
Addicted to SAN

There are many replication tools available and no two are the same. It is important to have a clear understanding of your requirements, and then look for a solution that doesn't add an additional layer of complexity.



Introduction

While an irrefutable tenet of physics argues that matter cannot exist in two places at one time, data definitely can and often should. To that end, data replication is an established and straightforward way to protect critical business data. From periodically vaulting large blocks of data to inexpensive cloud storage, to real-time replication of virtual machine images in stretch cluster environments, replication is a reliable and flexible data transfer mechanism.

Today, every storage, backup or virtualization solution vendor offers some form of replication tool, and each tool has a niche in the market, yet many require matching source and target hardware and high capacity bandwidth, or can only support part of your physical or virtual environment. As budgets continue to be cut or frozen and hardware cycles get longer, how can you ensure that you have the right solution to support what you have now and what you will have in the future, and how can you avoid vendor lock-in giving you greater flexibility and negotiating power as you move forward? This paper addresses two keys areas; storage replication and host based replication, and identifies key differences between them. It also highlights the benefits of choosing a host-based solution.

Due to its high level of error checking, synchronous replication was quickly seen as a viable way to replicate data from one array to another.

By frequently copying data from one database to another, database replication serves as the underpinning of highly utilitarian high availability and disaster recovery technology. Distributed databases also allow users to obtain global views of timely data, thus reducing conflicts and inconsistency. While there are many flavors of replication technology, they can generally be grouped into two distinct forms distinguishable by their underlying communication systems.

Synchronous versus Asynchronous

The use of synchronous replication is used in situations where transaction volume is elevated and instantaneous failover, data integrity and consistency are important. To be sure, SAN array-based synchronous replication has many merits. For one, it is another array feature managed in the same way as any other array feature, and therefore, it takes little effort to leverage replication. Because it is an array function, deploying it requires very little involvement of anyone else in the business other than the storage managers. It is the storage group, after all, that makes it work. Provided by the hardware manufacturer, array-based replication is supported by a single vendor so a great deal of finger pointing is eliminated when problems occur.

Due to its high level of error checking, synchronous replication was quickly seen as a viable way to replicate data from one array to another, and storage vendors quickly added options to their storage systems to accommodate additional functionality. Initially this was common to high end arrays, but over time, it has become prevalent in midrange and lower end arrays, including iSCSI.

The most profound limitation of array-based replication appears to be the inflexible requirement for homogenous source and target storage systems. Interestingly, storage vendors often don't support replication between their own heterogeneous storage arrays. And, in the majority of cases replication is an option that is charged for by device or replicated capacity.

Array-based replication is typically replication of entire data volumes at the block level and only deals with the data stored on its own volumes. It does not deal in any way with the host servers connected to it. Block-based replication support works between all platforms that are array attached and leverages compression algorithms, deduplication and encryption. Block-based replication is only replication and does not inherently provide any high availability solution for the servers and clients that are accessing the data. Array-based replication solutions usually require an additional clustering solution to provide high availability.

Through the deployment of virtualization technology, entire server environment can be consolidated into a single server.

The least expensive and probably most straightforward approach to data replication is host-based replication. Since its execution environment is less isolated than array-based replication, applications and server issues can affect the stability and performance of host-based replication. Products will differ in their platform support and often include features such as throttling, compression, de-duplication, encryption, high-availability (HA) capabilities and management options. Host-based replication, a software only approach, is flexible and works with any type of server, storage, virtualization platform or cloud infrastructure; and is capable of supporting a wide range of platforms depending on the replication product and vendor.

The method of replication used in asynchronous host-based or software-based solutions relies on a driver that filters I/Os so files or block level changes can be routed to target servers. Host-based replication is generally file-based, allowing granular selection of data for replication at the file and directory level. Host-based replication solutions reside on the host server and some, like those from Vision Solutions, provide server and application failover for the host server. Host-based replication can replicate at the byte level because operations are replicated from the file system layer instead of the hardware layer. Having granular selection of data to replicate when replicating at the byte level typically reduces the amount of data transmitted during replication, a feature that conserves bandwidth.

Even though it competes in scenarios where array-based replication solutions can be deployed, the fact that it is pure software allows it to extend into areas where array-based replication cannot compete. First, it is the ideal replication method anyone who does not want to invest in SAN or iSCSI storage or can't afford to purchase two arrays. It is also suitable if you cannot afford the bandwidth required for synchronous replication to work. Second, it is for replication to, from or between Cloud storage, Cloud services like the Amazon Elastic Compute Cloud (EC2) simply can't deal with hardware-based replication, but they allow running replication software to exchange data between clouds, and between the cloud and on-premise servers.

The Virtual Effect

Through the deployment of virtualization technology, entire server environment can be consolidated into a single server. While this approach dramatically reduces all of the costs associated with a large server footprint, a single point of failure poses a much greater risk of dramatically affecting productivity and profitability. Consider what would happen to a virtual infrastructure in the event of a connectivity outage to the primary storage unit? The entire virtual host would become unavailable and that can put upwards of 8-10 production workloads at risk. That said, virtualization has actually simplified the backup and recovery process as you only need to backup the virtual disk image to restore the entire virtualized environment, versus trying to backup and recover bits and pieces of a server. Moreover, the disk image can be restored to different hardware or virtual host server.

This is why solutions have been developed to easily move virtual workloads between virtual hosts for high availability. Products like VMware® vMotion, VMware® HA, and Microsoft® Live Migration provide the ability to transfer virtual machines in real-time between similar virtual platforms. VMware® vMotion and VMware® HA are really limited to single site solutions unless you are able to implement storage virtualization across multiple storage arrays,

which of course comes at a price. VMware® does offer Site Recovery Manager which utilizes the replication functionality of the attached storage to replicate the virtual disk images between devices but this solution will be limited in the ways described earlier in this paper. Microsoft® Live Migration, however, is built upon failover clustering technology that allows shared storage between the virtual hosts. This fact, unfortunately, represents a single point of failure. Host-based replication solutions are hardware, storage and virtual vendor neutral, and allow virtual machines to be moved in real-time across any hardware or virtual platform, either locally or between sites.

Vendor Lock-in

One of the biggest issues with vendor lock-in is that if the relationship turns sour, or you decide to move to another vendor, the costs associated with moving away are likely to be high.

Manufacturers of computer peripherals, printers for example, have known for a long time that if they attract customers with low priced gear, they can make a substantial profit over time by selling high profit consumables. This strategy is known as vendor lock-in. As the customer, you are generally unable to opt for more cost effective supplies because they are incompatible with your hardware. There may be benefits to working with one vendor but one would intuitively arrive at the conclusion that numerous disadvantages are far more likely. Vendor lock-in can increase vendor pricing power, but on the flipside, it might mean that you benefit from greater discounts for volume purchases and that you have fewer relationships to manage. In the same vein, the customer should become more important to the vendor. Unfortunately, unless you are spending a considerable amount of money with them, this will not happen. Instead, it can increase exposure to vendor failure and in some cases, limit innovation.

One of the biggest issues with vendor lock-in is that if the relationship turns sour, or you decide to move to another vendor, the costs associated with moving away are likely to be high. This is something worth considering before you invest in two new storage arrays and replication software from the same vendor that will need replacing every three to five years. Additionally, you may not actually need the same high end storage for the second array. Perhaps iSCSI would be more suitable but almost no one offers the ability to replicate between their own storage platforms. As such, host-based replication comes into play again and it offers the flexibility needed to avoid vendor lock-in. You can either purchase high-end arrays from different vendors and use a host-based solution to replicate between them, or you could choose to stick with one vendor and utilize a host-based solution to replicate between a fiber attached SAN and an iSCSI SAN. It could even be iSCSI to NAS if you so desire. Either way you can make choices about your vendor, or vendors, and have the flexibility to purchase the storage that fits your needs from a price and technology point of view. Remember that synchronous replication is not likely to be required on more than five percent of the data stored in your business and in many cases it is much less than five percent.

Re-use or throw it away?

Host-based solutions will provide the most flexible approach to replication, allowing you choose the vendor you want to work with and adapt your infrastructure.

Finally, host-based replication is hardware, storage and virtual platform agnostic so it is possible to recycle existing hardware for your high availability or disaster recovery targets and replace your existing production systems with the latest hardware of choice. Not only does this mean you can extend the life of some of your hardware from the usual three to five years to upwards of ten years, but you also have the flexibility both now and in the future to use the infrastructure that most suits your needs at that time. For example, you may not be ready to make the move to a cloud computing construct right now but want the option to do it in a year without having to replace all of your existing computing assets. As a fringe benefit, you get the flexibility to move between Cloud providers as easily as moving between storage vendors.

Summary

Host-based solutions will provide the most flexible approach to replication, allowing you choose the vendor you want to work with and adapt your infrastructure as your business needs change without having to replace your existing investments. Of course, there are many solutions on the market and no two are the same, so there are still a number of decisions to make. For example, you will also need to consider whether the solution complicates your environment further, or does it simplify management and alerting through a unified console.

Once you have decided that host-based replication is right for your organization you need to build a checklist of functionality and features that it must incorporate. The list should include, but not be limited to the following:

- Operating system support
- Application support
- Recovery Point(s)
- Recovery Time
- Buffering/queuing of data
- Bandwidth limiting and/or scheduling
- Failover
- Unified console
- Failover/recovery testing

Easy. Affordable. Innovative. Vision Solutions.

With over 25,000 customers globally, Vision Solutions is one of the industry's largest providers of high availability, disaster recovery and data management solutions for Windows, IBM i (i5/OS), AIX, Linux and Cloud environments. Vision's MIMIX, Double-Take and iTERA brands keep business-critical information continuously protected and available. With an emphasis on affordability and ease-of-use, Vision products and services help customers achieve their IT protection and recovery goals, which in-turn improves profitability, productivity, regulation compliance, customer satisfaction and quality of life.

Vision Solutions oversees a global partner network that includes IBM, HP, Microsoft, VMware, Dell and hundreds of resellers and system integrators. Privately held by Thoma Bravo, Inc., Vision Solutions is headquartered in Irvine, California with development, support and sales offices worldwide.

For more information call 1 (800) 957-4511 or visit visionsolutions.com.



15300 Barranca Parkway
Irvine, CA 92618
800-957-4511
801-799-0300
visionsolutions.com

iTERA

MIMIX

Double-Take® AVAILABILITY™

© Copyright 2011, Vision Solutions, Inc. All rights reserved. IBM and Power Systems are trademarks of International Business Machines Corporation. Windows is a registered trademark of Microsoft Corporation. Linux is a registered trademark of Linus Torvalds.