

## Disk-based Data Protection Options

### The Changing Data Protection Landscape

This paper will compare and contrast the benefits of these new disk-based data protection options to allow you to select the best technology for each of the data types to be protected in your environment.



## Introduction

Over the past 40 years of computing, the standard way of protecting primary disk storage data was to backup that data onto one or more magnetic tapes on a daily basis. Magnetic tape will continue to play a critical role in data protection, offering the lowest cost storage medium for maintaining offsite backup data and long term archive data.

But in the last few years, the rapidly falling cost of SATA magnetic disk technology has come closer to the cost of magnetic tape, making it a viable option for specific data protection applications like local data center backup. While SATA is at the core of disk-based data protection systems, there are many system-level implementations, each with its own strengths and weaknesses. This paper will review these systems, which include:

- **RAID Disk Backup:** Simple SCSI, Fibre Channel, or iSCSI to SATA RAID disk subsystems. These low-cost disk storage subsystems, when combined with the latest advances in existing backup applications, can dramatically reduce the amount of manual magnetic tape administration, and increase the performance of restore operations.
- **Virtual Tape Libraries (VTL):** VTL systems represent the integration of low-cost SATA RAID storage with magnetic tape libraries. To the backup application, the SATA disk storage appears as a virtual tape library with virtual media and virtual tape drives.
- **Intelligent Disk-based Data Protection:** These products appear to existing backup, archiving and replication applications as a shared, scalable, virtual pool of SATA disk storage capacity. Existing backup applications can write data directly into the resulting file system to perform backups, usually across a gigabit Ethernet network and through NAS protocols like CIFS and NFS. More than just disks, these systems scale processing power with increased storage capacity. This scalable processing power affords Intelligent Disk-based Data Protection systems with the ability to reduce storage capacity consumed through data reduction algorithms.

## Data Protection Technologies

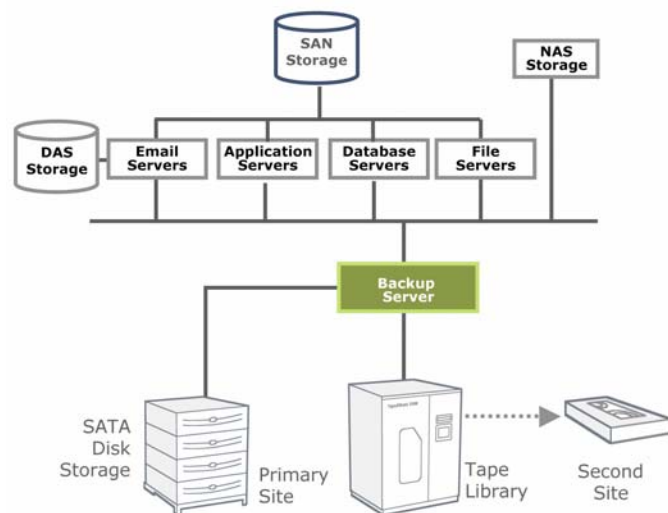
Next, we'll review the three disk-based backup systems and compare and contrast their features and benefits.

1. RAID Disk Backup
2. VTL Backup
3. Intelligent Disk-based Data Protection

### 1. RAID Disk Backup

In the latest releases of traditional backup applications, it's now possible to have email, database and file data backed up to lower cost SATA RAID disks as well as tape. To the backup application, the disk storage system appears as one or more DAS or SAN volumes. Backup data is written to disk as a collection of files within a disk volume's filesystem.

Some backup applications support a disk to tape staging process. In this model, all backup jobs, whether full or incremental, are backed up first to the disk subsystem. Once this has been completed, the disk data is written to one or more tapes. As more backups are performed, the oldest backups are removed from the disk subsystem since they have been successfully written to tape.



Another popular way of using RAID disk storage is to direct all incremental backups to disk while sending weekend full backups to tape. This allows backup tapes to be shipped offsite on a weekly basis and also allows for fast restores from disk-based incremental backups.

Finally, to increase restore times for all backed up files created or modified within the last few months, you can maintain 13 full backups and a number of weeks of incremental backups on RAID disk. For just 3TB of email, database and file data in your environment, you would need as much as 50TB of RAID disk storage capacity, assuming 4+1 RAID 5 disk storage. Even at \$3/GB to \$5/GB, the disk storage system would cost from \$150K to \$250K.

One false assumption that is often made is that RAID disk-based backups are always faster than tape-based backups. The sequential write rates of both RAID disk and tape are about the same, at tens to hundreds of MB/sec.

## Strengths

- Faster restores with the backup data that is still stored on disk.
- Leverages 2:1 compression of backup software applications to double the effective disk storage capacity. This compression ratio is much less than the data reduction rates of Intelligent Disk-based Data Protection systems which are covered later in this paper).
- Higher backup job success rate for disk-based backups.
- Data can be easily “test restored” from disk.

## Weaknesses

- Cost is a major weakness for this solution. For every terabyte of data to be protected, an organization may have to purchase anywhere from 5 to 20 times more disk storage to accommodate weeks of full and incremental backups.
- For disk subsystems less than a few terabytes, disk storage management is not too complicated, but as disk storage subsystems grow to the tens of terabytes, the management of LUNs, volumes and filesystems can become very complex.
- Some backup applications do not handle a disk subsystem running out of available capacity well. This requires constant monitoring of the available space on disk.
- SAN and DAS disk storage systems are not shareable among multiple backup servers without the addition of global SAN filesystem software.
- SAN and DAS volumes are created at a pre-allocated capacity. This may cause some volumes to have an excess of available capacity for new backups while other volumes are out of available capacity.

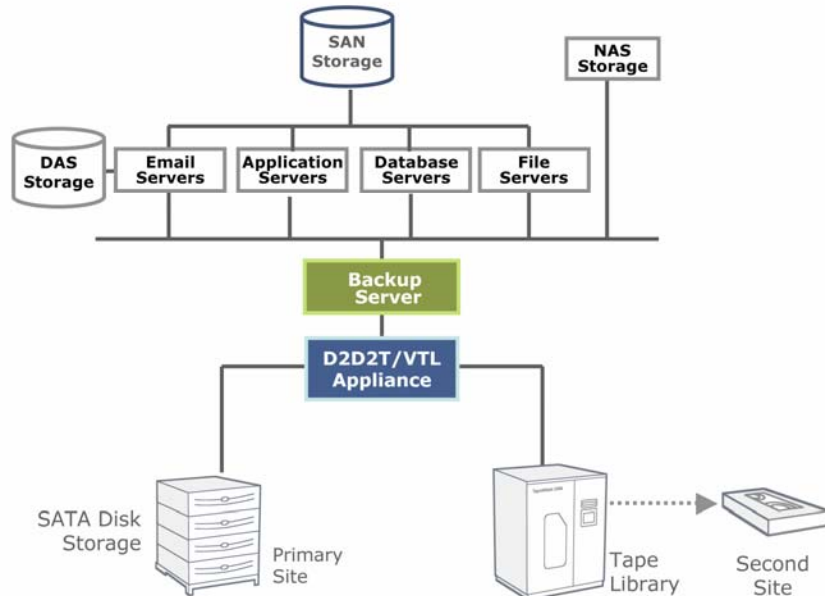
## 2. VTL Backup

VTL appliances take SATA RAID storage and cause the pool to appear as a tape library unit with one or more virtual tape drives and virtual tapes.

As an example, a 3TB RAID5 SATA array may provide 2.5TB of usable storage. 2.5TB of storage capacity equates to about 12 LTO-2 tape cartridges or 6 LTO-3 tape cartridges. As backups are written to these virtual tapes, they are written from the virtual tapes (disk) to real magnetic tapes, usually when an eject command is issued to the virtual tape, signifying the end of a backup run or a full virtual tape. Restores may be satisfied from the limited collection of backup data that is retained in the virtual tape (disk) pool.

## Strengths

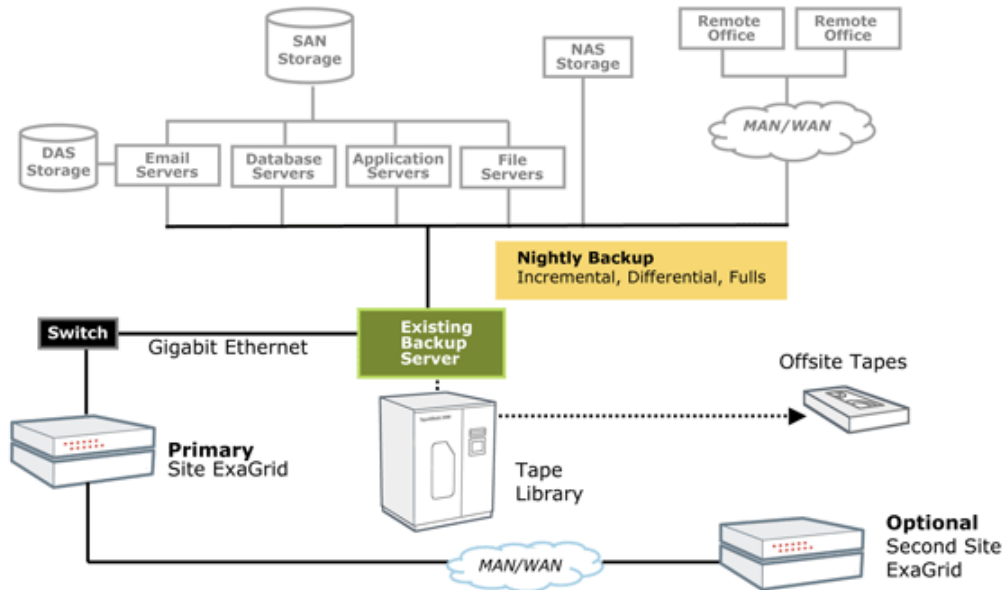
- Seamless integration with all backup applications – disk really looks and acts like tape
- Most VTL appliances have disk subsystems that are optimized to perform backups in the thousands of megabytes per second range, making these VTL appliances much faster than magnetic tape.
- Some VTL appliances perform 2:1 data compression.
- Faster restore times than tape, but only for the small percentage of recently backed-up files that are still resident in the disk array (6 to 12 tapes worth for 3TB of RAID disk).



## Weaknesses

- Many of the vendors in this space do not provide any form of data reduction.
- Cost is a major weakness of this solution. For every terabyte of data to be protected, an organization may have to purchase anywhere from 5 to 20 times more disk storage to accommodate several weeks of full and incremental backups.
- VTL solutions increase the overall cost and operational complexity of traditional backups.
- Some backup applications will treat the disk like tape, which could slow restore operations as various tape search and rewind commands are emulated on disk.
- Some VTL products present complex media labeling consistency issues between barcodes on virtual and real tapes
- Sharing a VTL among multiple servers may require a dynamic drive sharing software license from the backup vendor.
- The VTL appliance, not the backup application, is responsible for actually making physical tapes. This can lead to having to monitor backup job progress at two management points, the backup application and the VTL appliance management application.

### 3. Intelligent Disk-based Data Protection



Intelligent Disk-based Data Protection systems are disk-based data protection systems that integrate easily with your existing backup applications.

Intelligent Disk-based Data Protection leverages grid-based technology to provide a simple-to-expand, virtualization of disk storage into a common shared “repository” for all data to be protected. These repositories are not just simple RAID disk subsystems. They are intelligent servers with CPU processing power as well as one or more terabytes of disk storage capacity.

As the amount of data to be protected grows, additional servers can be added, which not only increases the available storage capacity but also increases the overall processing power of the repository.

What is all of this repository processing power doing for Intelligent Disk-based Data Protection?

- Unlike disk-to-disk-to-tape (D2D2T) or VTL systems, Intelligent Disk-based Data Protection systems deliver data reduction capabilities through the use of byte-level data de-duplication technology. With conventional RAID disk or VTL solutions, 3 terabytes of email, database and file data maintained for 10 weeks would consume 50TB of disk storage, but the same 3TB of data to be protected would only require a little more than 3TB of Intelligent Disk-based Data Protection disk storage capacity.

- Each server within the repository is capacity balanced with all other servers in the repository to ensure that space is always available and that data is evenly distributed across all servers as new servers are added

Single site deployments of Intelligent Disk-based Data Protection systems support the creation of offsite tapes for low cost site disaster recovery. Intelligent Disk-based Data Protection systems are also designed to be deployed across two data centers in order to supplement your existing offsite tape strategy.

## Strengths

- Data reduction delivers system solutions that are from 20% to 50% the cost of equivalent RAID or VTL solutions. Byte-level data de-duplication technology ensures significant reduction in data protection capacity consumption. For example, a 3GB database that is backed up fully every day may experience less than 1 percent change from day to day. Successive days of backup will only store the changes to the database resulting in significant storage capacity savings. At this rate, storing a year's worth of full backups takes just a little more than the capacity of a single full backup.
- Intelligent Disk-based Data Protection systems are simple to install, configure and manage.
- Intelligent Disk-based Data Protection solutions feature front-end InfiniteFilers (CIFS/NFS NAS servers with high performance SCSI RAID disk storage) that interface with the backup, archiving, replication and dump applications. Unlike RAID volumes that have a pre-allocated storage capacity, NAS storage acts as a pool of shared network storage where backup servers that need more storage can get more storage.
- Each InfiniteFiler can accept backup data at rates as high as 250GB/hour. Multiple InfiniteFilers can be added to the system to extend bandwidth to support 1.25TB/hour of backup data,
- Multiple backup servers can share a single InfiniteFiler. Simple global filesystem software can be used to redirect backups to alternate InfiniteFilers when an InfiniteFiler fails.
- The InfiniteFiler incorporates hierarchical storage management software which, to the applications that send data to it, appears like a disk storage system that never runs out of capacity. The result, backups don't fail because of insufficient storage media.
- The repositories are built with servers to form a data protection grid. This grid technology makes the system easy to scale by just connecting additional servers to the integrated gigabit Ethernet switch that forms the "backplane" of the grid.
- Intelligent Disk-based Data Protection eliminates the cost of additional robotics licenses or drive sharing licenses from backup vendors (some VTL implementations require the purchase of these additional licenses)
- The backup application becomes the single point of administration of disk-based and tape based backup copies.
- No virtual tape to physical tape media labeling inconsistency issues, as there can be with VTL solutions
- There is no need to pre-allocate storage capacity to a particular backup server.

## Summary

The three data protection technologies that were reviewed each provide unique strengths for the various kinds of data that require protection from accidental loss, system failure and site disaster. No one solution is best for all data to be protected, but disk can provide greater reliability, performance and security while reducing the amount of manual management of tape in your IT environment.

## About ExaGrid

ExaGrid® is a turnkey solution that works in conjunction with your existing backup applications and is about 25 percent the cost of standard SATA storage. ExaGrid includes SATA storage and provides data reduction through byte-level data de-duplication technology, which stores only byte-level changes for each backup instead of storing full file copies. This unique approach reduces the amount of disk space needed by as much as 20 to 1, resulting in a significant cost savings over standard SATA storage. With ExaGrid, you can replace onsite tape backups with a disk-based backup system, and still maintain your existing offsite tape strategy. ExaGrid is easy to install and use, and fits seamlessly into your existing environment, enabling you to leverage your technology investments.

ExaGrid Systems, Inc.

2000 West Park Drive  
Westboro, MA 01581

**1 800.868.6985**  
**[www.exagrid.com](http://www.exagrid.com)**

