USING DISK TO SOLVE BACKUP AND RESTORE PROBLEMS

Mark Thomason
December 15, 2003
**TABLE OF CONTENTS**

Introduction ........................................................................................................................................... 3  
Why Use Disk as a Data Protection Medium? .................................................................................. 3  
  Speed .................................................................................................................................................. 3  
  Flexibility ........................................................................................................................................... 4  
  Efficiency .......................................................................................................................................... 4  
  Expense ............................................................................................................................................ 4  
Why Tape Is Still Needed .................................................................................................................. 5  
Disk Enhanced Data Protection Methods .......................................................................................... 5  
  Backup to Disk .................................................................................................................................. 5  
  Image Backup .................................................................................................................................. 7  
  Disk Staging ...................................................................................................................................... 8  
  Synthetic Backups .............................................................................................................................. 10  
  Inline Copy ........................................................................................................................................ 10  
  Off-Host Backup .............................................................................................................................. 13  
  Instant Recovery (Virtual Backup) ..................................................................................................... 15  
Problem Scenarios ................................................................................................................................ 17  
  Improving Backup/Restore Performance Scenarios ........................................................................ 17  
  Disaster Recovery Scenarios ............................................................................................................. 19  
Summary ............................................................................................................................................. 20
Introduction

The use of disk for backup and restore offers the backup architect several new options for getting the job done much more efficiently. In fact, there have never been so many ways to backup, restore, store, or position data around the enterprise in such a cost effective and efficient manner. Although several of these technologies have been around for a few years, they can now be coordinated from a central point that nearly all businesses are familiar with today – the backup and restore application.

This article discusses the need for using disk as a data protection medium, explores several disk based and disk enhanced technologies, and then applies them to common backup and restore problems. It is not meant to be an implementation paper, because there are far too many variables to consider, but this paper should serve as a guide and a starting point for deciding which technology to use.

Why Use Disk as a Data Protection Medium?

Several years ago, the storage administrator’s toolbox for backup and recovery was fairly empty. Tape backup was the main tool that administrators used to protect their data since disk was so expensive. As disk prices fell and their capacities steadily climbed, more administrators started using disk as a backup tool to get around the limitations of tape drives.

Over the past few years, several new disk storage technologies have emerged, allowing administrators to manipulate data on disk in a variety of ways. These technologies, together with the continuing drop in disk prices, have added many new and affordable tools to the storage administrator’s toolbox.

In this section, we will review the various reasons for using disk for backup and restore purposes, but the main reason for using disk as a data protection medium is that it enables faster, more flexible backups and restores. Although tape has made significant improvements in throughput and capacity over the past few years, it is still a sequential access medium and is very inflexible compared to disk.

Speed

- **Faster recoveries and backups** – In most cases, backing up to and restoring from disk is faster than using tape. Using disk has the following performance strengths:
  - Disk volumes, especially RAIDs, have a very fast read performance that rivals throughput performance of the newest tape drives.
  - Disk uses snapshot technology, and disk backups and restores are virtually instant.
  - There are no tape mounting or positioning delays compared to a mechanical tape library/device.

- **Potential backup window elimination** – When using disk snapshots and/or off-host tape backup technologies, backup windows can be virtually eliminated since the impact to the production server is “light” (low impact) and quick.
Flexibility

- **Reduces exposure to data loss** – The benefits of disk backups enable faster backups (nearly instant with snapshots) which can be stored very efficiently. This allows administrators to schedule more frequent backups, which lowers the exposure of data loss.

- **Flexible utilization of disk devices** – Disk devices used as backup can support simultaneous backups, restores, and duplication operations if needed. This is impossible with sequential tape devices.

Efficiency

- **Simultaneous backups to disk do not require multiplexing/interleaving** – To back up multiple sources to a single tape drive, a technology called multiplexing, or interleaving, is used. While multiplexing can greatly increase the efficiency of tape devices, its main disadvantage is slower restores. Using disk eliminates the need for multiplexing since disk is inherently a random access device.

- **Data protection operations using disk are potentially more reliable** – Although both disk and tape drives are mechanical devices, tape drives, tape libraries, and especially the tapes themselves have been troublesome in the past, causing delays or failures in backup or restore operations. By using disk as a high-frequency, short-term data protection medium and tape as an archival medium, these problems can be reduced.

- **Creating disaster recovery media from backups made to disk is more efficient** – While duplicating existing tapes to store offsite has been available for years, new data protection features using disk can create synthetic full backups for offsite storage from incremental or differential backups. This scheme requires less data to be duplicated, which leads to lower cost and less tape-drive wear.

- **The added speed and flexibility of disk makes it easier to test disaster recovery procedures** – Disaster recovery plans are useless if they don’t work, so validation that they do work is mandatory. However, most DR plans are not tested due to the time or inconvenience of this complex process. Using disk for recovery can ease the inconvenience and speed up the testing, which makes DR testing a bit easier.

Expense

- **Reduced data protection total cost of ownership** – Through the benefits mentioned above, using disk can increase the efficiency, speed, and flexibility of your data protection scheme. These benefits, along with possibly lowering tape related failures, could help lower the total cost of ownership of storage management.

- **Inexpensive ATA disk drives.arrays offer a compelling case for using disk as a data protection medium** – Prices for ATA drives and arrays have come down to an affordable point for data protection uses, especially when the added benefits of using disk are factored in.
**Why Tape Is Still Needed**

From the list above, it may appear that tape may not have much of a future in data protection. However, tape technology still has some key advantages that disk technology has not (yet) solved:

- **Tape requires less electricity and cooling than disk** – A tape sitting in a box or library needs no electricity or cooling.

- **Tape can handle rough treatment compared to disk** – Tapes can be thrown in a box, survive large drops, and tolerate shipping better than a disk drive, which needs to be carefully packed for shipping.

- **Tape is still much cheaper than disk** – New large capacity tapes are bigger and cheaper than today’s disk drives.

- **Tape is still the best medium for long-term and off-site storage** – Due to the advantages above, tape is still the most efficient way of protecting a business from disaster or for long-term storage. Never underestimate the throughput of a UPS truck.

**Disk Enhanced Data Protection Methods**

This section will introduce some of the popular and leading-edge data protection techniques that use disk to improve performance or efficiency.

**Backup to Disk**

**Cost to Implement: Low to Medium**

Backup to disk is a basic method to utilize disk for data protection. In general, a backup to disk solution writes the same data to a file on a disk volume as it would have to a tape. Therefore, when a backup to disk operation is finished, a single file the size of the backup will exist on the target volume that contains all the files that were backed up.

![Backup to Disk Method](image-url)
### Advantages to Backup to Disk

- **Fast, disk-speed backups and restores** – In addition to the read and write speed advantages of disk drives over most tape drives, disk storage units do not have the mounting or loading delays that tape libraries have.

- **Simultaneous operations (backup, restore, dup)** – Since disk is a random access medium, it can support simultaneous backup and restore operations. However, users should take the time to profile the disk storage unit to see how many operations it can handle before performance degrades to unacceptable levels. It is very easy to overload the disk storage unit with multiple operations.

### Challenges to Backup to Disk

- **Space for backup images must be carefully managed** – Disk implementations have a finite space limit, so careful management of the space available for backups is mandatory. Users should carefully plan what data they plan to use disk backups for and how long they need to retain the images. Using a volume manager, which allows on-line growth of a volume, would help this situation.

- **In most cases, backup images on disk must be migrated to tape for long-term data protection** – Since disk is a finite resource, is more expensive than tape, and the cost of replicating all data to a disk-based DR site is rather large, duplication of disk backup data to tape is a necessity for long-term and DR preparation.

- **Disk storage unit capacity cannot be expanded as easily or cheaply as tape** – Buying a new box of tapes can easily expand tape capacity. However, expanding disk capacity can be significantly more expensive both in acquisition and management costs.

---

**VERITAS NetBackup Advantage**

VERITAS NetBackup™ software has supported backup to disk operations since NetBackup 3.1. Its flexible architecture allows up to 256 simultaneous operations to one disk storage unit. NetBackup software also supports the ability to simultaneous back up to multiple disks OR tape storage units, offering a high degree of flexibility in the data center. NetBackup software can apply one of 25 defined retention periods to disk images, which is similar to tape images. To help with the management of disk volumes, VERITAS Volume Manager can be used to dynamically increase a volume’s size if space gets low, ensuring that backups complete without failure.
**Image Backup**

**Cost to Implement: Low**

Image backup transfers data as a collection of raw disk blocks vs. individual files such as a traditional backup. The main reason for performing an image backup is to avoid the expensive overhead of repeatedly asking the operating and file system for a specific file, which takes up several CPU cycles and system resources; as this is how a traditional file based backup works. Instead, an Image Backup works by first creating a static image of the volume before the backup started (for example, taking a snapshot). Then it starts a transfer of the raw file system data starting and ending at specified block numbers (for example, transfer the contents of block 0 to 10000 to a specified backup device). Blocks of data that change during the backup process are cached so that the image backup always backs up the original data. The efficiency of this transfer method enables image backups to transfer data near the disk subsystem’s maximum throughput.

Although the image backup method is very efficient, there are some major challenges to effectively using it (see below). Depending on the OS, hardware, data type, and number of files, a traditional file backup may perform just as well as an image backup. Therefore, as a general guideline, image backups are normally used when the file system has many files on it (for example, more than 500,000) in applications such as Web servers and large file servers.

![Figure 2. Image Backup Method](image)

<table>
<thead>
<tr>
<th>Advantages to Image Backup</th>
<th>Challenges to Image Backup</th>
</tr>
</thead>
<tbody>
<tr>
<td>• VERY fast and efficient full backups and restores, especially of volumes that contain many files.</td>
<td>• Restores of individual files from an image backup require advanced technology — To restore a single file from an image backup, the backup application must have a “map” of the image to know which blocks make up the specified file; and typically the blocks that make up a specified file will be scattered throughout the image.</td>
</tr>
<tr>
<td>• Low host server impact of backups – Image backups use less CPU resources.</td>
<td>• Safe backups of in-use or on-line data require advanced technology – If blocks of data are changing on the disk being backed up, then a snapshot method must be used to freeze the data so that a consistent image is backed up.</td>
</tr>
<tr>
<td>• Less disk drive wear and tear – Image backups can drastically reduce the amount of disk thrashing (seek) vs. a traditional file backup.</td>
<td></td>
</tr>
</tbody>
</table>

VERITAS ARCHITECT NETWORK
**VERITAS NetBackup Advantage**

The VERITAS NetBackup 5.0 Advanced Client option includes an image backup feature called FlashBackup that supports UNIX and Windows platforms. FlashBackup addresses all the challenges above by using snapshot technology that safely backs up live volumes and mapping technology that allows restoration of any number of individual files or directories in one pass. NetBackup software goes one step further than the competition by offering differential and incremental image backup, which increases both the backup speed and flexibility. Backup performance improvements of 600–1000% are commonplace for Web and file servers containing millions of files.

**Disk Staging**  
**Cost to Implement: Medium**

Disk staging is a method that uses disk as a cache during a backup to tape. The goal of this scheme is to blend the best attributes of disk and tape into one automated feature. This feature works basically in the following way: The target data is backed up to the disk cache first, and then the backup image is migrated or moved to tape after a predetermined time.

In the data protection market today, there are several incarnations of disk staging that are implemented in both software and hardware. The hardware implementations typically appear as virtual tape drives to the backup application – the backup application doesn’t really know it is using up the hardware-based disk cache as the device handles the migration to tape and management of the images. Software disk staging implementations can range from a very simple and manual process to a highly sophisticated and automated process, depending on the vendor.
Using Disk to Solve Backup and Restore Problems

Advantages to Disk Staging

- Flexible backup schemes utilize the best of disk and tape – This allows efficient and flexible schemes such as nightly backups to fast disk (to shorten backup windows) and then transferring to tape during the daytime.

- Efficient use of disk and tape resources – Fast backups to a finite amount of disk combined with “infinite” tape capacity. In addition, since disk supports multiple backups, disk can be used when tape devices have limited availability.

- Faster restores from disk-based images – While data is still on disk, restores can be quickly performed with no tape device or library latency.

- Can act as a buffer/cache for very fast tape drives – Some of today’s high performance tape drives are capable of very high throughputs. If the incoming data is too slow, the drive’s performance will plummet due to “shoe shining” the media. Using tape multiplexing can help, but there may be instances where it is impossible to “feed” the drive enough data to stay efficient. In these cases, disk staging may be a perfect solution to cache the data on disk before streaming it to a high-speed tape drive.

Challenges to Disk Staging

- Additional management needed to coordinate the migration process – Unless the backup-to-disk solution is automated, the solution will require additional management to coordinate the movement of data.

- Large disk cache size needed – Unless the backup to disk solution has an automated way of managing the images in the disk cache, then the disk cache must be large enough to fit all backups targeted to the cache until the cache can be flushed of expired images.

VERITAS NetBackup Advantage

VERITAS NetBackup software has two methods for performing disk staging:

1. NetBackup 5.0 software now includes a completely automated disk staging solution based on a new device – Disk Staging Storage Unit (DSSU) – which can be easily targeted with a backup policy. The DSSU is configured with a disk volume(s) to use and a target device(s) to migrate the data to. Once backups are completed to disk, they are migrated to tape on a schedule. If the DSSU becomes full during a backup, disk images that have been migrated to tape are deleted. If a restore is started and the image is still on disk, then the data will come from the disk backup; if not, a tape restore will automatically be accomplished. The VERITAS NetBackup software inline copy can also be used to duplicate the data, either before the data is backed up to disk or when it is copied to tape.

2. The NetBackup Vault option includes a disk staging feature where staged disk backups can have a unique retention period from the copy(ies) on tape. This allows an additional level of flexibility to control how long images are kept at a location for those enterprises which have strict SLAs on recovery time.

When combined, NetBackup software’s disk staging features offer performance and flexibility that are unmatched in the industry.
**Synthetic Backups**

**Cost to Implement: Low**

A synthetic full backup is assembled from a previous traditional full backup and the subsequent incremental and/or cumulative (in other words, differential) backups. The goal is to create a full backup image from existing backup data that is nearly identical to a full backup performed at the time of the last incremental or differential backup – with no data movement from the actual client (for example, no client or LAN impact). The ability to create synthetic fulls can be a big advantage for IT organizations under regulatory or SLA guidelines to generate frequent full backups when it is too expensive in resources to perform "real" full backups.

<table>
<thead>
<tr>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
<th>Sat/Sun</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inc</td>
<td>Inc</td>
<td>Inc</td>
<td>Inc</td>
<td>Syn Full</td>
</tr>
<tr>
<td>Full</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inc</td>
<td>Inc</td>
<td>Inc</td>
<td>Inc</td>
<td>Inc</td>
<td>Syn Full</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is also possible to create synthetic cumulative backups, which consolidate several incremental backups into one synthetic cumulative image. This is advantageous when incremental backups are performed often and the overall dataset is too big to efficiently do frequent full backups. These synthetic cumulatives can then be used to create synthetic fulls on a monthly or quarterly basis.

To create a synthetic full backup, selected data from the initial full backup and the subsequent incremental or differential backups are copied to a new tape in the form of a single image. If all backup data is on tape, then two tape drives are needed to act as a source for the original backups and destination for the synthetic backup. To construct a synthetic backup, the backup application not only must keep track of which files were present on the volume for each backup, but it also must know what data was moved or deleted since the full backup.

Using backup-to-disk with synthetics can dramatically raise the usability and efficiency of synthetic backups.

If backup-to-disk is used for the smaller, more frequent incremental backups, it reduces or removes the need to mount potentially many tapes to create the synthetic fulls or synthetic cumulatives (differentials). This ultimately saves time, uses disk and tape drives for what they are best at (tape for long writes/reads and disk for short, high frequency writes/reads), and reduces the wear and tear on tape drives and libraries.
Using Disk to Solve Backup and Restore Problems

Advantages to Synthetic Backup

- **No impact to client resources** – Synthetic backups are created using existing backup data and do not impact the original client resources (for example, CPU or disk volumes).
- **No network impact** – Since data is not moved from the client, synthetic backups do not impact network utilization.
- **Reduced recovery (DR) time** – Using synthetics may reduce the number of tapes needed for disaster recovery, which may reduce the time needed for recovery.
- **Efficient use of disk and tape resources** – Using disk for small, high-frequency incremental backups and tape for large full backups can leverage the best characteristics of the media while still being able to create synthetic fulls or cumulatives (differentials).
- **Conveniently create synthetic single image full backups for SLA and regulatory needs** - When it is not possible to impact the client resources for a full backup, synthetic backups can still be accomplished since they do not impact the client.

Challenges to Synthetic Backup

- **The metadata (for example, security, attributes) about the files in a synthetic backup may be outdated** – Since synthetic backups manipulate data that has been backed up in the past, they may not reflect a file’s current metadata if any changes to security or file attributes have been done. Here are some exceptions and workarounds:
  - On Windows platforms, changes to security and encryption attributes will reset the archive bit, assuring that these changes will be captured on incremental backups, but changes to other attributes will not.
  - On UNIX platforms, changes to ACLs on file/directories change the ctime, not mtime (which is the usual trigger for inc backups). If this is an issue, the administrator can set ‘USE_CTIME_FOR_INCREMENTALS’.
- **Synthetic full and cumulative (differential) backups are only as up-to-date as the last actual incremental backup** – Since synthetic backups only use existing backups, they are only as current as the last incremental backup. This is called a coverage gap.
- **Synthetic backups usually require two tape drives** – Unless all original backup data is on disk, creating a synthetic backup will require the creation of two tape drives – one for the source data and one for the destination data.

**VERITAS NetBackup Advantage**

VERITAS NetBackup 5.0 software now includes synthetic backups as a new standard feature. Although other competitors may have synthetic backup capabilities, NetBackup software is the only product able to create a single image synthetic backup which appears similar to a real full backup (vs. a collection of full/inc backup images). In addition, NetBackup software is the only application which can create a cumulative (differential) backup from a collection of incrementals.
**Inline Copy**

*Cost to Implement: Medium*

Inline copy is the ability to write backup data to multiple destinations (disk or tape) simultaneously. Before inline copy, backup applications duplicated data as a secondary process after the initial backup from the client was finished. Therefore, the goal of inline copy is to enable flexible and efficient duplication of data to several types of media at the same time. Inline copy can greatly ease an organization’s electronic vaulting efforts by combining two operations (backup and duplication) into one, thus conserving time and IT resources.

An application of this feature would be to perform a backup once from the client to three copies: 1) one copy on disk for fast restores which expires within a week, 2) one tape copy for long-term recoveries onsite and, 3) one tape copy for offsite storage that may never expire. In this example, the data center is well positioned to handle a variety of recovery needs in basically the time it takes to perform the initial backup without the expense of performing post-process duplication later.

![Figure 5. Inline Copy Method](image-url)
Using Disk to Solve Backup and Restore Problems

<table>
<thead>
<tr>
<th>Advantages to Inline Copy</th>
<th>Challenges to Inline Copy</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Creation of duplicate copies simultaneously with the initial backup</td>
<td>• Backup speed will be tied to the slowest backup device – Since a single backup stream is going to multiple</td>
</tr>
<tr>
<td>– Instead of performing duplication as a post process after the backup, inline copy</td>
<td>destinations, the stream can only be as fast as the slowest device used.</td>
</tr>
<tr>
<td>can be used to create duplicate copies during the initial backup.</td>
<td>• Enough disk or tape resources must be available at the time of backup – For each inline copy requested,</td>
</tr>
<tr>
<td>• Simultaneous creation of disk and tape copies for both short-term and long-term recovery</td>
<td>there must be a resource available to accept the stream.</td>
</tr>
<tr>
<td>– Enables quick and easy creation of backup copies to suit different recovery needs.</td>
<td></td>
</tr>
</tbody>
</table>

**VERITAS NetBackup Advantage**

While VERITAS NetBackup software has had InLine Tape Copy in the Vault option since NetBackup 4.5, it is now a standard feature of NetBackup 5.0 software and has been expanded to include both disk and tape media. NetBackup software still supports up to four simultaneous streams and each can have a customer retention period. InLine Copy can also be implemented with the new Disk Staging feature to provide duplication of disk copies, tape copies, or both, providing unmatched flexibility from any backup product on the market today.

**Off-Host Backup**

*Cost to Implement: High*

Off-host backup uses a variety of technologies to logically move the work of performing a backup to another client/host. Therefore, the goal of this feature is to move the impact of a backup off of the host to another device or server. This backup method requires advanced technology and special equipment, so it is not trivial to set up; however, the benefits are large for those shops that cannot tolerate any backup impact to their host systems or LAN network.

Although there are several methods to perform off-host backups, a basic method is shown in Figure 6. In this example, a client’s SAN-attached disk is mirrored to a third volume. Then the mirror is split and logically imported to a backup server, which then backs up the volume as though the volume was normally hosted on the backup server. After the backup is finished, the volume is resynced with the mirrors on the disk array in preparation for the next backup.

Other methods for performing off-host backups include: serverless (uses a network device that supports SCSI-3 Extended Copy vs. a backup server to move data); replication (uses a replicated volume to the backup server vs. a mirror); or SCSI based (uses shared SCSI mirrored disk with the backup server vs. SAN attached disk). All of the methods share one common need: the volumes must be snapped (frozen) and possibly mapped so that the backup application can safely back up and restore the data.
Advantages to Off-Host Backup

- **Negligible impact to host during backup** – The host client has only a small impact when snapping the volume. After that, the rest of the work occurs on the backup server or network device.
- **Negligible impact to LAN network during backup** – Off-host backup methods do not use the LAN to transfer the primary data, keeping the LAN free for user traffic.
- **Fast backup** – Since the backup of the third mirror occurs on a dedicated host (server or device), the data transfer throughput can be significantly faster than from the client.

Challenges to Off-Host Backup

- **Expensive to setup due to hardware cost** – Typical installations of off-host backup usually require SAN network and SAN attached disk. However, prices for these are dropping quickly.
- **Several “moving parts”** – Off-host backups may have to rely on several applications, arrays, or devices to accomplish the operation. This adds risk to the backup operation. However, with proper planning and testing, these risks can be minimized.

---

**VERITAS NetBackup Advantage**

VERITAS NetBackup software has been blazing the trail for off-host backup for several years now, and with the latest release it is now easier to use. Off-host backups, along with the former Array Integration Option (which is used to manipulate hardware arrays), are now part of the new Advanced Client. In addition to the packaging change, this feature adds integrated GUI support and intelligence to make configuration and maintenance easier.
**Instant Recovery (Virtual Backup)**

**Cost to Implement: High**

Instant recovery, or virtual backup, is a blend of volume-based snapshot technology and backup/restore technology.

![Figure 7. Instant Recovery Method](image)

Snapshot technology is a method of capturing changes to an online volume on a block level. A snapshot gets its name by being able to quickly “take a picture,” or freeze, an online volume. After the snap, changes to the volume still occur to the disk, but the old data is copied to a new “volume” so that it can still be accessed (like a normal disk volume in some implementations). This implementation is commonly called a copy-on-write snapshot. Therefore, the snapshot volume is simply a collection of changed blocks from the original volume. Snapshots can be fairly “light” in storage utilization depending on the data changes and thus can be used to “back up” changes made to a disk at a small resource cost. There are many hardware, software, and operating system vendors that offer this technology, and that differ by implementation and features. The main disadvantages for using snapshots are: 1) if the original data becomes corrupt, all snapshots that refer to the corrupt volume are also corrupt, and 2) the snapshot data is not a long-term backup solution – and data must be copied somewhere else for disaster recovery.

Snapshots by themselves are very useful for short-term data protection, but they should be coordinated with other data protection efforts to provide the best combination of protection and efficiency. Therefore, it makes sense to treat snapshots as another enabling technology for a central backup and restore application to schedule and catalog. Marrying these technologies greatly raises the benefit of snapshots as the table below outlines.
Advantages to Instant Recovery

- **VERY fast and flexible restores** – Backup applications can use snapshots to provide the following types of restores:
  - *Block level restore* – restore only the blocks that have changed for a particular file or the whole volume.
  - *File promotion* – instantaneously restore a file from one of the snapshots back to the original volume. This provides a very fast, granular recovery of a file whether it was deleted or an old version.
  - *Image rollback* – instantaneously restore the entire volume back to a previous state by designating which snapshot to rollback to.
- **VERY fast and efficient initial backups (snapshots)** – Snapshots generally take only seconds to perform and typically do not impact the host using the volume.
- **Integration with long term backups** – after the snapshot is accomplished, block level backups of just the changes within the snap or a traditional backup of the data on the snapshot by a backup application can be performed, making snapshots part of a long-term recovery scheme.
- **Centralized data protection catalog** - After the snapshot is competed, the data can be cataloged to provide seamless integration into the backup application’s normal restore catalog.
- **Integration with applications** – Database applications can be quiesed and snapshots performed to provide the above benefits for applications too.

Challenges to Instant Recovery

- **Data corruption can affect all snapshots** - If the original data becomes corrupt, all snapshots that refer to the corrupt volume are also corrupt.
- **Not a long-term backup solution** – Snapshots are limited by both storage space and the number of snapshots the application can keep.
- **Low portability** – Snapshots are usually bound to the array/system they are created on – they typically can’t be moved to another array.
- **Reading the snapped volume impacts the performance of the original volume** – Since the snapped “volume” contains only the differences between the original data, reading it (performing a backup) ultimately impacts the performance on the original volume.
**VERITAS NetBackup Advantage**

Instant Recovery functionality was initially introduced with VERITAS NetBackup 4.5 software. In NetBackup 5.0 software, the Instant Recovery feature is now part of the Advanced Client Option as part of a greater set of Advanced Backup techniques. NetBackup software’s Instant Recovery offers the flexibility to leverage snap technology from several hardware, software, and OS vendors, as well as the VERITAS Volume Manager and File System products. From one policy, NetBackup software users can specify how many snapshots to keep, the snapshot source, and what to do with the data after the snapshot if finished (backup to tape, disk, or leave as a snapshot). When VERITAS Volume Manager and File Systems are used, NetBackup software users can leverage an advanced set of Instant Recovery features, such as: Block Level Incremental backup/restore, File Promotion, Volume Replicated Snapshots (using VVR), and scheduling third mirror creation and retention. Oracle Database backups are also supported by Instant Recovery. Instant Recovery is supported on selected Solaris, HP, and Windows platforms.

**Problem Scenarios**

The following scenarios serve as examples for how the technologies discussed in this document can be used to solve common backup and restore needs. The solutions mentioned here are suggestions and do not comprise an exhaustive list.

The table below summarizes these scenarios and the suggested technologies used to address them. Additional details are included in the sections below.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Backup to Disk</th>
<th>Image Backup</th>
<th>Disk Staging</th>
<th>Synthetic Backups</th>
<th>Off-Host Backup</th>
<th>Instant Recovery</th>
<th>Inline Copy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow Tape</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>High Frequency Backups</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Slow Client</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Small Window</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Fast Recoveries</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Quick DR</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Multiple Copies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

**Improving Backup/Restore Performance Scenarios**

**Scenario 1: Tape Drives are Too Slow**
If tape-drive bandwidth is a throughput bottleneck, consider using these technologies to improve backup/restore performance:

- **Disk staging** – Using disk staging will improve backup speeds by backing up to faster disk during the backup window and then backing up to tape when speed or time is not so critical.
• **Synthetic backups** – Using this technology may reduce the need for regular full backups, thus requiring the tape drive to back up less data.

• **Instant recovery/virtual backup** – Using this technology could serve as a method for performing short-term incrementals on disk (snaps). Furthermore, this may reduce the amount of data needed to back up since the snaps can be backed up as more efficient block-level incrementals.

• **Backup to disk** – If backup data is not needed for long-term storage and disk is plentiful, simply backing up the data to disk might be a good idea.

### Scenario 2: Need High Frequency Backups (Near-Continuous Backup)
If the data being protected requires high frequency backups to ensure a small window of vulnerability, consider using these technologies:

• **Instant recovery/virtual backup** – Using this technology could serve as a method for performing short-term incrementals on disk (snaps). These snaps can be backed up to disk or tape as a secondary operation when time isn’t so critical.

• **Synthetic backups** – This technology is effective for summarizing several incremental backups into a synthetic full or differential.

• **Incremental image backup** – If the file system being protected has many small files or a few very large files in which only a small percentage of the large files change, image backups with followup incremental image backups may do the trick.

### Scenario 3: Slow Backup from Client (due to file structure, slow disk or network, load on server, etc)
If the throughput from the source client is slow (due to a complex file structure, slow disk I/O, heavy resource load on the server, etc.), consider using these technologies to speed up the backup:

• **Disk staging** – Instead of using tape multiplexing, disk staging can be used to cache a slow data stream to disk before moving the data to a high-speed tape drive.

• **Off-host backups** – This technology is well suited for this scenario since it moves the impact/load of a backup to a backup server or device.

• **Synthetic backups** – Reducing the amount of data needed from the client may help solve the problem. In this case, use synthetics to summarize the incremental backups into a synthetic full or differential.

• **Instant recovery/virtual backup** – This technology has a light impact on the client and could serve as a method for performing short-term incrementals on disk (snaps). These snaps can be backed up to disk or tape as a secondary operation when time isn’t so critical.

### Scenario 4: Backup Window Too Small
If the window of time allotted to back up data is too small, consider using the following technologies to reduce the amount of data or increase backup throughput:

• **Synthetic backups** – Reducing the amount of data needed from the client may help solve the problem. In this case, use synthetics to summarize the incremental backups into a synthetic full or differential.
Using Disk to Solve Backup and Restore Problems

- **Off-host backups** – This technology can completely move the impact of data backup to another server or device to increase backup performance.

- **Instant recovery/virtual backup** – Using this technology could serve as a method of performing short-term incrementals on disk (e.g., snaps). Furthermore, this may reduce the amount of data needed to back up since the snaps can be backed up as more-efficient block level incrementals.

- **Disk staging** – Assuming slower tape drives are used, disk staging can improve backup speeds by backing up to faster disk during the backup window and then backing up to tape when speed or time is not so critical.

- **Backup to disk** – If backup data is not needed for long-term storage and disk is plentiful, simply backing up the data to disk might be a good idea.

- **Image backup** – If the file system being protected has many small files that are dragging down the speed of a traditional backup, image backup (along with Incremental Image) could greatly speed up the backup process.

**Scenario 5: Need to Recover Data Very Fast**

If recovery time must be very quick, consider using these technologies:

- **Instant recovery/virtual backup** – Instant recovery can restore file(s) or an entire volume within seconds from a snapped volume.

- **Backup to disk** – Restores from data that has been backed up to disk can occur very quickly since the media does not have to be mounted (as with tape based restores).

- **Disk staging** – If the data being restored is still staged to disk (has not been deleted after being copied to tape), restores can occur very quickly similar to backup-to-disk.

**Disaster Recovery Scenarios**

**Scenario 6: Need Quick Disaster Recovery**

For instances where time is essential to bring back a full volume or backup image, consider using these technologies:

- **Instant recovery/virtual backup** – Instant recovery can restore an entire volume within seconds from a snapped volume. Be careful to choose a snapping technology that supports persistent snaps (that come back after a reboot).

- **Off-host backups** – Off-host backups using an alternate client configuration backing up a third mirror can be configured to not resynchronize the mirror back to the primary host after the backup is finished. If a full disk recovery is needed after the split has occurred, the third mirror can then be used to recover the volume by then performing the resync.

- **Synthetic backups** – During a disaster recovery, it is far quicker to restore one full backup image vs. a full and several incremental backups. Synthetic backup can be used to greatly reduce the amount of backup images needed for a restore by summarizing the incremental backups into a synthetic full or differential.
• **Image backup** – Recovering a full backup of a volume that has many files on it can be very time consuming if a traditional file-by-file method is used. If the file was backed up with image backup, the entire volume can be quickly restored via image technology.

**Scenario 7: Need Multiple Disk and Tape Copies for Onsite and Offsite Storage**
Disasters can come in all sizes; therefore it can be equally important to plan for events where a tape is lost or an entire site is lost. Having the right type of backup media placed onsite and offsite in an electronic vault is very important to ensuring a DR occurs without failure.

• **InLine Copy** – This technology can efficiently create several copies of data on multiple types of media simultaneously, enabling the admin to easily position data for fast restores (backups on disk), offline restores (tape in a nearby library), and total site disasters (tape in offsite electronic vault).

• **Disk staging** – Staging data to disk for quick recovery could help compliance with SLAs, while being a part of an electronic vaulting scheme to create multiple tape copies for offline storage. If just one tape copy is needed, then disk staging by itself may be enough, or combine it with InLine Copy technology to create multiple tape or disk copies for maximum flexibility and DR preparation.

**Summary**

While tape is still needed for data protection, using disk for backup and restore has the advantages of being faster, more flexible, more efficient, and more cost-effective. In recent years, several new disk-storage technologies have emerged, allowing administrators to manipulate data on disk in a variety of ways. Additionally, the continuing drop in disk prices has led to affordable tools that make the best out of these new technologies, and VERITAS NetBackup™ software is one such tool. NetBackup software can be used for each of the disk-enhanced data protection methods described in this paper, and it is an effective and comprehensive solution for solving backup and restore problems.
VERITAS Software Corporation

Corporate Headquarters
350 Ellis Street
Mountain View, CA 94043
650-527-8000 or 866-837-4827

For additional information about VERITAS Software, its products, VERITAS Architect Network, or the location of an office near you, please call our corporate headquarters or visit our Web site at www.veritas.com.