STORAGE SWITZERLAND

THE VALUE OF HIGHLY AVAILABLE BACKUPS

George Crump, Senior Analyst

The backup process is often considered a secondary process. Because of the number of things that can go wrong, missing a backup or having the completion of the process delayed by a few hours is something the IT Managers have learned to live with. The problem with this is that when something does go wrong, especially due to a backup device failure, it causes a ripple effect that's felt throughout the data center. Steps to build redundancy in the backup process have been expensive but ineffective. Backup virtualization should be looked at as an alternative to standard backup techniques to gain the value of highly available backups.

What Can Fail?

Backup, potentially unlike any other process, interacts with almost every component in the data center. Networks, interface cards, server memory and storage systems all participate in the backup process. Because of the need for production to continue, in most cases there is redundancy built into these systems. The exception, typically at the end of the backup chain, is the target storage device. While disk and tape have some level of redundancy, most almost always have single points of failure as well. But even redundancy has limitations. For example, if a tape drive fails or a drive in a RAIDed disk backup device fails, the speed at which data can be backed up is impacted significantly. In the past this has often been considered an acceptable risk, after all it is a second copy of data. Because of other overnight processes and compliance concerns, it's now more critical that the backup job complete on time, every time. Also, because of the sheer size of the backup job, there is not enough time to get "caught up". If a backup process is missed, it can be rescheduled or it has to be skipped. Of course this puts data and compliance at risk.

What Do You Gain From Highly Available Backups?

The backup is most typically performed at night when the data center is quiet via a backup application that automates the process. The problem is that there are other automated processes also running during this data center quiet time, processes that are dependent on the backups completing successfully and on time. Some of these other processes may monitor the backup process and wait for it to be completed before they are run. A backup failure may lead to them not running at all, and a delay may mean these jobs impact the performance of applications when users return to work. Finally, many of these processes require that the backup job be complete before it's safe for them to perform their work. This is often because these overnight processes or batch jobs manipulate data, and the desire is to have a secure version of that data prior to another process or batch job beginning.

Processes other than those that don't manipulate data, will run independent of the backup but require timing assumptions be made by the process owner. These will either fail if the backup process is still running or be significantly delayed due to resource competition with the backup process itself. For example, imagine a database reindexing process having to run while that database is still being backed up, because one of the drives in the tape library failed and a significant percentage of backup throughput was lost.

Affordable Yet Highly Available Backup Devices

To address these challenges backup managers are interested in putting more availability options into their backup investment. The goal is to do more than just keep the backup process running but to keep it running at top speed. The problem is how to accomplish this goal without liquidating the IT budget.

The first option is to go with 'dual everything', which means putting in two independent tape libraries or disk backup systems and then writing the backup job to both, simultaneously. There is an obvious expense associated with this strategy and frankly it's overkill for most environments. Additionally, this doesn't really speed the process along, since the backup job is being written twice to identical units. In fact, in some cases, the process can actually slow down due to the "mirroring" of the data stream done by the backup application. This means that the backup server has an extra burden placed on it and may slow down as a result. Also, some applications require the purchase of an additional agent (more expense) to perform the actual mirroring function.

A second option that backup managers will consider is a Virtual Tape Library (VTL) device, the goal being to stream data to disk, then to quickly write it to tape. The problem is that most VTL solutions must wait for the entire backup job to complete before movement to tape can occur. Many cannot move data to tape via their own transport, and must copy the data back through the backup server, which then copies it to tape. The VTL is also full of failure points. As is the case with a normal backup process, the devices attached or integrated into the VTL can fail. More importantly the appliance hardware that runs the VTL typically can't be clustered and is itself a single point of failure. While workarounds can be created, it is a very complex process to implement and manage.

Backup Virtualization Simple, Affordable, Highly Available

A logical alternative may be backup virtualization solutions like Storage Director™ from Tributary Systems. These devices have the intelligence to create near-realtime copies of inbound backup jobs in almost any manner required, then subsequently stream that data to tape very efficiently. For example, backup jobs can be sent to a disk system and then quickly streamed to two tape drives simultaneously. In most cases this can be done without requiring additional tape resources. The disk front end allows for better overall transfer speed and in most backup virtualization implementations, drive bandwidth is used more efficiently. An alternative option is to stream the inbound backup data to two hard disk caches simultaneously and then to two tape devices. With backup virtualization, the backup manager has the choice as to how many copies are made and how quickly they are produced.

Most importantly, backup virtualization creates a 'clusterable' front end for backup data to be sent. Multiple appliances can act as one, rerouting data automatically if there is an appliance failure. The result is no loss in backup performance because of a failed appliance.

Backup virtualization addresses the challenge created when trying to design highly available backup infrastructures through an easy to implement and manage clusterable configuration that scales to meet performance and high availability demands. In addition, it allows for the better utilization of existing backup resources and almost always leads to a significant reduction in both backup and recovery times.