

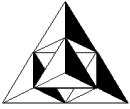
Understanding Data Backup Best Practices



July 2, 2001

Commissioned by the DLTtape Media Platform Council





META Group

208 Harbor Drive, PO Box 120061, Stamford, CT 06912 (203) 973-6700 Fax: (203) 359-8066

INTRODUCTION3

BUSINESS IMPERATIVES DRIVING DATA BACKUP3

BALANCING COST AND RISKS FOR BEST PRACTICES3

THREE CATEGORIES OF DATA5

POINTS OF RISK.....6

DISASTERS ARE NOT THE MAJOR SOURCE OF DATA LOSS6

LACK OF ENTERPRISE-WIDE BACKUP POLICIES7

DATA BACKUP BEST PRACTICES7

CATEGORY 1 BEST PRACTICES.....7

Backup Retention Strategy.....7

Media Rotation Strategies – Interleaving and Multi-Streaming.....9

Duplication vs. Cloning.....9

Vault Rotation and Archiving10

Tape Recycling11

VARIANCES IN BEST PRACTICES FOR CATEGORY 2 DATA11

VARIANCES IN BEST PRACTICES FOR CATEGORY 3 DATA12

BEST PRACTICES FOR VAULTING.....12

OPERATIONAL BEST PRACTICES13

OTHER SOFTWARE SELECTION CRITERIA.....14

ABOUT THE AUTHORS AND METHODOLOGY.....15

APPENDIX I – BACKUP ROTATION SCHEMES16

 GRANDFATHER-FATHER-SON16

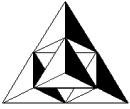
 TOWER OF HANOI.....17

APPENDIX II.....18

 GLOSSARY.....18

APPENDIX III.....20

 BACKUP BEST PRACTICES SUMMARY CHART.....20



INTRODUCTION

Business Imperatives Driving Data Backup

Data is everything to the Information Economy. Data is the representation of all activities within global enterprises, and the amount generated everyday, by software applications, multi-media, networks, servers, and mission-critical business systems such as financials, human resources, manufacturing, and marketing, is increasing exponentially.

The issue of data backup has gained urgency due to current growth in information technology environments and their corresponding usage. Despite the fact that business operations, and hence IT systems, are more distributed, decision timeframes for business have shrunk dramatically because of competitive pressures and heightened customer expectations. Data – uncorrupted, accurate and centralized – has to be available on demand, and be preserved for the next round of critical decisions that impact all levels of a business. However, given increased availability requirements, businesses are forced to perform backup in shorter time slots.

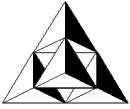
META Group evidence suggests that the costs associated with downtime, which further heightens the need for swift restoration, ranges anywhere from \$80,000 to \$2 million per hour. Implementing best practices and / or infrastructure to increase availability from 99.9% (9 hours of annual downtime) to 99.99% (52 minutes of annual downtime) adds between \$64,000 and \$16 million to top line revenue.

Many businesses today, especially those that are information-intensive, have begun to measure data access interruptions in concrete dollar figures (e.g. amount of dollar lost per unit time), demonstrating a heightened sense of awareness towards the critical need for data backup. Furthermore, given rapid business cycles and changes, these businesses are seeking data backup practices that must be highly efficient and repeatable to avoid the loss of large amounts of data.

However, at the end of the day, data is collected by fallible systems, which means that business operations run the risk of being severely, or even lethally, impacted if a company does not take the necessary steps to protect and back up its information assets securely and efficiently.

Balancing Cost and Risks for Best Practices

Operating systems, database systems, system software, application software, and individual files all carry important business information, but companies have limited manpower, time and financial

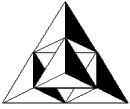


resources. In order for a company to maximize its data backup efforts, the first action is to determine data priority. It is important to recognize that not all data is at the same level of risk, or as valuable to the organization, as other data. For example, losing an individual employee's e-mail will certainly not have as significant an impact on the business as losing data from the company's core financial tracking system.

To establish data backup best practices while minimizing costs, enterprises should look towards implementing the *right* level of data protection, and not simply aim for the *best* level of data protection for anything and everything. Furthermore, the technical process of data backup must be coordinated with business processes. For example, an online trading company knows that daily backups need to be completed prior to the return of trading hours. To correctly establish data priorities and ensure compliance with business procedures, companies need to create comprehensive operational plans. An overarching operations strategy would contain guidelines for activities such as:

- Evaluation of application priorities to the business – to determine which sources of data would require higher redundancy;
- Analysis of business impact – to measure potential effects on the business in the event of data loss / access interruption;
- Synchronization of business processes – to coordinate data backup and archival tasks with operations milestones. For example, backing up data upon the acquisition of every 100 orders;
- Scheduling of backups – to enforce compliance with legal, financial, or regulatory requirements;
- Timely procurement of backup media – to ensure tape and tape drive demands driven by frequent backups are met;
- Periodic audits of data volume growth – to anticipate data growth and risks as the business scales;
- Centralized control of distributed data storage networks – to maintain consistency and integrity of data archives;
- Mandates for employee desktop backups – to protect knowledge / information assets on the individual level which is particularly important in the mobile and telecommuting world, where the end computing system can take on many forms such as a laptop, personal digital assistant, mobile phone, etc.;
- Establishment of communication channels / procedures – to organize communications in the event of data loss, and to provide for escalation procedures (to executive management or qualified third party services) if data loss widens.

Although backup and recovery (B/R) operations often focus on backup alone, best practices focus on two perspectives:



META Group

208 Harbor Drive, PO Box 120061, Stamford, CT 06912 (203) 973-6700 Fax: (203) 359-8066

- The protection of data from loss; and
- The swift restoration of data in the event of loss.

(For the purposes of this paper, we will not be discussing best practices for disaster recovery, i.e. data loss from unplanned / unforeseen business interruptions, caused by factors beyond human control.)

Three Categories of Data

To support data backup best practices, companies need to conduct the foundational work of categorizing their data assets according to the priority and value for the organization. Based on its research, META Group proposes the following segmentation:

- **Category 1 – Mission-Critical Data**
 - This class of data is mission-critical to the ongoing operation of the business. Without this data, some or all business operations can come to a halt. This category requires the highest level of redundant protection and the fastest possible restoration.
 - Some examples of mission-critical data include online transaction processing (OLTP) that records business transaction activities, and enterprise resource planning (ERP) systems that comprise the majority of the companies' critical operations such as financials, manufacturing, inventory, etc.
- **Category 2 – Business-Critical Data**
 - Category 2 data is business-critical to the organization. This data must be protected, but does not halt ongoing business operations when lost. Business-critical data does not justify the highest levels of redundancy, and can tolerate longer restoration timeframes.
 - Some examples of business-critical data include those found in data warehouse environments, or e-mail servers.
- **Category 3 – Employee-Level Data**
 - Category 3 data is unique to an individual employee, and is usually found on a desktop / laptop device. This class of data is usually considered the responsibility of the employee to protect, even though the data is corporate property.
 - Examples include employee e-mail, spreadsheets, documents, etc.
 - This category of data is often further sub-divided by employee type (i.e. critical research data created by pharmaceutical scientists is more highly categorized than internal IT help desk logs).

This categorization will also form the basis of our best practices discussion.



POINTS OF RISK

Disasters Are Not the Major Source of Data Loss

Industry research indicates only 3% of data loss is the result of natural disasters. Other more probable causes include human error (e.g. accidental deletions), hardware failure, software corruption, theft and viruses.

All three categories of data above are vulnerable to loss. While many companies may be tempted to think that data loss is frequently beyond their control (hence the great deal of attention paid to the issue of disaster recovery), industry research indicates only 3% of data loss is a result of natural disasters¹. Overwhelmingly, more probable causes include human error (e.g. accidental deletions), hardware failure, software corruption, theft and viruses. A 1996 study by ONTRACK Data International indicates that data loss could typically be attributed to hardware error (44%), human error (32%), software error (14%), and virus (7%).

In addition to outright data loss, data can be rendered unrecoverable for several other reasons, including:

- Loss of storage media – If IT operations are not centralized and data is dispersed among different locations, there is an increased probability of misplacing, or even simply mis-marking data cartridges;
- Media corruption or broken media – Due to poor handling, some media can be corrupted or broken, and the data therein becomes unrecoverable;
- Media usage and shelf life – Tape/media manufacturers typically make recommendations regarding appropriate usage (e.g. the maximum number of times the tape could be read / written upon) and shelf life to maximize effective tape longevity and to ensure data integrity. Failure to adhere to these recommendations could result in data loss.
- Loss of compatible devices – As the market rolls out new versions of hardware and/or software, and older versions become moot, some vaulted media can no longer be restored. Incompatible hardware and software can include tape drives, backup / recovery software, operating system versions, and application software packages.
- Corruption of the data – Viruses often delete data and undermine data integrity. A danger is that many backup routines often spread the impact of a virus by their very nature.

It should be apparent that the majority of data loss can be prevented if there are long-term operational and technology plans in place to anticipate, control, and coordinate data backup management needs.

¹ The Data Recovery Solution, ONTRACK Data Recovery, Inc. 1998



As such, businesses must take care to choose procedures and technologies that not only plan for major catastrophes, but also for the above-mentioned mishaps.

Lack of Enterprise-Wide Backup Policies

An enterprise-wide data protection plan is easier-said than done, and META Group research shows that most organizations have yet to apply corporate-wide policies (for example, only 5% of category 3 data is protected by organizational procedures). Besides the lack of knowledge about data backup best practices, there are broader challenges. On a business level, scattered and diverse operations / divisions frequently thwart the establishment of centralized control mechanisms. On a technology level, many organizations' IT infrastructures are still being built on heterogeneous platforms that dictate the implementation of different B/R software. Lodged between the business and technical issues are cultural challenges, such as language barriers and politics, that exacerbate the pain of creating an enterprise-wide data protection plan. In yet another unique twist, country-specific privacy laws also potentially impact the movement of customer data across borders, limiting data protection plans. In turn, these discrepancies hinder execution of a consistent backup policy.

DATA BACKUP BEST PRACTICES

META Group believes that if comprehensive data backup best practices are put in place, and applied to the data environment, including hardware, software, operating systems, applications, and the associated business processes, businesses should be able to significantly reduce the likelihood of costly – and potentially crippling – data loss. As mentioned earlier, different categories of data justify different levels of backup investment and effort. The three categories of data above will frame the rest of our best practices discussion.

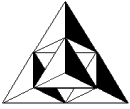
Category 1 Best Practices

Category 1 represents mission-critical data, and hence must be restorable to a current state for core business continuity, or to a point-in-time for audit purposes. In fact, some regulatory requirements may mandate specific retention schedules other than the general outline below.

Backup Retention Strategy

For this type of data, the best practice backup strategy can be segmented according to specific retention periods:

- **Weekly Full Backup** – Weekly backups should be performed on operating system data, B/R software data, content data, configuration files, registry files and application software data. Weekly data should be retained through the end of the month.



META Group

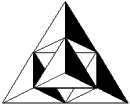
208 Harbor Drive, PO Box 120061, Stamford, CT 06912 (203) 973-6700 Fax: (203) 359-8066

- **Incremental Backup** – Changes made to files since the last full backup should be backed up on a daily basis. However, META Group does not recommend delta incremental backup (including only those files that have changed since the last incremental) for category 1 data, since delta incrementals take longer to restore and are riskier if there is media failure. From a data protection perspective, if an organization follows a comprehensive backup schedule, daily incremental data need only be retained for seven days or through the end of the week. Some backup software solutions default to three weeks for retention of incrementals, and for companies that need specific point-in-time restorations from a business perspective (e.g. for audits), this longer retention period may be necessary. There is no significant downside to three-week retentions other than a larger volume of data to manage.
- **Monthly Full Backup** – Similar to the weekly full backup, this monthly activity needs to include all data content and files. The data should be retained in the media set for 30 days, then vaulted off-site, after which it should be retained for one year, or through the end of the year.
- **Annual Full Backup** – All data and files are vaulted off-site immediately, where they are typically retained for seven years, or for a period specified by the company's regulatory demands. However, best practices dictate that the media should be returned for image refresh after approximately three and a half years, or the respective midpoint of the regulated storage period, to ensure data integrity. Most manufacturers recommend a finite shelf life for their media, which should be followed. But regardless of indicated durations, enterprises must take care to refresh data periodically on even the most robust tapes for maximum assurance.

Needless to say, the backup / recovery software chosen by the enterprise will need to be robust enough to support all of these retention schedules, and be able to give immediate notification should there be any glitch or failure during any one of the cycles. It should also enable backup administrators to specify custom retention schedules if necessary. Given this rigorous schedule, software that enables auto-loading and automated rotations, minimizing direct human intervention, is highly desirable.

No matter how rigorous the backup schedule, and how well-preserved data is, enterprises must still take one more step to ensure backup data recovery. Upon instances of significant hardware upgrade, the IT team must retain sufficient legacy hardware components, such as tape drives, interconnects, computers, and backup software, to perform data restoration in the future. Otherwise, even with the best storage media, there may not be the appropriate IT components to access the information.

Upon instances of significant hardware upgrade, the IT team must retain sufficient legacy hardware components and recording software to perform data restoration in the future.



Besides the above-mentioned media rotation schedule, there are other common rotation schemes that are employed by enterprises, such as the Tower of Hanoi. See Appendix I for a complete description of these methods.

Media Rotation Strategies – Interleaving and Multi-Streaming

Businesses, especially those with global operations, generate a vast amount of category 1 data every day. This high level of activity poses a great challenge to data backup operations, as there is only a finite amount of time within which the backup must be completed.

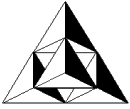
The technique of interleaving was designed to side-step this hurdle. Via the control of backup / recovery software, interleaving allows for multiple streams of data from different sources (tape drives) to be recorded onto the same storage media concurrently, thereby reducing total backup time and resources consumed. Although interleaving saves on backup time, it creates longer restoration timeframes. Different sets of server data are likely to have different rotation schedules. If this data is mixed on one tape, it becomes very difficult to differentiate and separate one set from the other. Also, while it is a more time-efficient method, META Group recommends against employing interleaving for mission-critical data, as mixing data save sets on the same media means that in the case of media failure or loss, multiple sets of data will be put at risk. Interleaving is acceptable for category 2 data, particularly if the data retention schedules are similar for the different save sets. For category 1 data, always use unique media without interleaving.

There exists another technique called multi-streaming, which allows backup software to transfer multiple streams of data to separate storage media at the same time. This method retains the time-efficiency advantage of interleaving, while reducing the risk of losing mission-critical data in one fell swoop. Hence, multi-streaming is the preferred method for Category 1 data. However, users must be aware that multi-streaming is more resource-intensive, requiring higher network capacity and a higher volume of storage media than interleaving.

Depending on a company's specific vertical needs or judgment, either one of the two approaches, or a combination of both, may be used on the different data categories. Companies should ensure that their backup software solution has the flexibility to initiate and manage both methods, and allow for administrative controls, such as the number of threads sent to each device, or the number of clients backed up at once.

Duplication vs. Cloning

Category 1 data can gain additional redundancy via duplication and cloning.



Duplication refers to the process whereby multiple copies of save sets are created online. The process is critical to category 1 data, especially if the backup software packages' archiving function automatically erases data off disks. Best practices recommend that one duplicate copy of the media set be sent to an off-site vault immediately, or at least to separate storage location to avoid failures. Keeping both on-site and off-site copies of the data enables more convenient access and speeds up restoration timeframes. The retention schedule should be as rigorous as the one outlined above for daily, weekly, monthly and annual backups. We recommend that full backups for quarterly or monthly data be duplicated, and consider duplication for annual full backups an absolute necessity. Unique media, without interleaving, should always be used when it concerns category 1 data. Finally, at least two copies of the annual full backup should be placed in separate vault sites for maximum protection.

Full backups for quarterly or monthly data must be duplicated, and consider duplication for annual full backups an absolute necessity.

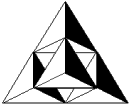
For duplication, the costs are slightly higher as it requires dedicated online resources. However, as it provides for the highest level of data protection to category 1 data, the benefits exceed the cost considerations.

Backup personnel should not use cloning for category 1 data as a substitution for online duplication.

On the other hand, cloning is the process of duplicating tapes offline, usually used when there is not enough tape drives available for online duplication. Since cloning utilizes drive resources on an as-available basis, the costs associated with cloning (especially the opportunity costs of focusing drive resources elsewhere) are not as high. However, in terms of best practices for category 1 data, cloning is not the ideal method because cloning creates a window of time when you only have one copy of the data, while waiting for drives to become available. Backup personnel should not use cloning for category 1 data as a substitution for online duplication, unless duplication is really not feasible.

Vault Rotation and Archiving

After data duplication, data is ready to be moved off-site for long-term storage. Software that helps in marking data for off-site vaulting is advantageous and proves essential to swift restoration. However, off-site vaulting of data is not the last step in backup. For category 1 data, even vaulted media should be refreshed periodically to ensure data is not corrupted, and is accessible. For example, due to regulations and auditing needs, there are some financial institutions that are required to keep their data



META Group

208 Harbor Drive, PO Box 120061, Stamford, CT 06912 (203) 973-6700 Fax: (203) 359-8066

for seven years. In those cases, we recommend that the vaulted data be recalled every three and a half years (the half-way point) for an image refresh.

Tapes that are vaulted should be monitored for data expiration, after which the storage media can be removed from the vault and recycled back to the scratch pool for re-use.

Depending on specific industry needs, some data may not be removed off-site for vaulting. For example, some medical / pharmaceutical companies only need several years of laboratory work to deliver a drug, but the process of bringing the drug to market can be far longer. For FDA auditing purposes, these companies may eventually be required to restore an experiment / protocol on-site. In these cases, companies would choose to archive data on-site for the convenience of data restoration rather than remove them to off-site vaults. However, best practices dictates that an off-site duplicate / clone is still necessary for disaster recovery.

Tape Recycling

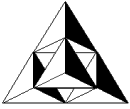
Enterprises are not expected to save any and all information forever. After passing their businesses' financial, legal or regulatory requirements, expired data can be erased, and the storage tapes are made available to the scratch pool for use again.

Storage media, no matter how robust, have finite longevity. Tape naturally wears out after a number of soft-corrected errors, or after it has endured a certain number of read/write passes that include friction with tape heads. Backup teams should closely follow the tape manufacturers' recommended guidelines, and discard the storage media after the maximum number of passes is exceeded, and after all of the data therein is expired. Again, enterprises should watch for backup software that enables the monitoring of media life and notification of expirations.

Variations in Best Practices for Category 2 Data

Business-critical data, such as those found in data warehouses and e-mail servers, are important to business operations, but not as critical as category 1. META Group recommends that the backup schedules and strategies for category 2 data follow a similar pattern as those for category 1, with daily, weekly, monthly and annual procedures. However, there are several aspects where category 2 data may not need backup processes that are as stringent.

In the case of business-critical data, duplication / cloning of data is optional, and as opposed to category 1 data, cloning of category 2 material is acceptable. Also, interleaving data onto the same media is not as serious a concern because data restoration performance and timeframe are less important. With the use of interleaving, backups (especially network backups) may become faster, and unique media are not required, which allows for denser packing of data onto tapes.



Variations in Best Practices for Category 3 Data

Category 3 data resides at the individual level. As such, it lacks the advantage of centralized control and monitoring, and backups are harder to enforce and track. Still, there are operational best practices that can be put in place to minimize the impact of data loss or corruption.

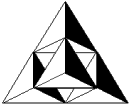
The first step is to prompt users to back up data on a regular basis to corporate resources, such as a networked file server or collaboration database. Backing up to corporate resources, rather than individual ones, provides several benefits. First, backups are managed by professional IT staff (ideally a centralized team with consistent processes, formats, and procedures) who have more advanced knowledge and tools at their disposal to ensure data protection. Second, corporate data is retained even after individual departures, preserving corporate knowledge. Third, when category 3 data is delegated to corporate resources, the business will have an easier time outsourcing the process to a storage service provider (SSP). (For the purposes of this paper, we will not be discussing best practices for outsourcing.)

The retention schedule for this level of data is much more flexible. Nevertheless, the corporation should set aside some integrated policies as guidance for employees, in order to ensure proper management of data volumes and formats. Retention for corporately managed data, such as project files, and some e-mails, should be managed at the server level for better control. For example, some organizations may choose to destroy e-mail after a specified time period for legal protection, rather than leaving it up to the employees. Other forms of user data should and can be retained until deleted by the user.

Best Practices for Vaulting

Vaulting procedures need to be part of a comprehensive data backup strategy, not only because the process provides for a higher degree of data security and preservation, but also because it is mandated in many industries, such as financial and legal institutions. Government regulations require data to be kept for a certain period of time as a matter of record-keeping, auditing, and other forms of accountability.

Vaulting not only meets regulatory requirements, it also makes good common sense as part of data backup and swift restoration best practices. Moving data to a separate, off-site location protects against data loss when, for example, the company's physical site is damaged by fire, water, or other natural disasters. There are companies that employ more than one data vault for additional protection. The important role vaulting plays in data restoration / recovery highlights the need to keep vaulted data up-to-date and refreshed. The vault's environment must also be monitored to ensure the proper



maintenance and protection of data against contamination and other forms of corruption. Also, companies should seek out software packages that provide robust monitoring of archived data.

Furthermore, vaulting should always include everything that is necessary to implement a fully functional system, so that if all equipment at the corporate site is damaged, the company will still be able to access and retrieve stored data at the off-site location. As part of best practices, make sure that the vault contains current copies of the operating system, the backup / recovery software, application software, and other data-related devices. Furthermore, enhanced versions of backup software must be able to read archives.

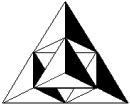
Operational Best Practices

Data backup technologies will never function to their maximum potential if an organization does not employ sound operational best practices to guide and augment technology usage.

As mentioned at the beginning of this document, business and operational planning are foundational for high quality data backup practices. Data backup technologies help, but they will never function to their maximum potential if an organization does not employ sound operational best practices to guide and augment technology usage.

To jump-start data backup activities, the company should conduct a survey of enterprise-wide systems and application data, and categorize them appropriately into the three categories outlined in this paper. Bear in mind that the category definitions provided above are not written in stone. If a company determines that, for example, its data warehouse information is critical to the business' survival, by all means it should classify that data as category 1. Also, if a company has enough resources to spare for data backup, it can potentially choose to employ more stringent backup policies, similar to category 1, for category 2 data.

On a related note, there also needs to be a survey to determine the extent and scalability of the company's enterprise backup resources, such as its tape libraries, backup software, backup servers, etc. This audit will help determine whether the desired best practices can be implemented, and if not, where the gaps are. Frequently, companies will find themselves with multiple operating systems and varying IT architectures across business divisions. While consolidation and standardization should take place, companies will need a backup software solution that can support multi-platform servers and operating systems to overcome this hurdle.



META Group

208 Harbor Drive, PO Box 120061, Stamford, CT 06912 (203) 973-6700 Fax: (203) 359-8066

One aspect of data backup that is important for these audits to uncover is the proliferation of backup software, hardware, formats and policies across divisions in the enterprise. To ensure consistent execution of best practices, the wide diversity of backup components needs to be reduced to a standardized set of vendors. A more homogeneous environment allows for easier facilitation of backup skills transfer between operations, more consistent backup processes, and improved remote support. Also, companies should look to reduce the number of backup / recovery vendors with whom they contract. Proliferation of vendor arrangements will only prolong implementation and training cycles, as interoperability issues mushroom. Furthermore, if these vendors support different backup formats, the confusion could exert a very negative impact on disaster recovery / data restoration efforts.

Centralization is increasingly critical as companies continue to build more far-flung operations, frequently with individual IT teams and architectures. Companies with truly wide-ranging locations could consider a strategy of centralized control with distributed processing. The goal is to ensure that the business, as a whole, has consistent data backup procedures, vendors / products, and infrastructure. One “master” location will hold the overall backup policy information (schedules, rotation practices, etc.) and then push out the policies to remote operations. The central location will also collect backup reports from all other sites, but the actual data processing and backup will be done independently at each site. Administrators will need strong reporting capabilities from the backup / recovery software in order to manage distributed data processing operations.

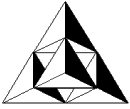
Finally, as part of operational best practices, META Group recommends that data protection and restoration operations be included with annual financial audit processes, to ensure that backup operations are run according to sound principles, have safe pre-cautions against data loss / tampering, and are meeting regulatory requirements.

Other Software Selection Criteria

Throughout this document, we addressed many backup functions that can be handled by backup / recovery software packages. There are a large number of solutions in the market today that collectively encompass all necessary enterprise backup needs. However, there is no single package to date that can do all things. Companies will need to carefully evaluate the extent of their backup / recovery requirements prior to making a final software selection.

Some other selection criteria that should be applied to the software of choice include:

- Does the software support backing up large file systems (e.g. over 4 GB)?



META Group

208 Harbor Drive, PO Box 120061, Stamford, CT 06912 (203) 973-6700 Fax: (203) 359-8066

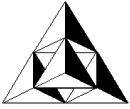
- Can the software break up large data sets into smaller ones for backup to multiple tape drives?
- Does the software allow for administrators to specify the size of the data streams, based on network congestion conditions? Or time of day?
- Does the software support data compression prior to streaming?
- Can the software support all industry-standard data formats?
- Can the software work with industry-standard hardware (tape drives)?
- Does the product support HSM (hierarchical storage management) processes?
- Can the software be used for monitoring / managing vaulted (off-site) assets?
- Does the product offer fail-over protection (i.e. re-route backups to another device)?
- How easy is the software to install across multiple clients?
- How does the package execute file restoration procedures?
- Can the software make duplicates?
- Does the software have backup diagnostics and / or guidance?
- Does the software assist with media retirement and replacement (i.e. track hard / soft errors)?

Finally, be sure to evaluate how easy it is to customize the software for the company's unique business requirements, as data backup is by no means merely a formulaic procedure.

About the Authors and Methodology

Information in this document is based upon META Group's daily, intimate interactions with and feedback from leading end-user organizations and IT vendors. The analysts and consultants involved in the creation of this document have had extensive experience providing strategic research and advice to global enterprises in their data management implementations.

META Group is a leading research and consulting firm, focusing on information technology and business transformation strategies. Delivering objective, consistent, and actionable guidance, META Group enables organizations to innovate more rapidly and effectively. Our unique collaborative models help clients succeed by building speed, agility, and value into their IT and business systems and processes. Connect with metagroup.com for more details.



META Group

208 Harbor Drive, PO Box 120061, Stamford, CT 06912 (203) 973-6700 Fax: (203) 359-8066

DLTape and the DLTape logo are trademarks of Quantum Corporation.

APPENDIX I – BACKUP ROTATION SCHEMES

Grandfather-Father-Son

The backup rotation scheme that is described on pages 7 and 8 of this document, and is considered one of the most common media rotation schedules. Grandfather-Father-Son (GFS) refers to monthly-weekly-daily backup activities respectively.

The GFS scheme begins with the daily backups. Typically, four tapes are labeled for a particular day of the week (e.g. Monday through Thursday). Each tape is recalled for use each week on its labeled day.

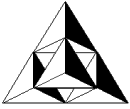
For weekly backups, data personnel usually set aside five backup tapes that are labeled according to the specific week in the rotation. On the day that a daily (son) backup is not done, a full backup is done on the weekly data. Each “father” (weekly) tape is re-used on a monthly basis, after the monthly (grandfather) backup has been completed.

Finally, monthly data (stored on separate tapes labeled for the particular month of the quarter – e.g. Month 1, Month 2, and Month 3) is recorded on the last business day of each month and is returned to the scratch pool for reuse quarterly.

As a result of this rotation (assuming that the company plans to keep any particular round of data for about 2-3 months), about 12 save-sets are created.

Month 1

Mon	Tues	Wed	Thurs	Fri
------------	-------------	------------	--------------	------------



<p>Grandfather-Father-Son Media Rotation Scheme: The white squares represent the most recent backups while the shaded squares represent previous backups. Only the daily tapes have been reused. Note that the weekly backup is performed on Friday.</p>
--

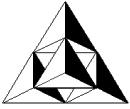
Tower of Hanoi

The Tower of Hanoi scheme is based on the mathematical game that bears the same name. The game uses three poles and three different-sized rings, and starts with all three rings on one single pole. The challenge for the game-player is to move all the rings to another pole, without ever placing a larger ring on top of a smaller one. To accomplish this feat, a specific “rotation” sequence must be determined.

Similarly, in data backup, the organization employs several media sets, e.g. Set A, Set B, Set C, and so on, but only one media set is used on any given day. Set A is backed up every other day; Set B begins on a day when there is no Set A being backed up, and is repeated every four days. Set C begins on a day when there is neither Set A nor Set B activity, and repeats every eight days. Media set D starts on the first non-A, B, or C backup day and repeats every 16 days.

				W1
				W2
				W3
		Wed	Thurs	W4
Mon	Tues	Month 1		

The advantage to the Tower of Hanoi scheme is that with each new media set added to the rotation, the organization’s backup history is doubled. Also, this scheme allows for easier and more recent data recovery, due to the frequency of backup for each media set.



META Group

208 Harbor Drive, PO Box 120061, Stamford, CT 06912 (203) 973-6700 Fax: (203) 359-8066

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A		A		A		A		A		A		A		A	
	B				B				B				B		
			C								C				
							D								
															E

Return to Day 1

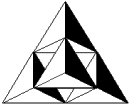
APPENDIX II

Glossary

Cloning: The process by which tapes are duplicated offline, usually when there is not enough tape drives available for online duplication. Cloning utilizes drive resources on an as-available basis lowering the associated costs; however, while waiting for drive resources to become available, only one copy of data exists increasing the risk of data loss.

Duplication: The process by which multiple copies of save sets are created online. Duplication requires dedicated online resources, increasing associated costs, but provides the highest level of data protection.

Interleaving: The technique by which multiple streams of data from different sources (tape drives) are recorded onto the same storage media concurrently. Interleaving reduces total time to backup and resources consumed but creates longer restoration timeframes.



META Group

208 Harbor Drive, PO Box 120061, Stamford, CT 06912 (203) 973-6700 Fax: (203) 359-8066

Multi-streaming: The process by which multiple streams of data are transferred to separate storage media at the same time. Because multi-streaming produces multiple storage media concurrently, it is time efficient and reduces the risk of losing mission-critical data in one fell swoop. However, multi-streaming is also more resource intensive, requiring high network capacity and a high volume of storage media.

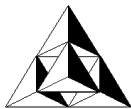
Vaulting: Tapes are stored and managed offsite for long-term storage.



APPENDIX III

Backup Best Practices Summary Chart

	Category 1 Data	Category 2 Data	Category 3 Data
Definitions	<i>Mission-Critical Data</i> OLTP (online transaction processing) ERP systems	<i>Business-Critical Data</i> Data warehouse Email servers	<i>Employee-Level Data</i> Individual email Spreadsheets Documents
Backup Retention Strategy – Weekly Full Backup	<i>Retain through the end of the month</i> <ul style="list-style-type: none"> Operating system data B/R software Content data Configuration files Registry files Application software data 	<i>Retain through the end of the month</i>	N/A – Data resides at individual level, backup schedule cannot be enforced or tracked
Backup Retention Strategy – Incremental Backup	<i>Changes made to the files since the last full backup on a daily basis</i> <ul style="list-style-type: none"> Delta incremental backup not recommended 	<i>Changes made to the files since the last full backup on a daily basis</i>	N/A – Data resides at individual level, backup schedule cannot be enforced or tracked
Backup Retention Strategy – Monthly Full Backup	<i>All data content and files</i> <ul style="list-style-type: none"> Retain in the media set for 30 days Retain in offsite vault for 1 year or through end of year 	<i>All data content and files</i> <ul style="list-style-type: none"> Retain in the media set for 30 days Retain in offsite vault for 1 year or through end of year 	N/A – Data resides at individual level, backup schedule cannot be enforced or tracked
Backup Retention Strategy -- Annual Full Backup	<i>Immediate offsite vaulting for all data and files</i> <ul style="list-style-type: none"> Retain for 7 years or regulatory timeframe 	<i>Immediate offsite vaulting</i> <ul style="list-style-type: none"> Retain for 7 years or regulatory timeframe 	N/A – Data resides at individual level, backup schedule cannot be enforced or tracked



	Category 1 Data	Category 2 Data	Category 3 Data
Operational Best Practices for Decentralized and Individual Backup	N/A – Backup is strictly scheduled, centralized, and monitored	N/A – Backup is strictly scheduled, centralized, and monitored	<i>Consistent backup to corporate resources (networked file server, collaboration database)</i> <ul style="list-style-type: none"> • More flexible retention schedule • Corporately managed data (project files, some email) managed at the server level for better control
Interleaving	<i>Not recommended</i> <ul style="list-style-type: none"> • Unique media required 	<i>Acceptable</i> <ul style="list-style-type: none"> • If the data retention schedules are similar for the different save sets • Unique media not required 	N/A
Multi-streaming	<i>Recommended</i>	<i>Not necessary</i>	N/A
Duplication	<i>Critical</i> <ul style="list-style-type: none"> • Unique media always required • At least one duplicate copy immediately vaulted offsite • Rigorous retention schedule • Duplication for annual full backups considered mandatory 	<i>Optional</i>	N/A
Cloning	<i>Not recommended but should be used if duplication is not feasible</i>	<i>Optional, Acceptable</i>	N/A
Vault Rotation and Archiving	<i>Recommended offsite current copies</i> Operating system B/R software Application software Other data-related devices	Image refresh every 3.5 years	N/A



META Group

208 Harbor Drive, PO Box 120061, Stamford, CT 06912 (203) 973-6700 Fax: (203) 359-8066

	Image refresh every 3.5 years		
--	-------------------------------	--	--