

STORAGE

Buying Guide:

HOST-BASED REPLICATION



inside

- 2 What you need to know about host-based replication
- 8 Special factors to consider

Replication is now a key part of most companies' data protection and disaster recovery arsenals. Although there are several deployment options, host-based replication has emerged as a cost-effective and relatively easy to implement alternative.

Buying Guide:

HOST-BASED REPLICATION

Replication is now a key part of most companies' data protection and disaster recovery arsenals. Although there are several deployment options, host-based replication has emerged as a cost-effective and relatively easy to implement alternative.

BY JACOB GSOEDL

KEY
FEATURES:
HOST-BASED
REPLICATION

SPECIAL
FACTORS TO
CONSIDER

*d*ATA REPLICATION has grown steadily over the past 15 years into a critical component of most companies' disaster recovery (DR) and data protection architectures, especially for applications and data that require recovery point objectives (RPOs) and recovery time objectives (RTOs) that are too stringent for tape-based systems to meet. The replication chore can be performed by storage arrays, on hosts or in the network.

While array-based replication garners approximately 80% of the worldwide revenue in storage-based replication, according to Steven Scully, research manager for continuity, disaster recovery and storage orchestration at IDC, host-based replication, with its revenue percentage share in the low teens, is predicted to have the faster compound annual revenue growth rate until 2012 (close to 20% vs. 10% for array-based replication). If we look at the number of installations, host-based replication, as a result of its low price, simplicity and resulting attractiveness to small- and medium-sized businesses (SMBs), is by far the most widely deployed replication method.

With three competing replication approaches that differ in where the replication occurs—array, network or host—and numerous replication products within each category, any storage replication evaluation needs to start by determining the category that best fits the requirements before choosing the most suitable product within that category.

WHEN HOST-BASED REPLICATION IS A GOOD FIT

Your existing storage infrastructure will largely determine what replication approach to pursue. In environments with homogeneous arrays from a single vendor, there's a good chance the storage arrays already support replication or that the storage vendor offers replication as an add-on.

When considering array-based replication, you must take an inventory of the storage arrays found in all locations that are part of the replication network. More often than not, companies standardize storage arrays in large locations, with no firm standards for smaller sites, requiring a substantial storage investment to equip the remaining locations with storage arrays capable of replication. In environments with heterogeneous storage, array-based replication may be cost prohibitive, so storage-agnostic host-based replication becomes an attractive option.

In environments with heterogeneous storage, array-based replication may be cost prohibitive, so storage-agnostic host-based replication becomes an attractive option.

Network-based replication is an alternative to host-based replication, but usually at significantly higher cost and greater complexity than a host-based replication implementation.

The number of servers that participate in replication is another critical factor. With host-based replication, clients need to be installed and managed on each server, so it becomes less attractive for replication scenarios with a larger number of servers. If you need to support replication networks with a large number of nodes, then centralized array- and network-based replication, which don't require any software installed on servers, usually have the upper hand. Likewise, if you need replication for more than one operating system, or replication of operating systems beyond Windows and Linux, most host-based replication products don't measure up to their OS-agnostic array- and network-based replication brethren.

Contrary to array- and network-based replication, host-based replication adds a few percentage points of overhead to servers' CPU utilization and increases server complexity while potentially lessening reliability. This is one of the reasons why mission-critical applications and applications

KEY
FEATURES:
HOST-BASED
REPLICATION

SPECIAL
FACTORS TO
CONSIDER

with a high transaction volume are more often paired with array-based rather than host-based replication.

The sweet spots for host-based replication are heterogeneous storage environments and SMBs that run Windows and Linux servers. SMBs often have budget and resource constraints, but still need enterprise-level DR and data protection that's only available with replication.

APPLIANCE-ASSISTED OR SOFTWARE-ONLY HOST-BASED REPLICATION

All replication products send a copy of data written to disk on the source server to replication targets. More specifically, the replication facility looks at the address of incoming transactions and, if part of a replication volume, sends them to both disk and the replication target. Where this splitting occurs—host, array or network—determines the replication category. In host-based replication products, the splitting task is performed by the software agent on the server.

While the majority of host-based replication products perform all tasks related to replication on the server, a few vendors have supplemented their products with an appliance in the network to offload replication tasks by forwarding transactions to the appliance instead of sending them directly to destination servers so the appliance can bear the brunt of the replication work. As a result, appliance-assisted host-based replication minimizes the overhead placed on servers by delegating tasks to the appliance. On the downside, these products are more expensive and, since they involve a hardware component, more complex to set up and manage than their software-only host-based replication counterparts. EMC RecoverPoint, FalconStor Software Inc. Continuous Data Protector and InMage Scout are examples of appliance-assisted host-based replication products.

While the majority of host-based replication products perform all tasks related to replication on the server, a few vendors have supplemented their products with an appliance in the network.

Although the three products have much in common, they differ in several key areas. Some appliance-assisted host-based replication products also support network-based replication, as is the case with EMC RecoverPoint. In environments with intelligent Brocade or Cisco fabric switches, RecoverPoint can move the splitting task from hosts to the intelligent switches, eliminating the need to install agents on servers. Another factor to consider is the number of appliances required. While RecoverPoint replicates between RecoverPoint appliances, requiring an appliance (a pair for redundant setups) at the source and destination site, InMage requires only a single appliance

KEY
FEATURES:
HOST-BASED
REPLICATION

SPECIAL
FACTORS TO
CONSIDER

(or pair for redundancy) that's capable of directly replicating to destination servers, yielding a lower overall cost. By replicating between two appliances, EMC RecoverPoint minimizes the overhead on servers to a greater degree than is possible with a single appliance, especially on the replication target.

Contrary to appliance-assisted host-based replication, software-only replication is usually less expensive and less complex, and can be rolled out quicker with less planning required. With no hardware component to be deployed, it only requires the replication software to be installed on servers. Some examples of software-only host-based replication products include: CA ARCserve Replication, CommVault Simpana Replication, Double-Take Software Inc. Availability (Double-Take has been acquired by Vision Solutions Inc.), EMC RepliStor, Neverfail Ltd. Continuous Availability Suite, Quest Software Inc. vReplicator, SIOS Technology Corp. SteelEye DataKeeper, Symantec Corp. Veritas Volume Replicator.

FILE-BASED vs. BLOCK-BASED REPLICATION

Replication products can replicate blocks of data on volumes or logical unit numbers (LUNs), or replication can be performed at the file-system level. In the case of file-based replication, a so-called filter driver captures byte-level file changes as they occur. For block-based replication, changes to blocks on disk volumes are captured in real-time on a write-by-write basis by a filter driver that sits below the file system and usually above the volume manager.

Block-based replication is platform-agnostic and works across various operating systems. Conversely, file-based replication is platform-dependent and requires the development of filter drivers for each supported platform. As a result, block-based replication products are more likely to support a larger variety of platforms than file-based replication products.

Being file-system-aware enables file-based replication products to use file-system commands and file-system metadata during replication. For instance, EMC RepliStor allows flexible inclusion and exclusion of content to be replicated (e.g., certain file types, like *.tmp or *.doc). File-based replication also allows vendors to search the data stream for events like file-save operations, or Exchange and SQL Server start and shutdown events, to determine consistent application states and mark these events to enable application-consistent recovery.

Block-based replication products, on the other hand, depend on proactive consistency, where applications are proactively put into a consistent state through APIs like Microsoft's Volume Shadow Copy Service (VSS) to take application-consistent snapshots, which are then replicated to replication targets.

As far as application integration and support is concerned, almost all host-based replication products support the file system (replication of files

KEY
FEATURES:
HOST-BASED
REPLICATION

SPECIAL
FACTORS TO
CONSIDER

and folders) and Microsoft apps like Exchange, SharePoint and SQL Server. Support for apps beyond these varies greatly by product and vendor.

SYNCHRONOUS vs. ASYNCHRONOUS REPLICATION

Another evaluation criterion is the replication mode a product supports.

Synchronous replication replicates data first and only commits it to disk on the source system after confirming that it was successfully replicated to the target. Synchronous replication depends on sufficient bandwidth and low latency, and supports replication distances ranging from 50km to 300km. It's typically used in apps where zero RPOs and RTOs are required (such as high-availability clusters) and for mission-critical applications that demand 100% synchronicity between primary and replicated data.

Asynchronous replication writes data to disks on the replication source first and, depending on the implementation approach, commits data to be replicated to memory or a disk-based journal that is then copied in real-time or scheduled intervals to replication targets. Unlike synchronous replication, it's designed to work over long distances and greatly reduces bandwidth requirements.

All host-based replication products support asynchronous replication, but only a few support synchronous replication, namely EMC RecoverPoint, SIOS Technology SteelEye DataKeeper and Symantec Veritas Volume Replicator.

All host-based replication products support asynchronous replication, but only a few support synchronous replication.

DISASTER RECOVERY

While all replication products are deployed for data protection and disaster recovery, products vary greatly in the DR options they offer. Some questions to ask while evaluating DR features are:

- How does the product support recovering from a failure of the source system? Does it require rebuilding the source system and then replicating data back from the replication target to the source, or does it provide a failover option?
 - If there's a failover option, what's the process of failing over to the replication target? Is it a manual process or does the product automate failover? What about client failover?
 - Does the product make the necessary Active Directory (AD) and other

KEY
FEATURES:
HOST-BASED
REPLICATION

SPECIAL
FACTORS TO
CONSIDER

changes so clients are redirected to the failover system, or is it a manual process?

- Does the product provide a clean way of failing back to the source system? Does it ensure data consistency and completeness after failing back to the primary system?
- Does it have DR testing options?

All host-based replication products can be used to recover from catastrophic events but there may be significant differences in how quickly they're able to restore service and the degree of automation in the process.

For VMware virtual server environments, VMware provides vCenter Site Recovery Manager to automate failover and integrate with third-party replication products via Storage Replication Adapters (SRA). As a result of their architecture, most host-based replication products can't integrate via SRA. EMC RecoverPoint, as an appliance-assisted host-based replication product, is one of the exceptions, enabling customers to leverage Site Recovery Manager for fully automated failover.

All host-based replication products can be used to recover from catastrophic events but there may be significant differences in how quickly they're able to restore service.

CDP AND NETWORK CONTROLS

Support for continuous data protection is another relevant aspect when evaluating host-based replication products, especially if recovery to prior points in time is required. Most of the products in this buying guide provide some level of CDP, but they vary in the list of applications they support to enable application-consistent rollback. For instance, CA ARCserve Replication provides Data Rewind for Windows, Linux and Unix, supporting application-consistent rollback for Microsoft Exchange, SQL Server and SharePoint. Similarly, EMC RecoverPoint, FalconStor Continuous Data Protector, InMage and Neverfail provide CDP and application-consistent point-in-time recovery for a wide range of applications.

The ability to control the bandwidth required by the replication process is a must-have feature and almost all host-based replication products have some type of bandwidth throttling mechanism. The better a product is able to minimize required replication bandwidth through compression, replication scheduling, etc., the more cost-effective and less disruptive

KEY
FEATURES:
HOST-BASED
REPLICATION

SPECIAL
FACTORS TO
CONSIDER

KEY
FEATURES:
HOST-BASED
REPLICATION

SPECIAL
FACTORS TO
CONSIDER

SPECIAL FACTORS TO CONSIDER

The existing backup infrastructure needs to be taken into consideration when evaluating host-based replication products. We're seeing a trend of backup application vendors adding replication-based data protection to their suites. They view replication as a feature within their backup suites rather than a standalone product. In fact, the majority of backup software vendors already offer host-based replication options for their products: BakBone Software Inc. NetVault Real-Time Data Protector (now part of Quest Software Inc.), CA ARCserve Replication, CommVault Simpana Replication, EMC Corp. RepliStor and Symantec Corp. Net-Backup PureDisk extend these vendors' respective backup suites with a host-based replication option. Similarly, FalconStor Software Inc. has integrated Continuous Data Protector with all major backup suites.

Giving further credence to the notion that traditional backups are converging with replication-based data protection is that some host-based replication product vendors have added backup applications to their portfolio, namely Double-Take Software Inc. Backup (Double-Take was recently acquired by Vision Solutions) and Quest Software vRanger Pro Data Protection Platform (DPP).

Options inherent to operating systems and apps should also be considered when choosing replication products. For instance, Microsoft Distributed File System (DFS) replication in Windows Server can be used to replicate folders and keep them synchronized. Likewise, database servers, such as Microsoft SQL Server and Oracle database servers, provide robust replication options for their respective databases. Even though they don't have many of the features of general-purpose replication products, they perfectly perform the single-purpose task they are designed to perform and are usually the replication option of choice of DBAs and Active Directory administrators.

the product will be. For instance, InMage has integrated wide-area network (WAN) optimization to optimize bandwidth usage and reduce the impact of latency for long-distance replication. ☉

Jacob Gsoedl is a freelance writer and a corporate director for business systems. He can be reached at jgsoedl@yahoo.com.