<td>5</td>
</tr>
<tr>
<td>Several options available</td>
<td>5</td>
</tr>
<tr>
<td>Synchronous replication</td>
<td>6</td>
</tr>
<tr>
<td>Asynchronous replication</td>
<td>7</td>
</tr>
<tr>
<td>Semisynchronous replication</td>
<td>8</td>
</tr>
<tr>
<td>TCO</td>
<td>9</td>
</tr>
<tr>
<td>Several components play a role.</td>
<td>9</td>
</tr>
<tr>
<td>Connectivity</td>
<td>9</td>
</tr>
<tr>
<td>Connection type</td>
<td>9</td>
</tr>
<tr>
<td>Considerations for assessing service providers</td>
<td>10</td>
</tr>
<tr>
<td>Protocol</td>
<td>10</td>
</tr>
<tr>
<td>Data encapsulation</td>
<td>11</td>
</tr>
<tr>
<td>Remote connection options</td>
<td>12</td>
</tr>
<tr>
<td>Dual data center configurations</td>
<td>12</td>
</tr>
<tr>
<td>Multiple data center configurations</td>
<td>12</td>
</tr>
<tr>
<td>Methods of replication 1</td>
<td>12</td>
</tr>
<tr>
<td>Weighing the pros and cons</td>
<td>12</td>
</tr>
<tr>
<td>Methods of replication: Controller-based</td>
<td>13</td>
</tr>
<tr>
<td>NetApp® Snap-based replication</td>
<td>13</td>
</tr>
<tr>
<td>EMC SRDF replication</td>
<td>13</td>
</tr>
<tr>
<td>Hitachi Data Systems Universal Replicator</td>
<td>13</td>
</tr>
<tr>
<td>Methods of replication: Fabric-based</td>
<td>14</td>
</tr>
<tr>
<td>Methods of replication: Server-based</td>
<td>14</td>
</tr>
<tr>
<td>Overview</td>
<td>14</td>
</tr>
<tr>
<td>Market leaders</td>
<td>15</td>
</tr>
<tr>
<td>Methods of replication: Application-based</td>
<td>15</td>
</tr>
<tr>
<td>Overview</td>
<td>15</td>
</tr>
<tr>
<td>Oracle® application</td>
<td>15</td>
</tr>
<tr>
<td>DB2 application</td>
<td>15</td>
</tr>
<tr>
<td>SQL application</td>
<td>16</td>
</tr>
<tr>
<td>Clustering impact on replication</td>
<td>16</td>
</tr>
<tr>
<td>Overview</td>
<td>16</td>
</tr>
<tr>
<td>Veritas Cluster Server™</td>
<td>16</td>
</tr>
<tr>
<td>Oracle RAC application clustering</td>
<td>17</td>
</tr>
<tr>
<td>IBM Cluster Systems Management™</td>
<td>17</td>
</tr>
<tr>
<td>Methods of replication: Logical</td>
<td>18</td>
</tr>
<tr>
<td>volume managers or SAN</td>
<td>18</td>
</tr>
<tr>
<td>Overview</td>
<td>18</td>
</tr>
<tr>
<td>Symantec Storage Foundation</td>
<td>18</td>
</tr>
<tr>
<td>IBM High Availability Cluster Multiprocessing</td>
<td>18</td>
</tr>
<tr>
<td>(HACMP) replication using LVM</td>
<td>18</td>
</tr>
<tr>
<td>Linux Logical Volume Manager</td>
<td>18</td>
</tr>
<tr>
<td>Methods of replication: Tape cloning</td>
<td>19</td>
</tr>
<tr>
<td>Symantec Veritas NetBackup and EMC NetWorker</td>
<td>19</td>
</tr>
<tr>
<td>Methods of replication: Virtual tape</td>
<td>19</td>
</tr>
<tr>
<td>Data Domain VTL Replication Technology</td>
<td>19</td>
</tr>
<tr>
<td>Summary</td>
<td>20</td>
</tr>
<tr>
<td>Datalink can play a role</td>
<td>21</td>
</tr>
</tbody>
</table>
INTRODUCTION

Replication as part of data protection

For most companies, the difference between a solid data protection plan and a weak data protection plan can equate to hours of lost productivity, system downtime, lost data, disenchanted customers, decreased revenue…the list goes on. For this very reason, prudent organizations invest the resources necessary to ensure their IT infrastructure is built to withstand scenarios that range from small blips of downtime to natural disasters. Before computers, this perhaps meant keeping all important files in a fireproof area under lock and key. In today’s world, it often involves replication of data.

Data replication plays an integral part in a corporate business continuity plan. A replication strategy should take into consideration the structure, usage, and business recovery objectives of the data. The replication strategy should be flexible enough to evolve as the business requirements change.

Replication overview

The definition of replication is to make an exact copy of the original. Replication technology can be used to make both local and remote copies. For this paper, we will focus on the creation of remote copies. Copying data from one location to a remote location can be achieved in a number of ways.

Historically, companies have used tape-based technologies to physically move data from the source location to remote location(s). However, this is becoming especially challenging for many companies as recovery point objectives (RPOs) and recovery time objectives (RTOs) become more stringent. And even though tape-based disaster recovery (DR) solutions are generally robust, they are subject to media failure and degradation. As a result, tape-based DR solutions cannot support increasingly stringent data recovery service level agreements (SLAs).

Therefore, many organizations are augmenting or replacing their existing tape-based DR solutions with disk-based solutions such as replication. Some organizations leverage both tape- and disk-based replication solutions to meet specific SLAs or regulatory requirements. Depending on the regulatory requirement that applies to storing and protecting data, the longevity of the media can also come into question. In this case, tape, optical, disk or replication solutions (or a combination thereof) may be deployed.

Replication technology can be used to make both local and remote copies.
Types of Replication

Types of Replication

Methods by category

Replication of data can be achieved with one or more of these six methods:

1. **Controller-based replication.** This method is focused on the bulk data transfer of the files or blocks under the storage frame from one array to one or more other arrays. Controller-based replication is independent of any one application, meaning that multiple applications or data structures (e.g., databases, file systems or applications) running on a single server or multiple servers can be replicated simultaneously to a secondary server by leveraging tools associated with the storage frame. This type of replication solution allows organizations to "commoditize" the server and application environment and, in some cases, the fabric environment as well.

   Traditionally, controller-based (or hardware) replication is homogeneous, with data copied between disk arrays from the same manufacturer and often within the same family of products. A dedicated transport channel is commonly required to link the two sites.

2. **Fabric-based replication.** This method is focused on the bulk data transfer of the blocks or chunks within the storage network from a source fabric to a destination fabric. Generally, there is performance impact, sometimes just minor, seen within the fabric for "in-band" solutions. This performance impact can be propagated up to the application layer. As a result, this replication type is well-suited for situations where the performance requirements are not as stringent. A major benefit of a fabric-based replication solution is that it commoditizes the storage controller and back-end storage. This means the organization is not tied to the typical costs associated with controller-based replication solutions, and can further reduce costs by using lower cost mid-range storage devices or even entry-level SATA arrays as the replication targets.

3. **Server-based replication.** This method is focused typically on the bulk data transfer of files. Although, with some server-based replication products, the actual block changes for a given file can be replicated from point A to point B. In some cases server-based replication models are substantially cheaper to deploy and manage than other types of disk-based replication. The cost benefits break down, however, when replication must be accomplished on many servers simultaneously. There is also a performance cost associated with this type of replication model that is manifested at the server layer. The classification of the data coupled with the rate of change of the data determines the performance impact. Typically, large block or large file sequential data with high change rates (i.e., OLTP transactional databases, eBay, etc.) are not served well by server-based replication models because of the performance impact on the server.

4. **Application-based replication.** This method is specific to an application, such as a database or web server, and is typically done at a transaction level (whether a table, row or field) by the application itself. If multiple applications are being used on the same server, an application-specific replication solution must be used for each individual application. This replication is independent of the storage frame, fabric and the server type since it occurs at the application layer. The replication is limited solely to the respective application and does not accommodate data structures outside of the control of the application-based replication solution.
5. **Logical volume managers or SAN as replication.**
This method uses the data mirror as an alternative to replication, providing fast recovery in the event of failure. Using a logical volume management (LVM) tool, organizations can create a mirror and keep local and remote parts of the mirror in disparate sites anywhere up to 10 kilometers apart. Organizations that have implemented SAN technology can leverage a mirror to provide data redundancy across a wide area. However, as the distance between the two sites grows, latency begins to impact data consistency and application availability. Additionally, there is substantial cost associated with the infrastructure in connecting two remote sites via Fibre Channel. One way to reduce this cost is to leverage a Fibre Channel protocol conversion utility (FCIP or iFCP). Later in this document, we will outline some differences between FCIP and iFCP.

6. **Tape cloning as replication.** This method copies data from tape to tape and is not normally associated with replication but, for the most part, it accomplishes the same thing. For many years, organizations have generated clone copies of backup tapes as a security tool against site failure. Recovery from cloned tape DR solutions can be a very slow and labor-intensive process to manage. The potential for lost tapes is also becoming a more critical security risk. However, it remains one of the most economical approaches to DR solutions.

**Virtual tape library replication.** VTL to VTL replication provides a viable option for a single replication strategy when dealing with diverse data sets and recovery objectives. Diversity in certain data types, such as database data, file shares or content addressable data, impacts the replication strategy. VTL to VTL replication provides an umbrella to accommodate all data structures through its passive replication of environments with diverse data.

Organizations will ultimately need to look at the type of data that is being replicated when they select the mode of replication.

---

**Replication Modes**

Several options available

Replication implemented at the array, network, or host-level works in one of three modes: synchronous (in database terms most like a two-phase commit), asynchronous (much like a middleware product that leverages a queuing engine), or semisynchronous (a combination of synchronous and asynchronous replication). With each method offering various advantages and disadvantages, organizations will ultimately need to look at the type of data that is being replicated, as well as many other factors, when they select the mode of replication.
Replication Modes

Synchronous replication

In a synchronous replication environment, data must be written to the target device before the write operation completes on the primary device. This assures the highest possible level of data currency for the target device. At any point in time, the target will have the exact same data as the source.

Synchronous replication, however, can introduce substantial performance delays on the source device, particularly if the network connection between the sites is slow or unavailable. Some solutions combine synchronous and asynchronous operations, switching to asynchronous replication dynamically when the performance slows, and then reverting to synchronous when the communication problems are resolved.

When considering a synchronous replication model, an organization must look at the rate of data change, available bandwidth, and distance between sites.

Rate of Data Change—Environments that have a high rate of change will require a broader pipe, either IP or FC. This is the physics of remote replication and usually the most expensive part.

Available Bandwidth—The connection speed between the primary and secondary site will allow for only so much data to be transferred before becoming a bottleneck. Organizations need to consider the volume of data being moved to determine if there will be an impact on the application and then weigh that against the available bandwidth of the connection. The bandwidth must be sized to handle anticipated peak loads.

Distance/Latency—Because every write is written “in-order” to the remote side before the application sees an “I/O complete” message, there can be challenges with latency and slowdown to the application when there is distance between the sites. This is because Fibre Channel requires four round trips to complete every I/O. A loss of approximately one millisecond occurs for every 100 miles traveled. Synchronous replication is, therefore, effectively limited in distance to about 200 miles. After that, application performance begins to degrade to a point where the synchronous replication seldom meets the performance requirements for which the solution was designed. In an OLTP environment or a high transaction oriented environment, synchronous replication is normally not feasible over very long distances. There are, however, variations in impact on performance, depending if the replication is hardware- or software-based.

Software-based synchronous replication, for instance, includes a performance impact on server I/O as the host and target servers must both acknowledge the write commit before the next transaction can be processed.

Hardware-based synchronous replication, on the other hand, offloads the I/O for replication to the arrays that are hosting the storage for SAN-attached hosts. Critical high volume databases are generally better served by hardware-based replication as the wait state between write commits is substantially reduced.

In most cases, the trade-off for the increased performance of a hardware-based replication strategy comes in the requirement to lock into a proprietary replication strategy between like arrays or the fabric. For example, a hardware-based replication solution from EMC® will require EMC arrays at each site. That said, some storage virtualization products, such as Hitachi Data Systems® Universal Replicator, do not necessarily lock in a proprietary replication strategy between like arrays or the fabric. Instead, they can provide replication services to unlike arrays.

Given that hardware-based replication most often leverages Fibre Channel (unless a translation device is used), hardware connectivity requirements for replication between arrays over distance involves investment in separate infrastructure to support replication on the WAN. This is often referred to as channel extension.
Replication Modes

Asynchronous replication

With asynchronous replication, the source system does not wait for a confirmation from the target systems before proceeding. Data is queued and batches of changes are sent between the systems during periods of network availability. Server-based replication tools from vendors such as Symantec™ (Veritas™), Fujitsu, EMC and Topio support asynchronous replication by using a journaling engine as a queuing mechanism to maintain data integrity and ensure data is written to the target in the same order it was written on the source. Some controller- and fabric-based replication applications or appliances support asynchronous replication as well.

The key factors that affect synchronous models of replication also affect asynchronous models, but—because of the queuing engine—not quite in the same manner. In most cases, the queuing engine is bound by the available memory to which it has access. The more memory the larger the queue; therefore, rate of change, available bandwidth and latency do not present as much of a concern. However, if these factors impact the queue in a manner where the queue can not fully empty, the queue itself could become a source of contention, which might require a broader pipe or a closer site.

An asynchronous replication architecture is not as consistent as a synchronous environment, however it can span the earth. The replication is scheduled between the host and target, and changes are cached to be dumped as an “all at once event,” eliminating the issues of distance latency and I/O during busy times on high volume databases.

Software-based asynchronous replication—This form of replication engages a queuing mechanism, typically shared memory from the server, to write the next transaction for processing. Like its synchronous counterpart, software-based asynchronous replication results in a performance impact on server I/O. Also, if the rate of change is greater than the bandwidth, then the queue will fill up and there will be performance issues. Software-based tools handle queuing over great distances. So, generally speaking, the farther data must travel, the more viable this replication type becomes.

Hardware-based asynchronous replication—Hardware-based asynchronous replication offloads the I/O for replication to the arrays that are hosting the storage for SAN-attached hosts. Like software-based asynchronous replication, the hardware-based method engages a queuing mechanism to write the next transaction for processing. The key difference is that the hardware method uses the cache that’s local to the array for queuing (versus at the server level). Hardware-based asynchronous replication also requires like hardware. In order to function properly, the hardware at site A and site B must both be from the same vendor.

It may be difficult to meet all the specific needs with just one method of replication, be it synchronous or asynchronous.
Replication Modes

Semisynchronous replication
Organizations may use literally hundreds of applications, each having different levels of criticality, performance implications, etc. As a result, it may be difficult to meet all the specific needs with just one method of replication, be it synchronous or asynchronous. In these cases, it may be necessary to combine these two replication types for semisync replication. This hybrid approach allows an organization to use the same pipe for different types of replication, thereby allowing for customization to meet the unique demands of the business while at the same time capitalizing on efficiencies.

Software-based semisync replication – Software-based semisync replication can be achieved through sync and async remote copies.

Sync – This method copies all writes to the remote client with an I/O complete only when the data is received at the remote host. This assures maximum transaction consistency and is appropriate for database environments.

Async – This method provides an I/O complete to the host when data is written locally. Transaction consistency is achieved by sequencing each write at the primary site and writing the data at the remote site in the same sequence as the primary.

Hardware-based semisync replication – Hardware-based semisync replication can be achieved with sync, async, and Adaptive copy capabilities.

Sync – Like the software-based method, every write written to the primary side is written "in-order" to the remote side before an I/O complete message is received.

Async – This allows the target site to be one write behind the primary site, which eases latency somewhat.

Adaptive – This method breaks off a mirrored volume and uses it as an image to copy to the remote site. The scheduling of this function should factor in how frequently copies should occur (how current the data will be) and the bandwidth available to do the copy.

The point at which an organization realizes the ROI on a replication strategy varies, depending on the value of the data and the application.
TCO

Several components play a role

The TCO of replication solutions includes many elements—from the obvious to the not so obvious. Generally, TCO can be separated into infrastructure costs and operational costs. The elements listed below include both infrastructure and operational costs and should be included in the TCO equation:

- Hardware required for replication
- Location expenses (floor space, line leases, electrical)
- Personnel to manage the system
- Connectivity: Ongoing costs associated with maintaining an appropriately sized network connection between sites
- Software: Replication technology costs (hard-licensing, soft-management)
- Maintenance: Software and hardware support
- Implementation: From server builds to final installation
- Trending analysis: Understanding the “change” to the rate of change over a period of time to ensure viability of the replication solution

Some of these expenses may be internalized or absorbed through the repurposing of IT resources for the completion of a site-to-site DR strategy. Expenses that can be internalized tend to be the implementation and some layer of physical infrastructure expenses. It is exceedingly rare that an organization will have unutilized dark fibre available between the production site and the DR site and, as such, this is inevitably a new expense item for most organizations.

The point at which an organization realizes the ROI on a replication strategy varies, depending on the value of the data and the application. For instance, through eliminating or minimizing downtime, etc., an organization will recoup the up-front infrastructure costs for a revenue-generating application much more quickly than for a non-revenue generating application.

In general, managing the TCO can be a complex task for organizations. Often they may find it beneficial to work with an outside firm that has experience in defining and quantifying the components that comprise TCO.

CONNECTIVITY

Connection type

Once organizations have designed their replication solution and identified an appropriate disaster recovery site, they can then determine the connection type between the facilities and select a bandwidth provider or service provider.

The bandwidth or service provider responsible for maintaining the connection between the two sites must have an understanding of the organization’s requirements for connectivity and Quality of Service (QOS). After all, a two-hour recovery SLA will produce different requirements than a twominute recovery SLA. In clearly articulating its needs, an organization can achieve the optimal solution for the situation. A DR site is of no value if the links between the two locations are intermittent or susceptible to failure. The SLA may go as far as to understand the physical routing of redundant links. Redundant links offer little value if a single event can disrupt both links simultaneously.
Considerations for assessing service providers

It’s important to note that bandwidth providers sell circuits; whereas service providers are in the business of meeting client needs for connectivity. Following are some key questions that can help in the process of assessing service providers.

- What kinds of pipe connections are offered by the service?
- What kinds of redundancy connections are available across the WAN?
- How quick is the recovery period in the event of a crash or failure? Has this been tested under a variety of failure modes?
- Are recovery procedures documented?
- What kind of power protection is used?
- Is there a remote mirror site for connectivity management? Is it at another physical location?
- What is the uptime history for the service? How long has the uptime history been measured?
- Does the provider have a provision for the payment of liquidated damages in the event that SLAs are not met?
- How is QOS measured? What are the nominal SLAs provided as part of the connection service?
- What references are available for other clients using the service?
- What kind of database software does the service allow access to for connection review and past status on dropped packets?
- What are the bandwidth charges? How are charges measured?
  - By MB of transfer?
  - By throughput on the pipe?
  - By number of pipes for connection (charge for redundancies)?

Protocol

With replication technology, data is moved from point A to point B, regardless of the type of replication or vendor that an organization ultimately selects. This requires a dedicated transport layer, either Fibre Channel or IP-based. In selecting a replication method, following are considerations regarding protocol usage.

- Controller-based replication can be achieved through either Fibre Channel or an IP-based interconnect. If the transport channel is IP-based, this will require a number of other pieces of proprietary hardware, such as storage bridges and routers.
- Server-based replication is implemented in software at the CPU level and is independent of the disk array used. The replication is done using standard protocols such as TCP/IP across an existing network infrastructure.

With replication technology, data is moved from point A to point B, regardless of the type of replication or vendor that an organization ultimately selects.
**Data encapsulation**

Burying Fibre Channel cable underground to move data over a great distance would be extremely costly. Meanwhile, a TCP/IP pipe can be leased rather economically. Conversion devices (such as a Brocade switch or Cisco blades in a director) convert the Fibre Channel protocol to TCP/IP protocol and allow organizations to take advantage of the lower cost transport method.

In determining the connection type for TCP/IP protocol, the two most common transport mechanisms are Fibre Channel Internet Protocol (FCIP) and Internet Small Computer System Interface (iSCSI).

FCIP as a transport mechanism dedicates an organization to using encapsulated Fibre Channel over TCP/IP. This method is more robust and faster than iSCSI, but it also has distance limitations and is more expensive. With iSCSI, the more common transfer protocol, data does not travel at a fixed speed as it does with FCIP. That is, several variables (such as how much pipe is available) can affect the transfer speed. This means it is crucial to determine the amount of replicated data and the time window to access this data. This will ensure accurate sizing of the TCP/IP pipe between the two locations. Too small a pipe is worse than no replication at all, whereas too large a pipe wastes money on unutilized bandwidth. The critical strategic analysis of what data requires a specific level of protection, and an understanding of how much that data will grow, is imperative in building and scaling the appropriate infrastructure to support replication operations. The chart below outlines some of the key pros and cons of the two connection types.

iSCSI allows two hosts to negotiate and then exchange SCSI commands using IP networks. By doing this, iSCSI takes a popular high-performance local storage bus and emulates it over wide-area networks. Unlike some SAN protocols, iSCSI requires no dedicated cabling; it can be run over existing switching and IP infrastructure.

<table>
<thead>
<tr>
<th></th>
<th>Rate Speed (Mbits/s)</th>
<th>Rate Speed (Mbytes/s)</th>
<th>Likely Speed (Mbytes/s)</th>
<th>Time to Replicate 100 GB (hours) 100GB (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>1.544</td>
<td>.183</td>
<td>.175</td>
<td>162.54</td>
</tr>
<tr>
<td>10BaseT</td>
<td>10</td>
<td>1.25</td>
<td>.8</td>
<td>35.56</td>
</tr>
<tr>
<td>Full TR</td>
<td>16</td>
<td>2</td>
<td>1.5</td>
<td>18.96</td>
</tr>
<tr>
<td>T3/DS3</td>
<td>45</td>
<td>5.6</td>
<td>3.6</td>
<td>7.90</td>
</tr>
<tr>
<td>100BaseT</td>
<td>100</td>
<td>12.5</td>
<td>8.5</td>
<td>3.35</td>
</tr>
<tr>
<td>OC-3</td>
<td>155</td>
<td>19.1</td>
<td>15</td>
<td>1.90</td>
</tr>
<tr>
<td>OC-12</td>
<td>622</td>
<td>78</td>
<td>62</td>
<td>.46</td>
</tr>
</tbody>
</table>

*Figure 1. Speed based an circuit type. Theoretical maximum does not include transport overhead.*
Remote Connection Options

REMOTE CONNECTION OPTIONS

Dual data center configurations

The most typical configuration for remote replication utilizes two data centers. The second data center might be an in-region recovery site that is less than 25 miles away, such as a hot site serving a campus-level or metro-level server cluster. Or, the second site might be an out-of-region recovery site hundreds of miles away. Dual data center synchronous replication can provide a very fast recovery time (low RTO) and good data currency (low RPO) when recovering at an in-region hot site. In-region replication, however, does not protect from regional local disasters such as floods or hurricanes, etc. Meanwhile, out of region synchronous replication, which protects from the aforementioned disasters, presents limitations as well. These limitations lead organizations to turn to multiple data center (MDC) strategies in many cases.

Multiple data center configurations

Multiple data center (MDC) strategies provide fast recovery and data currency for local site failures as well as protection from regional disasters.

MDC cascaded synchronous replication supports a current copy of production data at an in-region site. Using asynchronous replication, the in-region site forwards the data to a long distance site. This type of replication often “follows the sun.” For instance, data may be replicated from Los Angeles to New York, then from New York to London, London to Paris, etc.

MDC multi-target synchronous replication supports a current copy of production data at an in-region location while a separate asynchronous replication process copies data from the primary site to the long distance site using a separate dedicated replication network. Multi-target also supports the ability to have a journal at the intermediate site, which will allow replication and disaster recovery to continue asynchronously from the metropolitan site to the remote site after a primary site failure or failover. This type of replication is often used in financial sectors. Data may be replicated simultaneously from New York to London, Paris, Tokyo, etc.

METHODS OF REPLICATION

Weighing the pros and cons

In architecting a replication solution, organizations need to carefully weigh their requirements against the pros and cons of the various methods of replication and determine where trade-offs can be made. To some degree, these same considerations must be taken into account when looking at the different technologies that fall within the solution types, as each of the technologies varies in features and benefits. A summary of some of the leading technologies for several types of replication follows.
Methods of Replication: Controller-based

**NetApp® Snap-based replication**

NetApp provides a simple, efficient, and cost-effective solution for replicating data to local or remote sites for disaster recovery protection. The NetApp philosophy uses a strategy for preserving and retrieving critical data residing in an array of distributed storage systems, desktops, and laptops by leveraging features built into the operating system of the storage array. "Snap features" of the software allow for rapid recovery (in sync, async or multisync states) while lowering the TCO.

**EMC SRDF replication**

Based on the high-end Symmetrix storage architecture, EMC SRDF software is a widely-deployed suite of products built for remote storage replication and leveraged for disaster recovery and business continuity. SRDF uses the local Symmetrix cache as a queuing mechanism to stage block level changes on the source and transmit those changes to the target in either synchronous or asynchronous mode. SRDF leverages a configuration of Symmetrix systems to maintain multiple, real-time copies of logical volume data in more than one location. The source and target systems can be in the same room, in different buildings within the same campus, or hundreds of kilometers apart depending on the mode of replication being deployed.

**Hitachi Data Systems Universal Replicator**

HDS Universal Replicator (HUR) software, in combination with the asynchronous mode of TrueCopy software, supports improved performance and cost reduction in multi-site business continuity configurations. Universal Replicator software uses disk-based journaling and an optimized pull-style replication engine, leveraging the Hitachi TagmaStore® Universal Storage Platform (USP) for replication services. This platform allows virtualization of third-party storage arrays under the control of the USP. By virtualizing a tiered storage solution, replication services can be better aligned with RTO requirements. With this architecture, Universal Replication commoditizes the storage and, as illustrated below, data can be moved between dissimilar storage platforms.

![Figure 2: High-level overview of controller-based replication](image1)

![Figure 3: An application of Universal Replicator’s (asynchronous remote copy) disk-based journaling. OLTP data volumes at the primary data center are being replicated to a second Universal Storage Platform at the remote data center.](image2)
**METHODS OF REPLICATION: FABRIC-BASED**

Often driven by an appliance, fabric-based replication is an in-band replication engine that resides in the SAN fabric. Typically, the appliance captures packets or streams of data as they enter the fabric and replicates them to the remote site asynchronously. Inband solutions allow for the commoditization of the server and storage environment. However, with any inband replication solution, there can be a performance impact on the primary application depending on the performance requirements of the application.

Two leading technologies in this area are FalconStor IPStor Replication and IBM® SAN Volume Controller.

**METHODS OF REPLICATION: SERVER-BASED**

**Overview**

Server-based replication is achieved through software that runs on the server as a service or process. With server-based replication, establishing a host to target relationship is generally completed first at the production site, where the source host data can be initially copied to the target over the internal LAN. In some cases, organizations may do initial “seeding” of the data at the remote site via tape media. During that initial configuration, the relationship between the two servers is established, inclusive of the replication relationship and schedule (full volume copy, file system copy or file based). Once the target is installed at the remote site, network throttling is configured to allow for optimization of replication.

In a large scale replication solution, the organization should ensure that a central management console is available for the oversight of hosts and targets. Additionally, report tools within the console should have the ability to aggregate analysis of the throughput, performance, and consistency of replicated hosts. General considerations for server-based replication include:

1. As a general guide, it is recommended that there be at least 10% more disk space on the target than what is being replicated.
2. When choosing a custom target path, it is important to not select any path with data that should be saved. Replication will, in most cases, overwrite/delete all data in the selected directory.
3. When deploying to a remote machine, it is a requirement to have root or administrator access on both machines that are deploying from (host) and to (target).

Server-based replication products may work extremely well for environments where there are several applications and servers. A management problem arises, however, when there are hundreds of servers from which and to which to copy data. Buying software licenses for 200 servers at the primary location and 200 servers at the remote site is an expensive proposition. Additionally, manually aggregating replication schedules and maintaining and reporting on a per server basis is a “hidden” expense in the use of a one-to-one management tool for server replication.

**Market leaders**

The leading technology for server-based replication in the open systems space is Veritas Volume Replication™ from Symantec. In the mainframe space, the market leader is Softek®.
Methods of Replication: Application-based

Methods of Replication: Application-based

Overview

Several application technologies provide replication features. Figure 5 depicts how this type of replication functions. Some of the common application based solutions are also profiled below.

Oracle® application

Oracle Data Guard can provide replication services controlled by Oracle either synchronously or asynchronously. Data Guard maintains standby databases as transitionally consistent copies of the production database. These standby databases can be located at remote disaster recovery sites thousands of miles away from the production data center, or they may be located in the same city, same campus, or even in the same building. If the production database becomes unavailable because of a planned or an unplanned outage, Data Guard can switch any standby database to the production role, thus minimizing the downtime associated with the outage, and preventing any data loss.

Available as a feature of the enterprise edition of the Oracle database, Data Guard can be used in combination with other Oracle high availability (HA) solutions such as Real Application Clusters (RAC) and Recovery Manager (RMAN) to provide a high level of data protection and data availability.

DB2 application

IBM® DB2 Universal Database version 8.2 offers a high availability disaster recovery (HADR) feature. HADR is a replication that takes place at the database level. The HADR requires two active systems: a primary and a standby. All database transactions take place in the primary system. The transaction log entries are continuously shipped to the standby machine via TCP/IP. The standby system receives and stores log entries from the primary system and applies the transactions to the database. If the primary fails, the standby can take over the transactional workload and become the new primary machine.
Methods of Replication: Application-based

SQL application
Microsoft® SQL Server supports three types of database replication. They are:
- Snapshot replication. This type of replication acts in the manner its name implies. The publisher simply takes a snapshot of the entire replicated database and shares it with the subscribers. Of course, this is a very time and resource-intensive process. For this reason, most administrators don’t use snapshot replication on a recurring basis for databases that change frequently. Snapshot replication is commonly used for two scenarios: 1) for databases that rarely change; and 2) to set a baseline to establish replication between systems while future updates are propagated using transactional or merge replication.
- Transactional replication. This replication offers a more flexible solution for databases that change on a regular basis. With transactional replication, the replication agent monitors the publisher for changes to the database and transmits those changes to the subscribers. This transmission can take place immediately or on a periodic basis.
- Merge replication. This allows the publisher and subscriber to independently make changes to the database. Both entities can work without an active network connection. When the users are reconnected to the network, the merge replication agent checks for changes on both sets of data and modifies each database accordingly. If changes conflict with each other, it uses a predefined conflict resolution algorithm to determine the appropriate data. Merge replication is commonly used by laptop users and others who can not be constantly connected to the publisher.

CLUSTERING IMPACT ON REPLICATION
Overview
Clustering indirectly affects how replication performs from the application viewpoint by making the application aware that data has been replicated. This is important in that if there is a failover from site A to site B, the application needs to know that site B is 100 percent viable. Clustering software is, in essence, a layer between the application and the server. Although it’s not a type of replication, clustering often occurs in conjunction with replication (particularly in the database world) and provides various enhancements.

Veritas Cluster Server™
Veritas Cluster Server (VCS) provides each server in a cluster access to common disk. The common disk includes a volume that is dedicated to the cluster and functions as the virtual node (the virtual server that is made up of the two separate servers and seen by end users). Clustering provides enhanced HA in that if one of the two servers in the cluster were to become unavailable, the functioning server would assume all of the other server’s processes and responsibilities. Most cluster solutions require TCP/IP connectivity for heartbeat and locking for control resources. If an organization wants to stretch the cluster between locations so that all of its applications transparently fail over, the distance requirements for synchronous replication apply. If clustering is completed over long distances with synchronous replication between sites, global clustering (whether it is OS-based clustering or a third party solution like VCS) will not maintain a reliable and consistent connection because of latency.
Clustering Impact on Replication

Veritas Cluster Server™ (continued)

VCS allows companies to link multiple independent high availability clusters at multiple sites into a single, highly available disaster recovery framework. It also enables administrators to monitor, manage, and report on multiple Veritas clusters on different platforms from a single web-based console. VCS is supported on Solaris™, HP-UX, AIX®, Linux, VMWare®, and Windows® operating system platforms.

Oracle RAC application clustering

Oracle Real Application Clusters (RAC) is the premier database clustering solution in the relational database management system (RDBMS) market space. Oracle's RAC configuration options and features offer companies a wide range of flexibility for designing their high availability solutions.

IBM Cluster Systems Management™

Cluster Systems Management (CSM) software provides a distributed system management solution that allows a system administrator to set up and maintain a cluster of nodes that run the AIX or Linux operating system. CSM simplifies cluster administration tasks by providing management from a single point of control. CSM can be used to manage homogeneous clusters of servers that run Linux, homogeneous servers that run AIX, or heterogeneous clusters that include both AIX and Linux.

Most cluster solutions require TCP/IP connectivity for heartbeat and locking for control resources.

Oracle RAC allows an Oracle database to run any package or custom application, unchanged across a set of clustered servers. This provides the highest levels of availability and the most flexible scalability. If a clustered server fails, Oracle continues running on the remaining servers. And when more processing is required, an organization can add more processing power without an outage.

The IBM General Parallel File System (GPFS) is a high-performance shared disk file system that can provide fast, reliable data access from all nodes in a homogenous or heterogeneous cluster of IBM UNIX® servers running either the AIX 5L or the Linux operating system. CSM in conjunction with GPFS provides a possible WAN-based replication solution.
METHODS OF REPLICATION: LOGICAL VOLUME MANAGERS OR SAN

Overview

Logical volume managers (LVM) also indirectly impact replication—not necessarily from an application standpoint, but more as a method to move data within the SAN. Distance limitations apply to LVM tools since latency affects how the LVM tool controls data flow. Typically an LVM tool used for replication would not span city to city replication services. More often, LVM tools are intra-data center in nature.

Symantec Storage Foundation

Storage Foundation uses remote mirroring for replication services. Choosing the mode of replication will determine the transport type. An asynchronous mode over great distance using TCP/IP between the sites requires a protocol conversion device. This has performance implications for the application. A synchronous mode over a short distance using Fibre Channel between the sites does not require a protocol conversion. In this scenario, performance degradation to the application can be minimized.

IBM High Availability Cluster Multiprocessing (HACMP) replication using LVM

HACMP/XD (High Availability Cluster Multiprocessing/Extended Distance) LVM is IBM’s new software-based technology for mirroring business-critical data over standard TCP/IP networks. Using the IP protocol, HACMP/XD GLVM enables an organization to have one or two copies of an AIX 5L logical volume at a remote site separated by unlimited distance. Automated failover to that remote site assures little risk of data loss. GLVM is tightly integrated with AIX 5L’s native Logical Volume Manager for maximum reliability and ease of use.

Linux Logical Volume Manager

Linux Logical Volume Manager version 2 (LVM2) manages storage space, allowing drive space to be added or expanded on the fly without system reboots or shutdowns, giving increased flexibility for operations.

With LVM2, organizations no longer have to worry about how much space each partition will contain. LVM2 also offers enhanced resilience (e.g., new transaction metadata format), user configurability (e.g., device name filters, metadata backup and archive paths, and metadata archive history), new output display tools, and many performance optimizations such as asynchronous snapshots as replication services.

Distributed Replicated Block Device (DRBD) on top of LVM is possible, but care must be exercised in which LVM features are used and how they’re used, otherwise results may not match expectations. Snapshots, for example, won’t know how to notify the filesystem (possibly on the remote node) to flush its journal to disk to make the snapshot consistent.
Methods of Replication: Tape Cloning

METHODS OF REPLICATION: TAPE CLONING

Symantec Veritas NetBackup and EMC NetWorker

Although not widely known as replication products, Veritas NetBackup™ inline backup and EMC NetWorker™ essentially accomplish the same thing as replication. In one sense they could be considered very asynchronous.

Data Domain VTL Replication Technology

Data Domain VTL replication services provide a mechanism for replicating compressed and deduplicated data. This type of technology accommodates a variety of data structures. In environments where synchronous replication is not required and where the business recovery objectives are diverse, this solution can be very effective and easy to manage. Data Domain uses the TCP/IP layer, allowing an organization to passively replicate data from site to site. This application is particularly successful when the business recovery objectives do not require instantaneous recovery.

Regardless of which—if any—replication solution an organization deploys, there are some key elements that must be considered when developing a disaster recovery plan.

Figure 6: Tape-based replication

Figure 7: VTL-based replication
Primary considerations for DR

Regardless of which—if any—replication solution an organization deploys, there are some key elements that must be considered when developing a disaster recovery plan.

1. Business continuity and disaster recovery planning require executive commitment. Without this commitment, it is extremely difficult to design a solution that meets the unique needs of the organization.

2. The data replication strategy should be flexible to accommodate changes in the business requirements.

3. A key step in developing a disaster recovery plan is to identify critical business processes. Once the core processes have been identified, the applications, databases, and supporting infrastructure should be identified and classified. Normally, replication is reserved for revenue generating or otherwise mission critical applications because the TCO associated with replication can be costly.

4. Once an organization has identified the core applications, all application dependencies from a data source perspective should be analyzed.

5. Organizations should perform a business impact analysis of mission critical systems and assess the critical data within those systems. Due to the cost of implementing and supporting a legitimate DR solution, businesses must analyze operational losses as they equate to system downtime.

6. IT managers should work with the respective business units to determine priority levels for restoring mission critical data and non-critical business systems. In the event of an enterprise recovery or data transfer event, bandwidth, time and technical resources are just a few of the critical factors. Key systems and revenue-generating application environments should be first in line during enterprise recovery.

7. IT staff should test critical restore procedures in a real-time environment with the participation of the enterprise’s end-user community to verify viability of the RPO and RTO to any specific SLA criteria.

8. IT staff should be trained how to perform IT recovery without key personnel.

9. Organizations should review contracts with disaster recovery vendors. This includes assessing SLAs on facilities and equipment contracts to confirm that they accurately reflect the current infrastructure.

10. It’s important to require “restore performance expectations” from disaster recovery vendors.

11. A disaster recovery plan should include identification of single points of failure within the IT infrastructure. This should include an analysis and a trending model of the IT infrastructure for replication viability.
Summary

Datalink can play a role

Datalink can help organizations needing assistance in developing and implementing sound data protection strategies.

A complete data center solutions and services provider for Fortune 500 and mid-tier enterprises, Datalink transforms data centers so they become more efficient, manageable and responsive to changing business needs. Datalink helps leverage and protect storage, server, and network investments with a focus on long-term value, offering a full lifecycle of services, from consulting and design to implementation, management and support. Datalink solutions span virtualization and consolidation, data storage and protection, advanced networks, and business continuity. Each delivers measurable performance gains and maximizes the business value of IT.

We have worked with many organizations to help define and implement replication-based DR solutions. We first meet with key management to gain an understanding of business requirements and how those requirements drive application needs and, in turn, storage needs. We can help organizations classify applications and understand the data behind those applications and ultimately build a customized infrastructure that provides maximum data protection.

For more information, contact Datalink at (800) 448-6314 or visit www.datalink.com.