Data Protection Methodologies in the Virtual Environment

11/28/07

25 S. Arizona Place, Suite 200
Chandler, Arizona 85225
Phone: (480) 722-1227
Email: info@vigilant1.com
Website: www.vigilant1.com

TABLE OF CONTENTS

1.0 Summary

There are three approaches to data replication in the Client-Server environment and focusing on the SMB (small to medium size business) market;

- Host based
- Array (storage device) based
- Network (appliance) based (a blend of hardware and software)

Before investing in each, and organization must perform some type of Business Impact Analysis (BIA))to determine cost justification. Consider two scenarios;

- Bank or organizations with large amounts of in-flight transactions would need real-time replication (low DRO- see section 4.7.1) if the cost of losing the data justifies the cost to implement the technology (Typically $30K+ leased line)
- A medical office enters Exam data into a Electronic Medical Record system (EMR) that averages 20 exams per hour, would only loose 2 exams with a RTO of 5 minutes for a typical Array based storage system ($10K). Host based, at a cost of $500 per server, would be more cost effective.

This document will review existing technologies related to data replication in the industry and provide insight on recommended uses and applications. A comparison of selected products will be used to demonstrate technology application.

2.0 Data Replication in a virtual environment

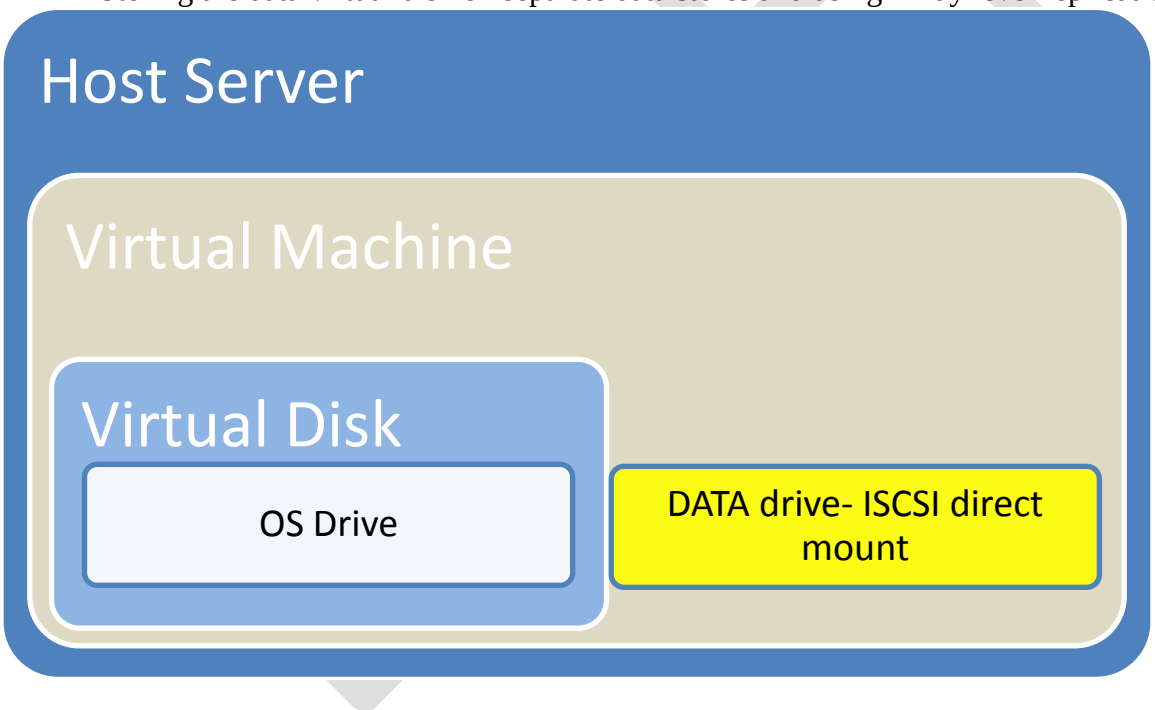
2.1 Overview

Virtualization removes the dependency of the workload on the hardware, resolving the hardware failure scenario, but exposes the next challenge in workload stability; Software /application level failures. Although the Windows 2003 platform and later Linux OS versions are very stable, the application environments system state can be corrupted by a temporary failure, a virtual machine relocation, a OS patch, virtual disk maintenance or software upgrade. Snapshots can protect these Virtual workloads from software failures, but put user data in-between snapshots at risk. Many applications are fault tolerant enough to handle this, but not every scenario is protected.

2.2 Data protection in the Virtual Machine Environment .

Using Server Virtualization in a SAN environment, be design, protects the Virtual workload from hardware failures. However, software failures and in-flight data still put user data at risk. A good strategy for protecting data from snapshots is to manage the data storage independently using these strategies;

- thru direct LUN mapping and using a storage device has snapshot capability
- Storing the data virtual disk on separate data stores and using Array level replication



3.0 Data management concepts

3.1 Overview

3.1.1 Disk Technology

To avoid making this yet another Hard Disk whitepaper, we originally were not going to include a lot of data on hard drives. But there are a lot of myths out there about disk, particularly SCSI, FC & SATA that need to be dispelled. The source of our data comes from several independent studies.
First, we will review the most destructive type of failure in computer systems, a disk failure. A report done on some these studies (CMU & Google) by Storage Mojo, "*Everything You Know About Disks Is Wrong*" essential removes the myth that more expensive "enterprise" drives have lower failure rates then desktop drives, that MTBF is accurate, and further asserts that once one failure happens, the second failure is actually more likely to happen.
Second, lets discuss RAID (redundant array of independent disk). Their a RAID-DP™ uniquely addresses the challenge of providing the highest level of data protection while requiring a minimal amount of storage and minimizing the performance lost from synchronizing data on hot stand-by disk.

	RAID 10	RAID 6	RAID 50	RAID-DP
Raw storage	10000	5625	5625	5625
Useable storage	5000	5000	5000	5000
Maximum storage utilization	50%	88%	88%	88%
Number of drive Failures with no data loss	2	2	1	2
RAID performance level	High	Low	Medium	High

Therefore, your best bet for cost effective data protection; replicate your data off-site somewhere and use RAID DP.

3.1.2 Synchronous vs asynchronous

Synchronous replication provides simultaneous writes of data to the both the local and remote systems at the same time. Asynchronous allows a delay of the data updates to the remote device.

3.1.3 Storage Virtualization

Storage virtualization, as the name implies, allows data storage to be isolated from the storage media (disk). Storage virtualization comes in to varieties;

- Virtual files systems;

In this scenario, the storage management systems emulates a file system and allows files to be distributed across multiple storage devices, including a several different NAS or SAN devices. Systems that access the files are unaware of the true physical locations and as a result, multiple devices can appear as one file system. When this is done by a separate storage appliance with its own metadata to store the file system topology, it is called In-band storage. When this topology is created and managed by the Network OS

(Windows, Netware, Linux, etc…), it is called Out-of-Band, since the file system topology is not dependent on a third party storage device. Because of the proprietary nature of In-band storage systems, the rapidly changing nature of network operating systems (Window, Linux,etc) and the potential performance problems associated Virtual Machine files being split across different devices, it is not recommended to us In-band file systems for Virtual Server applications and a thorough impact analysis should done before implementing in other file management scenarios.

- Virtual block storage;

The other, more useful application of virtual storage allows previously allocated, unused, space in a storage device to be re-allocated for other devices. In this scenario, the local device storage management system allows the client devices OS to allocate space using a pointer, but only the space used is actually provisioned on the disk or disk array. This allows unused disk space to be recovered and has shown an increase of up to 60% in disk utilization

3.1.4 Data replication

Replication is managed in several ways;
Block level replication is replication at the lowest or bit level and is the most efficient method of replications, because only the changed data within a file is updated.
File level replication requires that the entire file be re-written when a change to the file is made. This is very inefficient when large files are updated, as in databases and virtual machines.
Application level replication uses custom application level code to make synchronous writes to multiple data storage systems synchronously. Advance systems can accommodate for latency in the remote systems without causing a performance bottleneck on the local system. This is typically used for databases where client applications are load balanced across several database servers.

3.1.5 Snapshot technology.

Snapshot technology works on the same principle as block level data replication, by storing pointers to the bit-level changes in a file, and allowing administrators access to these pointers. When selecting snapshot technology, choose technology that allows data from a snapshot to be restored at both the partition and file level.

3.1.6 Strategies-Dealing with the Last mile (5 minutes) in replication

Based on the aforementioned data management concepts. A table can be developed to determine the strategies to use for data replication. The determining factor is latency. It should be noted that when the latency tolerances, or ***DRO(*** *Data Recovery Objectives*) in business continuity terms, is greater than 10 minutes, storage based data replication in most cost effective, at levels lower than 5-10 minutes, we consider this real-time data replication, and application level replication, or mirroring is best used. As an example, the cost of Microsoft SQL2005 Enterprise, which supports guaranteed database mirroring, is 5 times the cost of the standard edition, which supports periodic, asynchronous log shipping. A similar comparison of Oracle products indicates a cost factor 20x from Oracle 10G standard to a 10G RAC

A simple formula to determine the cost benefit of crossing the last mile can be found in Business Continuity concepts;

$$\frac{\text{Cost of lost data (@ DRO of Back-up system)}}{\text{Cost of replication system}}$$

Typically, only organizations with high transactional volumes, such as banks, require real time latency. Most other organizations are not going to typically lose the $25K or more in the five minutes of latency that is overcome by SQL2005 Enterprise or $75K by Oracle10G RAC. It should also be noted, that higher-end storage systems, in conjunction with dedicated fibre circuits can also meet real time latency requirements, but the cost is even higher.

4.0 Replication Concepts

4.1 Host Based

Host based replication is typically characterized by an agent or service running within a server (host) that synchronizes data at the transactional, file and block levels. A critical factor of host based system is the ability to clean capture "in-flight" regardless of the OS.

4.1.1 Application level

This is typically a mirror application such as MSSQL or Oracle. This can also be done in an application as well

4.1.2 Operating System.

Usually a host agent that replicates data as it is changed on the host at both the block and file level. This type offers a more granular restore capability, which is useful in protecting against a corrupted OS problem without having to roll back user data.

4.2 Array (storage device) Based

One of the most reliable methods is replication directly from one storage system to another. This type of replication is usually done at a partition level, bit by bit, giving it fairly low bandwidth usage. Array level replication usually requires at least 2 arrays from the same vendor. Most support many-to-one and one-to-many replication topologies

4.3 Network based .

Typically uses a dedicated appliance that attaches to the storage system directly, and uses disk I/O subsystems move synchronize data at the storage system level of a heterogeneous mix of storage systems.
It should be noted, that there are gateway appliances that can provide a 5-10x increase in WAN bandwidth thru proprietary compression algorithms. These compression appliances usally require a device (or agent) on both ends of the transfer to be most effective.

4.4 Replication Chart

Item	Host based -application	Host based -OS	Array	Network
cost	>10k/cluster	$500/host	Cost of 2 arrays (typically 15k or greater)	3k and above
Block level	n	y	y	y
Transactional	y	n	n	n
File level	y	y	y	y
Reliability		good	best	Good-to best
latency	Real time	60 min	>5min Lower latency possible with hi-end arrays/low latency leased lines	>60min

Bandwidth (ex; T1 or greater, DSL/Cable, DS3, any)	Low-medium (Varies by application)	medium (T1 or greater)	Lowest (dsl, cable)	Low (DSL/Cable with compression appliance, usually T1)
Examples	MSSQL 2005 Enterprise Oracle MySQL Enterprise Custom Applications Etc..	Symantec Msft CDP Acronis	Netapps IBM EMC Equilogic Adaptec	Double take Platespin infrastructure FalconStor Riverbed

Notes; data gathered from manufacturer data sheets, published articles and whitepapers, testing and historical data from real world systems; see references.

4.5 Comparison of software based replication tools.

Function	Platespin	Doubletake	falconstor	DPM2007
p2v	x	x	x	
v2p	x	x	x	
p2i	x	x	x	
i2v	x	x	x	
Licensing	per migration OR Source server	Have to have license for source and target	Don't know	
frequency of syncs	1hr or greater	1hr or greater	don't know	1hr
Limitations/notes				has to be on a domain.
Cost	500/host server	3K to 5k		500/host server

4.6 Array device info

Storage Arrays	Netapps	EqualLogic	EMC	IBM	3PAR
Model	S500	PS100E	NS500	n3700	S400, S800
Array bandwidth	65MBps	300MBps	u/a	u/a	
Estimated Pricing	Starts at $5000 2-4k FC upg	$51,405	$40,000 1TB	Starts at $50,000	$115,000.00
Virtual Storage	Supported	Supported	not listed	not listed	not listed
Interoperablity	Supported	not listed	Supported	Supported	Supported
iSCSI	Supported	Supported	Supported	Supported	Supported
Snapshot Technology	250 per Vol	512 per Vol	not listed	SnapVault	Supported
Wan Based Replication	SnapMirror	Supported	Optional	SnapMirror	not listed
V.T.L. Diskbased B/U	Supported	not listed	Supported	Supported	Supported
Tape B/U Fiber Channel	Supported	not listed	Supported	Supported	not listed
Interface	Fiber Channel and iSCSI Up to 64 Luns	iSCSI 3 GB ports/ Optional FC 3 ports	2 FC ports 4 iSCSI ports	2 iSCSI ports 2FC ports per controler	FC-64/128 iSCSI 8/16 ports
Min and Max Space	1TB to 6TB.	3.5TB	64TB	up to 16.8TB	192TB/384TB
Controlers	1	2	1	Single or Dual	2-4/2-8
RAID Configuration	Raid 4,6,DP	Raid 5,10,50	Raid 1,5	Raid 4 ,DP	Raid 0,10,50

Legend
"V.T.L." Virtual Tape Lun
"FC" Fiber Channel

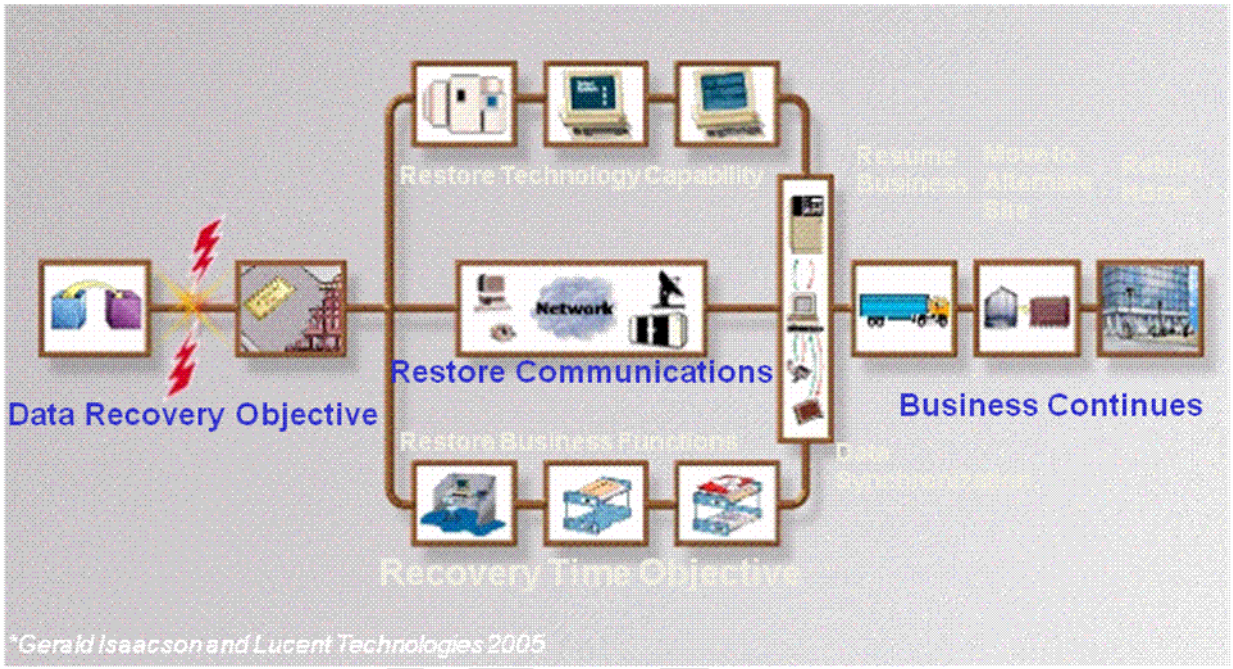
Table of devices with similar features. Price is influenced by capacity, features and performance. For the most part, performance is consistent for all systems and is a function of disk speed and controller capacity. As an example, the S500 cost less, but has less storage and does not support as many host (64 LUNs) . Of course in virtual land, that 32-64 servers, or 32 ESX host!

4.7 Summary

An organization's best approach to Data Replication would be to first perform some type of BIA, to understand DRO and RTO. This will help identify the impact to critical business functions and provide a cost that can be used to determine what solution is appropriate. Then been a table similar to the table in section 4.4 to derive a solution.
IT managers must keep in mind that cost increase drastically in the last mile (last 5 minutes of data latency).
Using a combination tailored to the environment is may sometimes be more cost effective, but must consider the total cost of ownership(TCO) over the long haul as well.

4.8 References

4.8.1 Disaster recovery.



High lights of the process;
Data Recovery Objective (DRO); The point in time to which data needs to be restored to resume operations.
RTO (Recovery Time Objectives) are the times derived from Business Impact Analysis (BIA) surveys that indicate how long the business can survive before critical services are interrupted and the business is impacted. For the purposes of the analysis we used industry standard analysis to mark as critical processes that will have a financial impact of greater than 100K if they are interrupted for a certain period of time or have significant legal, public safety or loss of public confidence impact.

4.8.2 Reference articles

4.8.2.1 STORAGE MOJO
"*Everything You Know About Disks Is Wrong*"

4.8.2.2 BYTE AND SWITCH;
HTTP://WWW.BYTEANDSWITCH.COM/DOCUMENT.ASP?DOC_ID=35892&PAGE_NUMBER=7

4.8.2.3 VIGILANT
Storage research
http://support.vigilant1.com/KnowledgeBase/Research/tabid/58/view/topics/forumid/66/Default.aspx

4.8.2.4 CRN
http://www.crn.com/storage/index.jhtml

4.8.2.5 STORAGE MAGAZINE.
http://searchstorage.techtarget.com/magazineFeature/0,296894,sid5_gci1258347,00.html

4.8.2.6 RIVERBED.
http:// www.riverbed.com/technology/

4.8.2.7 SILVER PEAK. ($9K-50K)
http://www.silver-peak.com/Products/at_a_glance.htm

4.8.2.8 STORAGE NEWS (RSS).
http://feeds.feedburner.com/techtarget/Searchstorage/NewsAndTrends

4.9 Addendum

4.9.1 Platespin Product overview table

PowerConvert Edition	Project	Standard	Universal	Enterprise
	Anywhere to anywhere conversions for server consolidation, hardware migration and data center relocation projects	Workload portability for provisioning, on-going management and hardware independent recovery	Workload portability for provisioning on-going management, replication and consolidated recovery	Enterprise-class workload portability for provisioning, on-going management, replication and consolidated recovery
Features				
P2V	x	x	x	x
P2P	x	x	x	x
V2V	x	x	x	x
V2P	x	x	x	x
Flexible Image Capture	x	x	x	x
Flexible Image Deploy	x	x	x	x
Incremental Transfer	TBD	x	x	x
Server Synch	X	x	x	x
Synchronization Schedules			x	x
File based Live Transfer	x	x	x	x
Block based Live Transfer				X
Purchase Details				
License type	per use	per workload	per workload	per workload
SKU	PC6-P-LIC	PC6-S-LIC	PC6-U-LIC	PC6-E-LIC
Price	$175/conversion	$425/workload	$795/workload	$1495/workload

4.9.2 Microsoft Data protection Manager

4.9.3 Doubletake product overview

	Uses	Cost
Doubletake for windows	Synch data between primary server and alternate server(s)	$3k W2k3 std $5K W2k3 Ent
Doubletake for Virtualization	1. 15 min or 32mb vmdk synch 2. Byte level from SAN	1. 3k/5pac Virtual infra 2. 10K- Virtual systems

Ref link; http://www.doubletake.com/products/default.aspx

4.9.4 Equalogic standard price sheet.

Model #	Drive Type	# C	Disk	RAW	RAID 10	RAID 50	RAID 5	Price
PS50E	SATA-II 7200 250GB	1	7	1.75TB	.65TB	.87TB	1.08	$22,500
PS70E	SATA-II 7200 250GB	1	14	3.50TB	1.30TB	2.17TB	2.60TB	$25,000
PS100E	SATA-II 7200 250GB	2	14	3.50TB	1.30TB	2.17TB	2.60TB	$35,000
PS300E	SATA-II 7200 500GB	2	14	7.00TB	2.60TB	4.33TB	5.20TB	$56,000
PS400E	SATA-II 7200 750GB	2	14	10.50TB	3.90TB	6.50TB	6.80TB	$66,000
PS3700X	SAS 10000 400GB	2	16	6.40TB	2.43TB	4.16TB	4.85TB	$60,000
PS3800XV	SAS 15000 146GB	2	16	2.336TB	0.89TB	1.52TB	1.77TB	$49,000
PS3900XV	SAS 15000 300GB	2	16	4.80TB	1.82TB	3.12TB	3.64TB	$67,000