

## KNOW YOUR DATA AND YOUR STORAGE FOR BETTER BACKUP AND RECOVERY

Backup and recovery continues to be an important but perplexing part in an otherwise solid data management and protection plan. The usual suspects include growing data loads, nonexistent backup windows, too much network traffic or poor performance, inefficient storage hardware, media management problems and poor backup reporting. However, these problems generally all stem from one common source: a lack of in-depth understanding of the storage environment and how it satisfies business objectives.

Many business environments have data consisting of mainly small and medium-sized files - and a lot of them. Others may have fewer larger files, and still others a combination of small and large. The complicated nature of data management makes backups a crucial issue in IT. Businesses need to be assured their backups are successful and that they're backing up what really needs to be backed up.

Data types may vary dramatically, such as email, databases, multimedia and graphics, and data may serve one or multiple related applications. Data also exists under differing degrees of confidentiality, such as company financial information, employee records or customer transactions. There may also be a significant amount of prohibited data if company policy is against employees keeping MP3 files or unauthorized software on the network.

Depending on the size and nature of the organization, some decisions about backup and recovery may or may not be flexible. Federal, state and local governments have imposed requirements such as Sarbanes-Oxley to ensure that enterprises are conducting business properly. You may be subject to regulations from the ATF, DOD, DOE, DOJ, FAA, FDA, IRS, SEC, and a handful of others.

In order to better manage disparate backup needs and requirements, and before acquiring new storage hardware or adding another layer of network complexity to achieve better backup, perform a thorough assessment of the storage environment, and the data itself, that correctly identifies the problems. Are backups taking too long because unchanged, obsolete data is being backed up too often? For example: Are there a small number of large files to back up, or a large number of small files? Is the data available during its scheduled backup, or is it in use, and if it is in use, is it always or sometimes? Are unnecessary or old data files being repeatedly backed up. Once you have thoroughly evaluated the data and the available storage resources, you will likely find more options than reactively purchasing hard disks or ever-larger tape libraries to solve a data management problem.

To make an intelligent decision on data protection, you need visibility into the backup and storage environment. The first step in gaining visibility is finding out why the data exists, who owns it, where it resides, when it was last accessed, what level of archiving versus availability the business application requires, and then determining whether current technologies can meet these needs.

Once the data itself has been evaluated in relation to application and usefulness, a more granular view of the environment is achieved by proceeding to an evaluation of the repositories where data is kept and where it must be duplicated for backup and recovery. This includes DAS, NAS, and SAN resources, file servers and application servers such as DNS servers, SQL and Exchange, as well as desktops and notebooks, which may contain data that is just as valuable and irreplaceable to the operation as data stored on servers.

As part of your assessment, note the storage capacity both in use and available for backup destinations to estimate the overall health and life-

span of these systems in order to help with capacity planning. This task may require a forecasting or trending tool depending on the size and complexity of the network, preferably one that will also identify wasted storage space or 'islands.' Some questions to consider include: Is server consolidation an option? Is it possible to compare allocated versus unallocated storage space? Which are the fullest systems, and what is the likelihood of maintaining them at present capacity?

Once the data and the storage environment have been evaluated, the next step is to consider the requirements for protecting the data to determine if these requirements can be reduced. Determine how necessary it is to access the data, how frequently requested the data is, and how quickly it must be recovered - including how tolerant executive management may be in the event of downtime.

An assessment of your storage environment and backup systems is the first step in establishing and clarifying effective policies and procedures for managing data. Policy-based tools to manage storage resources are gaining in popularity and usability. This category can include information lifecycle management, compliance with data retention regulations, assigning certain applications performance priority, assigning certain departments access priority, and other operational goals to be met.

Lifecycle utilities are frequently seen as an efficient complement to applications such as email or database programs in order to evaluate and match the frequency of accessing data to proper storage so that older data can be identified before it becomes oppressive to manage. Looking at data to view how the age of the files can be categorized (such as 0 to 30 days, 31 to 90 days, and 91 days or older) may help in making smart decisions about what data is eligible for archiving or even deleting.

Grooming or vetting data prior to backup will reduce the backup load; take the example of tape storage. For most enterprises where tapes are regularly taken offsite for archiving purposes, the data may be staged first to a disk system in a present-day equivalent of hierarchical storage management. Disk storage allows fast and easy recovery for frequently accessed files in the case of accidental data loss. In the case of data that need not be as frequently accessed, depending on the acceptable recovery time, manual tape retrieval may be perfectly reasonable for most operations. This data can be removed from onsite disk storage and archived on tape with little sacrifice in efficiency and a realistic timeframe for recovery.

When data management is happily married to storage, and current technologies are meeting your needs, maintaining harmony is still an ongoing challenge. In most cases, a storage environment will consist of SAN, NAS, DAS and backup resources, as well as critical 24/7 application servers like DNS servers and SQL servers, all working separately and in concert, inter- and intra-departmentally, core-to-edge and end-to-end, so it becomes more important to be able to perform efficient, dynamic monitoring of as many or as few points as are likely to be taxed.

For enterprises using a SAN, continual monitoring of the network's health is a priority for the best degree of data availability, backup success and restorability. A downed switch or a pulled-out cable will assuredly prevent a proper backup from being carried out. There are several effective management applications either pre-packaged with SAN hardware or available separately that can provide local views of the network topology to troubleshoot and even provide analysis and reports. However, you may or may not be able to integrate these with other reporting tools you rely on, and they may not be thorough enough to include switch I/O trending, asset management, alerting and notification, as well as performance monitoring.

Application monitoring should be considered an essential part of storage management. In the case of DNS, when a critical DNS server goes down, the quicker the problem is identified and corrected, the sooner data can be made available - meaning less financial loss is incurred and less recovery time may be required later. Similarly, monitoring SQL or Exchange resources, reduces total backup and recovery costs by eliminating application downtime, grooming data and even finding poorly allocated space.

Monitoring NAS filers as well as any accessible DAS will also allow file analysis, trend reporting, and capacity planning to assure a longer life and better availability to reduce operational expenses and assure simpler recovery.

With these tasks under control, it becomes possible to implement additional layers of enforcement on users to protect the health of the storage network, such as profiling the usage and allocations of individual departments who are prone to becoming space hogs, defining more comprehensive policies on retention of obsolete or duplicate data sets, identifying taboo files such as MP3s, as well as staying attentive to chargebacks.

Even with the most diligent management and reporting and the most functional utilities on the market today helping to head off backup and recovery problems at the pass, administrators occasionally will still be found scratching their heads over an unexpected failure. In fact, backup ailments may be even more inexplicable because of how healthy the storage body itself has become. This serves as a reminder to be vigilant, but also invites the question of whether such efficient diagnosis tools can be used on the backup and recovery process itself.

Backup software vendors, like most hardware vendors, will usually provide their own log or report function, though they are by no means the most powerful, reliable, or understandable. They may provide only partial information, such as when a backup failed, without telling where or why. They may indicate where a backup stopped before completion, but not indicate with any confidence that it will resume normally without operator intervention. Even if the information provided is adequate, they will still leave an administrator scrambling to compare logs and rewrite scripts daily once a point and/or time and/or reason for failure has been located and corrected.

Knowing when and why a backup failed and automatic backup comparison reports prevents backup headaches. Also key, of course, is information on successful backups including host name, client name, save set name, backup level, volume name, size, start time, end time, duration and speed.

When considering reporting utilities such as SRM or ILM software, performance monitors, SAN topology managers, or capacity planning - or an integrated suite of products that offers all these related but distinct functions within a department or across an enterprise - look for real-time monitoring for always-on management; off-the-shelf reports that require little time to process or can be easily customized via a wizard; a browser-based/web-accessible view; and the ability to save information over time for forecasting and trending.

Storage networks are heterogeneous, and storage departments are distributed. Due to organizational structure or budgetary restrictions, IT departments may need to solve a very specific pain point quickly without launching an all-inclusive storage management solution right away. If there seems a risk of biting off more than one can chew, or your responsibilities are limited to a single department or remote location in a larger operation, be aware that there are still several ways to control your storage challenges without a one-size-fits-all approach or by achieving what you need today and expanding tomorrow.

You may only require a consolidated, centralized view of backups across one or a small handful of servers, with or without looking at these backups in relation to the entire storage network. You may be in a position of having to control your costs and justify your usage at the request of an entity that is higher on the food chain. You may be facing service-level

agreements that require you to provide hundreds of reports each period to other departments within your organization. You may see the value in keeping your data groomed, systems balanced and potential bottlenecks scrutinized for the sheer altruistic joy of a job well done - that is, in addition to all of the above. In short, even if your control requirements do not stretch end-to-end, but only to your end, you may find that integrated management and reporting tools will provide as little or as much pain relief as you need, and are modular and scalable as your areas of responsibility and capability grow.

Performing a proactive diagnosis of data management problems is essential prior to considering any acquisition of storage or backup resources. Armed with in-depth information, it is possible to make a responsible decision instead of reactively or reflexively purchasing hardware. When evaluating your management options, consider toolkits that provide granular visibility into your environment and that offer a full complement of monitoring and reporting.

Storage challenges continue to grow as budgets diminish, but management software that goes the distance such as Profiler, allows you to save time, ensure data availability, prove backup success and address many other device- or application-specific pain points to monitor usage patterns, simplify administration, and ensure consistency in your organization. With the modular approach of Profiler, users can now implement as little or as much pain relief as they need and scale up as their network grows, while preventing unnecessary equipment acquisitions.

Tek-Tools, Inc.  
4020 McEwen Dr., Ste. 105  
Dallas, TX 75244  
972-980-2890  
[www.tek-tools.com](http://www.tek-tools.com)