



Adding disk to your Backup Solution



By: Jason Iehl, NetApp Consulting Systems Engineer and Chair of SNIA Data Protection Initiative (DPI)

DATA PROTECTION CHALLENGES

All data protection strategies are subject to the same challenges and outside pressures. The strategies that you choose must take into account a number of considerations. First, no organization is exempt from the pressures of managing and storing ever-increasing amounts of data. Your data protection strategy needs to be scalable to accommodate explosive data growth.

All global operations that must be up and running 24x7 are confronting shrinking backup windows. It is not uncommon for full backups to take an entire weekend and for an incremental backup to take all night. At the same time, backups are taking longer to complete and backup windows are getting smaller. Rising costs for backup management are the direct result of the data explosion. Getting data off primary storage and onto backup devices without creating performance bottlenecks can be a big challenge when you have tens to hundreds of terabytes to back up. Compliance requirements and legal discovery demands are growing. Data must be secure, unalterable, and immediately accessible in order to comply with regulations such as the Federal Rules of Civil Procedure in the United States, the European Union Data Protection Directive, and Japan's Financial Instruments and Exchange Law (J SOX).

A typical backup architecture needs to support a wide variety of environments, including networked attached storage (NAS), storage area network (SAN), and direct attached storage (DAS); data center and remote offices; a variety of operating systems; and a range of recovery point objectives (RPO) and recovery time objectives (RTO). When choosing a backup strategy, one must view it from a RTO framework. You should consider a range of solutions that consider your recovery point (RPO) and recovery time objectives (RTO) for your backup data. As an example some of your data may need to be recovered within minutes or even seconds to keep the business running and in other cases, the data may be critical however; as long as you get it back within 12 hours there is minimal impact.

CHOOSING A DATA PROTECTION SOLUTION

In choosing a data protection solution, it's important to understand the varying protection needs for the different applications and data in your environment. For each classification of data, consider these questions for both Operational Recovery and Disaster Recovery:

- **What is the budget you need to work within?**

- **How open are you to changing your existing backup infrastructure or processes to address the backup issues?**
- **Do you plan to continue using tape for onsite and offsite retention?**

With advances in data protection it may be possible to consolidate tape usage at a single site or extend the current tape investment.

- **How much data loss is acceptable in the event of a failure?**
This helps defines the protection frequency. The maximum time requirement between a failure event and a point-in-time to recover from is called the Recovery Point Objective (RPO).
- **How fast do you need to recover the data after a failure?**
The required time it takes to make data accessible after a failure is called the Recovery Time Objective (RTO).
- **What is the operational backup window? How much time do you have to perform backups?**

This defines the backup performance or backup technology required.

Your answers to these questions will help you identify suitable data protection technologies. Remember the answers may be different for each of the various applications that are being protected.

BACKUP TO DISK

Backup to disk, also known as disk-to-disk (D2D), has become a very broad topic and covers many technologies. These include but are limited to backup to conventional disk, Virtual Tape Library (VTL), Snapshots, and Continuous Data Protection (CDP). The first generation of disk backup is known as backup to conventional disk. This type of disk backup, in which backups are streamed to a conventional disk subsystem usually filled with low-cost SATA drives, has been around for a long time. Traditionally high acquisition and management costs have limited the use of conventional disk to only the most recent backup copies. The advantages of conventional disk are improved recovery times, and protection against media failures by means of RAID technology. Higher aggregated throughput can also be achieved using the ability of conventional disk to accept multiple backup streams without the need for multiplex data on tape media. Data on conventional disk can be accessed randomly, significantly increasing the speed of single file, directory or volume recovery. Yet, acquisition costs for conventional disk remain high due to the lack of advanced storage-efficient technologies (e.g., de-duplication).

BACKUP TO VIRTUAL TAPE LIBRARIES (VTL)

Now we get into more advanced methods of D2D backup. First of these are VTLs. VTLs that offer advanced D2D backup features. In this case the VTL is used the same a physical tape library within existing backup applications and they are non-disruptive to both backup administrators and the processes that they manage. Some VTLs also facilitate de-staging data from the VTL to physical tape to accommodate off-site tape vaulting for disaster recovery purposes or long-term archiving. Using a VTL in this configuration greatly improves the overall process of backing up the data and creating the cartridges needed for offsite disaster recovery. Of course another method to this option is to clone offsite. Enterprises for which VTLs are the best option exhibit one or more of the following characteristics:

- Preserve their investments in existing backup software and tape systems
- Use disk and tape in a tiered backup strategy
- Improve an existing tape backup environment that is difficult to change

To meet the backup window, it is essential that the VTL be capable of delivering the aggregate write performance required both today and over the planned life of the system (typically three to five years). To be cost-effective, the VTL must also provide capacity-efficient storage without sacrificing performance using either standard compression, data deduplication, or a combination of the two. To be easily manageable, the VTL should be certified for use with leading backup applications and designed to be managed by a backup administrator. It is important to consider the impact to your existing tape infrastructure. Ideally, you should search for a system that optimizes tape creation performance without increasing the consumption of physical tape media. Most important you should find a VTL solution that offers seamless recovery so that offsite cartridges are easily read back.

SNAPSHOTS WITH REPLICATION

For applications that need increased RTO, we come to disk-optimized solutions where there are many data protection services that capture changes to data to a separate storage location. There are multiple methods for capturing these changes, involving different technologies that serve different needs – they may capture data at the file level, the block level or the application level. This capture may be done on a point-in-time manner or continuously. Snapshot technology is a cost effective way of providing a solution to meet a small

retention time objectives. Remote replication of those snapshots adds a disaster recovery solution without impacting the current application environment.

Snapshots are used to make frequent point-in-time 'snapshots' to provide granular restore points and fast recovery times. By providing Recovery Points as frequently as every minute and by being very network and storage efficient, snapshots, which can refer to both full and delta replicas of data, and when used with frequent replication solutions are a very cost effective for a wide range of applications and data.

A Snapshot and replication solution can shorten backup times from many hours to a matter of minutes without impacting the backup window or application performance. Like many of the newer D2D technologies, snapshots provide more efficient storage through inherent De-Duplication of data—only changed blocks of data are transmitted and stored to provide what is essentially a full backup.

Backup data can be served directly to the backup system. Data can be restored from moments before a failure or disaster. Restores can happen quickly because end users can restore their own data. Snapshot backups can be used for other business needs such as test and development or decision support. It is important that your snapshot with replication solution is integrated with your application to ensure that the resulting snapshot backups are recoverable by your application. Look for a solution that meets your requirements first and then one that provides efficient use of storage capacity. For many environments, an ideal solution will be manageable by your existing backup tools, scalable to 100s of terabytes, integrated with leading enterprise applications, and work for both datacenter and remote office environments. Some solutions can facilitate faster recovery by storing data in native formats so that it can be recovered directly by authorized end users and administrators.

CONTINUOUS DATA PROTECTION (CDP)

Another data protection service that captures changes to data on separate storage device is CDP. CDP records every write with no disruptions to the systems being protected. There is no backup window, and no scheduling to manage. Recovery to any point-in-time may be appropriate for high-end applications with a near-zero RPO, or for environments where IT involvement in the data protection process needs to be minimized - RTO can also be improved with CDP. Recoveries take advantage of high performance disk speed, may be at very fine levels of granularity, and may provide advanced search functionality. A challenge that some,



though not all, CDP solutions may have a lack of application coherence and consistency, is not a challenge that is unique to CDP – indeed, it can also be a problem with snapshots. CDP systems that do not offer application consistency may roll back to any-point in time but they can not guarantee that the application will have been in a consistent state at the chosen recovery time. It is important, when implementing a CDP solution, to assess the disk and network capacity requirements. Because CDP captures every change to data, you will also need to manage your retention policies in line with available disk capacity.

COMPRESSION AND DE-DUPLICATION

With the amount of data growth in the organization it is important to be able to store that data for longer periods of time and more efficiently. This efficiency allows organizations to keep backup copies of data on disk longer, for greater RPO. Compression and de-duplication technologies are becoming more and more mainstream and are designed to help address this problem.

COMPRESSION

Data compression encodes the information to use fewer bits. This reduces the consumption of resources like disk storage and network bandwidth to accomplish greater efficiencies. Generally, software-based compression uses server CPU cycles to achieve its typical 2:1 compression ratio. Hardware-based compression mitigates this performance issue by offloading the compression function to a dedicated chip. The actual compression ratio depends on the type of data being compressed. JPEG files typically don't compress at all, while Microsoft Office documents can be compressed to a tenth of their size. Compression significantly increases the efficiency of some D2D backup solutions.

DE-DUPLICATION

De-duplication is a term used to identify data reduction technologies that eliminate the redundant copying of data at some level of granularity. This type of backup method generates duplicate data as part of the backup process—every full

backup is a duplication of a data set. De-duplication technology eliminates this redundancy without affecting data recovery. By discarding duplicate data strings and updating data pointers, de-duplication reduces disk capacity requirements and enables you to store more backups more cost effectively. All data can be safely recovered in its original format from the disk volume. De-duplication technology makes D2D backup solutions more affordable for a broader range of data. In many cases De-Duplication can work in conjunction with compression to yield even greater resource savings

There are three qualities that you should expect in a de-duplication technology:

- Based on the size of the data blocks being de-duplicated, the more duplicate data will be found.
- The process of de-duplication should not affect backup or recovery performance.
- Enterprise-class reliability and media protection features are a requirement, because a single data block error or media failure can affect multiple backups.

EXTENDING DATA PROTECTION STRATEGIES

D2D backup strategies continue to evolve bringing more value to your business. Advanced D2D backup solutions respond by integrating more specialized technologies. For example, making your backup data more discoverable responds to legal requirements and help streamlines the backup process. Indexing and accessing data can be much faster when backups are on disk. Disk backup makes it possible to quickly search for and locate individual files on storage systems throughout the network—an ability that is invaluable when carrying out an e-discovery search in response to litigation. Classifying data ensures a better data protection practices—you can set up different data protection efforts depending on the type of data. Choosing the right backup solution starts with a careful examination of your data classifications and availability needs. Along with this many D2D back up strategies are coordinated with Disaster Recovery strategies. The scope of data protection continues to mature with demands for insurance of data loss while leveraging additional value of backup data.

The Data Protection Initiative (DPI) within the Storage Networking Industry Association (SNIA) is starting work to bring common definitions and education to the area of de-duplication. Understanding what de-duplication is and the factors that determine an individual space saving ratio are important when designing a D2D scheme that takes advantage of this technology. Both of these organizations mentioned above will be working very closely with the SNIA-Europe and the Data Management Initiative (DMI) in Europe as well.

For more information about SNIA in North America you can visit www.snia.org, and for more information about SNIA in Europe – please visit www.snia-europe.org.