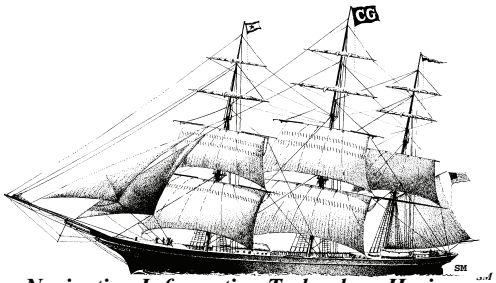
Clipper Notes™
VENDOR-AGNOSTIC EXPLANATIONS AND ADVICE FOR THE INFORMATION TECHNOLOGY BUYER



Navigating Information Technology Horizons℠

New in 2007 Report #TCG2007015 February 1, 2007

The Evolution of Backups — Part One – Improving Performance

Analyst: Dianne McAdam

History of Backups

The amount of data that IT organizations need to store grows every year with no end in sight. All of this data must be backed up to ensure that when corruption or accidental deletion occurs, the data can be quickly restored. As the amount of application data grows, the size of the backups that every IT organization must manage continues to grow.

While the amount of data that is backed up continues to grow, the time to complete the backups continues to shrink. Some of us remember the days when servers were brought down every night and backups were run before the beginning of the next online day. There was plenty of time to run backups. Servers stood idle in the early morning hours before the daily applications were brought online. Sunday afternoons were another period of idle time; servers stood idle unless that time was required to do maintenance on the operating systems. But that all changed when local companies started to reach outside of their geographic boundaries. Local companies that supported employees and customers in one time zone later acquired other companies several time zones away. Online applications that were once idle after 6 p.m. had to remain available to 9 p.m. Administrators started to feel the effects of shrinking backup windows, and some backups had to be cancelled when they could not complete it the time allowed.

The global economy has dramatically changed the requirements for application availability. While customers are sleeping in the United States, customers in other parts of the world are requiring applications be up and available for their use. The result – many administrators are faced with the task of completing backups during very limited, or even non-existent, backup periods. These shrinking backup windows coupled with high levels of data growth require administrators to find faster ways to complete backups.

Backup software vendors have not ignored this demand for faster backups and have made many changes to their software to improve performance. Disk and tape vendors have also continued to work on solutions that can speed up the backup process. In fact, we have come a long way from the days when applications were brought down every night and running full backups were the normal mode of operation. Read on to find out how.

IN THIS ISSUE

The Evolution of Backups

Full Backups

Originally, all backups were what we now refer to today as *full backups* – that is, every byte of data is backed up every single time on tape. In the days when online applications were brought down at 6 p.m., full backups easily were completed on time. But, as the amount of data grew larger and the amount of time for the backup window got smaller, full backups could not be completed in the time allowed. Many enterprises today combine full backups with partial backups to meet their schedules. Other enterprises continue to use nightly full backups for specific applications such as Microsoft *Exchange* or *SQL*.

Advantages of Full Backups

- Since every byte of data is backed up, full restores are quicker than other "partial" methods of backing up data.

Disadvantages of Full Backups

- A full backup requires the same amount of storage as the primary data. If full backups are taken every night then, by the end of the first week, one TB of data requires 7 TBs of backup space.

Incremental Backups

Incremental backups only back up the data that has changed since the last full or incremental backup. Incremental backups reduce the large storage requirements required by full backups but at a cost in restore time. For example, many enterprises will take a full backup on Sunday and incremental backups the rest of the week. Suppose a full backup is taken on Sunday night. On Monday night, the incremental backup records all data that has changed since Sunday night. Tuesday's incremental backup records all data that has changed since Monday night. This process continues until the next full backup is taken. To restore application data on Thursday, four restores have to be performed – first the full (Sunday) backup is applied and then each incremental (Monday, Tuesday and Wednesday) backup must be applied after that.

Advantages of Incremental Backups

- Incremental backups run quickly since only the changed data is backed up.
- Storage requirements are reduced significantly. If 10% of one TB of primary data is changed daily, then the weekly backup storage requirement is only 1.6 TBs. (1.0 TB +.1 +.1 +.1 +.1+.1 +.1 = 1.6 TBs).

Disadvantages of Incremental Backups

- The last full backup and all intervening incrementals must be applied to restore data, resulting in much longer restore times.

Differential Backups

Differential backups provide some of the advantages and disadvantages of full and incremental backups. Differential backups record changed data since the last full backup. On Sunday night, a full backup is completed. Monday only the data that has changed is backed up. Tuesday, the changed data from Tuesday and Monday are backed up. Wednesday's differential backup includes all the changed data since Sunday night.

Advantages of Differential Backups

- Differential backups run faster than full backups, because they do not have to do comparisons to the interim partial backups.
- Differential backups require less storage than full backups, because they replace the interim backups with the latest backup.
- Restores run quicker than incremental backups, since only the last full backup and the last differential backup are applied.

Disadvantages of Differential Backups

- Differential backups require more storage than incrementals. If 10% of one TB of primary data is changed daily, then the weekly backup storage requirement is 3.1 TBs. (1.0 TB + .1 + .2 + .3 + .4 +.5 +.6 = 3.1 TBs)
- Over time, differential backups will take longer to run than incremental backups.
- Restores take longer than full-backup restores; the differential backup must be applied after restoring the last full backup.

Synthetic Backups

Synthetic backups combine the quick restore times of full backups with the quick backup times of incremental backups. Synthetic backups are supported by many backup application vendors. Enterprises need only create the first full backup. Every backup after that is an incremental backup (sometimes called "incrementals forever"). The backup application then merges the first full backup with the next incremental backups to create a new full 'synthetic' backup. Periodically, the backup application creates a new synthetic backup from the previous synthetic backup and subsequent incremental backups. Why do we call this a *synthetic backup*? Synthetic backups are not created from backing up original data but are created by merging backup files.

Advantages of Synthetic Backups

- Synthetic backups initially run one full backup, then incrementals forever. Incremental backups run quickly since only the changed data is backed up.
- Synthetic full restores run quicker than other partial methods of backing up data.

Disadvantages of Synthetic Backups

- Each synthetic full backup requires the same amount of storage as the primary data.

Disk-based Backups

Incremental and differential backups provide needed benefits over traditional full backups by reducing backup storage requirements and reducing backup times. The emergence of lower cost SATA disk drives makes these disk systems financially attractive as backup targets. Full, incremental, or differential backups can be directed to disk rather than tape.

Some believe that replacing tape with disk speeds up the backup process since tape is slower than disk. Tape drives run most efficiently *when* they receive and write data in a continuous stream, but backup applications cannot always push data at a constant rate. The tape drive is forced to start, stop, and reposition itself for the subsequent block to eliminate wasted tape. This extra repositioning step – commonly referred to as a *back-hitch* – consumes valuable time and causes undue wear as the tape is moved back and forth under the heads. This stop-rewind-start process is also known as the *shoeshine effect*. In the case of shoes, the result is a nice, polished look; in the case of tape, it means wasted time and increased tape wear as the tape moves back and forth under the heads.

This particular problem has been around since the days of the first tape drives and, while it still exists today, vendors have learned how to compensate for it.

- One way is to add a large tape buffer, which essentially stores data so the tape drive can write the data at a constant rate to the media. This allows the tape drive to keep moving in a forward, not back-and-forth, motion.
- Virtual back-hitch eliminates back-hitching by writing data to a temporary part of the media, then writing it to its permanent location – again, keeping the tape moving in a forward motion.
- Another option, developed by backup software vendors, is to "interleave" or multiplex blocks from more than one volume at a time. The idea is to keep tape drives moving by sending blocks of data to it from more than one backup client at a time. While this type of technique improves backup performance (since data is continually streamed to the tape drives), it can actually slow down the restore process, since the data has to be "re-assembled" a block at a time from multiple backup jobs stored on the tape.

The bottom line - writing backups to disk can be faster than writing to tapes if the tape drives can't be kept streaming and restoring from disks is faster when backups have been "interleaved". Writing backups to disk can eliminate the need to handle tape cartridges manually – a boon for remote offices that do not have IT-savvy personnel on site. Let's examine how disk can be added to backup environments.

Disk as Disk

One way to introduce disk into backup environments consists of using a disk array as the target of the backup software instead of a tape drive. This method is supported by all major backup applications. Backed up data is written in disk, not tape, format. There are two common approaches – using existing disk arrays as the target and using specially-built appliances.

Conventional Disk Arrays

Secondary disk, usually with higher-capacity SATA disk drives, serves as the backup target for devices, and does not require specially-built appliances to be purchased. This can be a low-cost solution for customers that have extra disk capacity on hand. All of the major backup application vendors support writing backups to conventional disk arrays and publish manuals that step administrators through the changes required in the backup procedure.

Disk systems must be partitioned to support different operating systems and capacity cannot be shared across partitions. If the backup requirements for one operating system increase, then the disk system may need to be partitioned again. Conventional disk arrays have interfaces that allow administrators to configure disks, but these interfaces do not monitor backup activity. Customers need to monitor disk systems to determine if they have sufficient capacity for tonight's backup. They must also determine if the disks have been fragmented and need to be "defragmented" to improve performance.

Advantages of Conventional Disk Array Backups

- Disk-based backups can improve backup and restore times over more traditional tape-based processes.
- Customers can "re-use" extra disk capacity on site, and not have to buy additional hardware

Disadvantages of Conventional Disk Array Backups

- SATA disk, while less expensive than Fibre Channel disk drives, continues to cost more than tape.
- Implementing conventional disks as backup targets requires changes to the backup process, since the backup application is now writing to disk rather than tape.
- Disk arrays must be monitored for capacity and performance.

Special-Purpose Appliances

Some vendors have developed specialized software that ships with disk arrays to make these arrays easier to use as backup targets. Some of these solutions look like NAS filers and support *CIFS* and *NFS*; others look like Fibre Channel disk. Many of these appliances support additional options, such as compression or filtering (to eliminate redundant copies).

Appliances have many features not available with conventional disk, such as interfaces that monitor backups, and space-saving features, such as compression and filtering. However, these appliances generally come with a higher price tag than conventional disk systems.

Advantages of Specially-built Appliances

- Appliances have built in monitoring facilities to let administrators know when capacity is reaching a critical threshold.
- Administrators can easily monitor the activities of current backup jobs.
- Many appliances can add capacity dynamically without partitioning disk system.
- Many appliances have additional features, such as compression, which can save disk space and cost.

Disadvantages of Specially-built Appliances

- Like conventional disks, changes must be made to the backup process.
- Appliances have many features not available with conventional disks, but usually come with a higher price tag.

Disk as Tape

With virtual tape implementations, disk arrays respond to software commands just as if they were a tape drive or tape library, hence the name virtual tape. Usually, the disk array and software can be plugged easily into existing backup process. The normal processing of directing the backup stream to tape is simply redirected to tape. Since the disk arrays emulate tape drives or libraries, many maintenance tasks, such as setting up partitions and monitoring disk fragmentation, are not required.

Virtual tape libraries (VTLs) have gained market acceptance since they improve backup and restore times without requiring enterprises to alter dramatically their backup processes. Enterprises can install virtual tape libraries with little, if any, changes to their backup applications.

Virtual tape libraries can improve restore times, but many do little to reduce storage requirements. Some VTLs always compress backup data; some VTLs offer compression as an option. Compression rates can vary, but many enterprises see compression rates of about 1.5 or 2 to 1, or more. Since the compression algorithms in use by many virtual tape systems are similar to those used for years on tape drives, there is no noticeable reduction in storage requirements when writing to VTL rather than physical tape. However new data reduction technologies have recently been introduced by several vendors and the reduction factors average around 20 to 1 (as always, this is data dependent). Data reduction offers significant savings in the capacity required to store backups and will be covered in the second part of this paper.

Advantages of Virtual Tape Libraries

- VTLs are easy to implement and require few, if any, changes to the backup application.
- VTLs can improve backup and restore performance.
- Some VTLs have hardware compression capabilities that don't affect performance but save disk space.

Disadvantages of Virtual Tape Libraries

- VTLs vary greatly in price, but do add additional cost.
- Some VTLs have software compression methods that affect performance.

Combining Other Techniques with Backups

The need for long backup windows can be reduced significantly by combining other data protection techniques with more traditional backup processes. For example, snapshots can be created periodically during the day. These disk-resident snapshots can then be mounted on separate servers and used as the primary data source for backups. This off-host backup allows backups to be created without affecting the performance of online systems. Other data protection techniques, such as continuous data protection, or CDP, continuously capture updates on a separate storage device. The CDP engine can create a view of the data from any previous

point-in-time that can be the source for the backup process.[1]

Some backup vendors support encryption, which allows backups to be securely stored on tape or disk. Some also support encrypting data while it is being transmitted to a remote site.

Advantages of Combining Other Techniques

- Technologies, such as snapshots and CDP, can significantly reduce or eliminate the need for long backup windows.
- Some vendors have integrated snapshots with backups and have automated the process.
- Encryption provides additional security.

Disadvantages of Combining Other Techniques

- CDP and snapshot software requires additional licensing fees.
- Additional storage is required to store the snapshot or CDP data.
- Encrypted data does not compress well. If compression of data is needed, data must first be compressed, then encrypted.

Don't Forget about Disaster Recovery

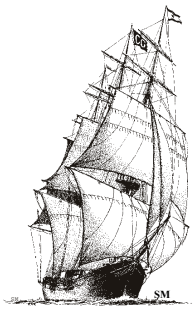
One important note! Disk-based backup solutions can be very effective in reducing the backup window and speeding up the restore time. But disk-based solutions are not disaster recovery solutions, unless the volumes are replicated over distance from storage system to storage system. Many disk backup products will reside in the same SAN as primary disk, rendering them ineffective for disaster recovery. A data center outage, such as a power failure, would make both the primary disk and the backup disk inaccessible.

Disaster recovery best practices dictate that a copy of data be stored at an off-site location. Copies of disk resident backups can be written to tape, which can be shipped to secure locations to satisfy disaster recovery requirements. Some disk backup solutions have support to create tapes for offsite storage. Other solutions do not have tape support and require additional steps to create physical tape backup copies. Some solutions can replicate data from one disk-based backup system to a second system. Each solution must be evaluated carefully to determine which processes, if any, need to be changed to satisfy the requirement of off-site storage.

Conclusion

Backup technology has evolved since the days when IT had to back up every byte of data to tape every night. Backup software vendors have added support for differential and incremental backups to improve backup times. Hardware vendors have combined lower-cost SATA disk with tape emulation to produce virtual tape libraries that improve backup and restore performance.

These improvements have eased the life of the backup administrator, but have not been able to reduce storage requirements dramatically. Data reduction software helps solve that problem. Several vendors are now delivering products (and others are promising products) that can reduce storage requirements dramatically, as well, for backup. We will cover that topic in part two of this report, also.[2]



[1] See **The Clipper Group Explorer** dated May 9, 2006, entitled *Continuous Data Protection – Now Ready for Prime Time*, available at http://www.clipper.com/research/TCG2006039.pdf

[2] See **The Clipper Group Explorer** dated August 15, 2006, entitled *The Evolution of Backups – Part Two - Improving Capacity*, available at http://www.clipper.com/research/TCG2006071.pdf.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Dianne McAdam **is Director of Enterprise Information Assurance for the Clipper Group**. She brings over three decades of experience as a data center director, educator, technical programmer, systems engineer, and manager for industry-leading vendors. Dianne has held the position of senior analyst at Data Mobility Group and at Illuminata. Before that, she was a technical presentation specialist at EMC's Executive Briefing Center. At Hitachi Data Systems, she served as performance and capacity planning systems engineer and as a systems engineering manager. She also worked at StorageTek as a virtual tape and disk specialist; at Sun Microsystems, as an enterprise storage specialist; and at several large corporations as technical services directors. Dianne earned a Bachelor's and Master's degree in mathematics from Hofstra University in New York.

- ***Reach Dianne McAdam via e-mail at dianne.mcadam@clipper.com or at 781-235-0085 Ext. 212. (Please dial "212" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group *Captain's Log***, **The Clipper Group Voyager**, **Clipper Notes**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.