



Measure Twice, Cut Once

ROB SIMS, President and CEO of Crossroads Systems has one answer to **IT budget cuts when it comes to long-term data storage**

With layoffs and budget cuts all too common during today's global economic crisis, organizations' IT departments are faced with a tough task: balance the ever expanding requirements to store and manage data with the reality of capital spending freezes. With belt tightening occurring over the last few years, IT managers, storage administrators and backup and recovery teams are forced to take a deep look at their priorities and explore new, creative ways to cut everyday costs.

Of course, one of the most obvious methods is to continue the use of existing functioning equipment and delay the purchase of new systems. Additionally, a general trend over the last couple of years has been in the area of maximizing system efficiency. The primary focus has been on server virtualization where companies are able to create multiple virtual servers from one physical server – greatly reducing their capital expenditures and power consumption. It is this same focus that is now required on their storage infrastructure and systems.

One old battleship in most organizations, storage infrastructure is their tape drives and libraries. Increasingly, corporations have been moving some of their data protection requirements away from tape and into disk systems whether that be disk-to-disk, replication, or virtual tape solutions that emulate tape drives and libraries but store the data on disk subsystems. While there are many benefits to a disk-based backup solution, there are still many long-term data retention requirements that have the need for tape solutions. Of course, there are several disk-based manufacturers that are pushing to utilize disk systems for long-term data retention requirements as well, completely removing the need for tape.

While the vendor community continues to push for "ripping-and-replacing" existing tape systems with disk-based solutions, the current economic realities coupled with the increasing need to pursue green technologies is having IT professionals dust off their existing tape systems and take another look.

However, the challenge with tape for the IT administrator hasn't changed. Tape has received a bad reputation over the past few years, somewhat due to one-sided press from disk vendors, but also from legitimate issues with the solution. Many of those issues have not been adequately addressed by the tape vendor community, leaving their customers looking for other ways to get the job done.

Regardless of the approach taken to solve

the tricky balance of shrinking budgets and increasing data requirements, the IT organization needs to maximize their knowledge of the environment in order to make the most informed decisions about their long-term data storage requirements. It's an approach rooted within the old adage "measure twice and cut once." If someone only has one piece of wood remaining for a project, it's clearly important to be sure your cutting measurements are correct – the first time around.

THE CASE AGAINST RIPPING AND REPLACING

Before the economic situation the business world is facing today, organizations were willing to write big checks for new disk-based systems in an effort to replace tape systems. Reliability was typically the main concern that companies were referring to as there was no way to pinpoint the components that were not tuned properly, conflicting or malfunctioning within a tape environment. An accumulation of various environmental inefficiencies were not detected by backup applications. Therefore, several storage administrators found themselves crossing fingers when trying to recover data.

Intricate marketing campaigns from some of the largest IT companies have pushed for organizations to solve their tape storage issues by changing to a miraculously trouble-free "SAN-in-a-box" (as long as the company pays fairly substantial, ongoing fees for the associated consultants). And at first glance, firms see this as a great option to avoid performance and management issues, and address their ever-growing amount of data.

Given the substantial failure rates that some firms may have experienced while trying to recover data stored on tape, accompanied by the gigantic marketing budgets of large disk-based solution providers, it's no surprise that several customers were convinced that buying extremely costly disk-based systems was the answer.

And although the price of disk-based systems has decreased a bit, hidden costs abound. These include costs associated with:

- Staff training and/or hiring outsourced resources that may need to become part of a vendor assurance program.
- For compliance purposes, additional units may need to be purchased to meet disaster recovery requirements – instead of just buying additional, less expensive tape cartridges.
- As a result of the integration with the operating system, disk-based solutions require

more frequent upgrades.

- Since tape is an actual storage location, it does not need to be powered up when the data is on tape. One can view the data on one tape without powering up all tapes. Whereas, with disk one must power up everything, all of the time (there are technologies for powering down disk systems, but those solutions come at a premium and will never reduce their power needs to that of tape).
- Most of the disk-based solutions for tape replacement use SATA disk drives which have a high level of failure. This requires costly maintenance contracts or the use of advanced RAID options (such as RAID6) whereby the system can recover from a situation when two disks fail. However, this is costly since there are now two disks used for recovery per RAID set and the performance is significantly impacted as well.
- For warmer climates, there are enormous power-cooling requirements.

Additionally, given today's M&A-based economy where heterogeneous environments are second nature, the difficulties of shifting to a new solution that is disk based can cause tremendous, costly hurdles to appear.

Of course, there are significant costs associated with the "rip-and-replace" process. Countless hours of internal planning coupled with the need for outside consultants can rack up costs in a hurry. This doesn't include the additional costs of labor required to reconfigure systems, test and deploy. At the end of this effort, the tape system is still required to deal with all of the old data that is already stored as it is extremely expensive, if not impossible to move all of that data to an online disk system.

The concept of replacing a functioning solution has been heavily promoted by large disk-based solution providers. Yet, as tape and disk both meet different needs by design, it is no longer viable for companies to throw away funds in such a manner or simply dispose of good technology that can easily be improved upon to meet disaster recovery needs and other corporate requirements. Disk and tape need to co-exist. Yet, to ensure efficiency within a storage environment, the frequent problems within tape environments need must be proactively and effectively addressed. After all, data keeps growing regardless of the economy and storage administrators are measured according to their ability to manage and recover all data.

Aiming to reduce new equipment acquisition costs, preserve older equipment, and lower

“PROACTIVE MANAGEMENT AND ACCURATE DATA TURNS A LARGELY UNSCIENTIFIC GUESSING GAME INTO AN INTELLIGENT PROCESS THAT BOTH IT AND CORPORATE EXECUTIVES CAN UNDERSTAND AND APPRECIATE.”

energy consumption, administrators can focus on reducing labor-intensive storage and backup strategies for nearly instant savings. Yet, with skeleton crews and extensive outsourced efforts, the option of ripping and replacing may be too challenging for managers that are truly concerned with cost management, business continuity, disaster recovery and compliance. So, the question to answer: Why rip out an existing system when it's easier and more cost-effective to simply improve a current system?

KEEPING THE OLD CAR RUNNING A BIT LONGER

Until a few years ago, no vendors addressed the need to proactively monitor tape – because companies showed constant willingness to simply buy more or upgrade to the latest technology, hoping the problems would be solved. This approach is no longer an option in the current realities of business. This is much like firing in the dark, hoping to hit the target. However, the required approach is one based on accurately measuring the current environment, identifying the areas in question, and solving the root cause issues. In other words, measure twice and cut once.

Traditionally, any effort made to improve the tape environment involved discussing the issues and listening to the solutions offered by the tape system suppliers. It's natural for tape storage vendors to recommend buying more tape libraries and upgrading an environment with new drives. Although a storage administrator may eventually have to consider some of these options depending on evolving needs, the fact is that most administrators using new tape monitoring tools discovered that they were only using a fraction of their current tape environment. It is not uncommon for tape libraries to have multiple drives sitting idle for months at a time, and those in use show average data throughput at a fraction of their native data rates. It is no wonder that so many companies

have backups close to or exceeding their time windows, since though their tape resources are used extremely inefficiently.

In many other areas of the network, IT professionals use monitoring tools and solutions to measure activity, efficiencies, and the overall effectiveness of an environment. The same approach is required in the storage environment. By implementing these solutions (in many cases these are also offered as services), corporations can be armed with the data required to make the appropriate decisions for their business.

Specifically in the tape sub-system, organizations need to implement tools that can correlate the data write/read errors between the tape drives and media to determine the cause of failure, measure the real-time utilization and performance of the tape drives, and validate the readability of the data stored on media.

One of the most frustrating challenges with tape is the multiple variables that exist for any failure. There is the tape media, tape drive, library, network, and backup or data server writing the data that can be the cause of failure. The proverbial chicken or egg issue with tape is, “what caused the media failure, the tape or tape drive”? Invariably, the user replaces the tape media since it is the easiest to change, and then they hope the problem will go away. However, research has shown that in many, if not most cases, the root cause of the failure begins with one or more of the tape drives. It is challenging to find these drives since it is quite possible that the defective drive wrote a piece of media and a subsequent drive failed to read the media at a later point, sometimes many months later. Since this is such a long term issue, which can foster over time – it is critical that an organization monitor all of the physical systems for every write/read operation.

The correlation of write/read errors by drive and tape will allow the storage administrator

to identify defective drives, and by using soft errors (those errors on tape which can be overcome via advanced error correction algorithms) it is possible to identify and remove drives and/or media before they can cause more serious data recovery issues.

Additionally, a drive that is experiencing issues in writing data (creating multiple soft errors) will cause the data transfer to slow significantly. This is due to the drive invoking rewind and rewrite operations in order to satisfy the minimum write error rate requirements. Without a monitoring solution, this issue would be completely unseen by the IT administrator – they would just see that the backup operation failed to complete in the time allotted or that it took much longer than anticipated. In fact a regional bank experienced this exact scenario, believing that their data had grown beyond their backup window. The bank purchased a new library and new, higher performing tape drives. However, after implementing a monitoring solution, they found defective drives which were the cause of their significant performance issues, resulting in the backup window problems. It was determined that had the real issue been solved, there would be no requirement to add new drives or a tape library for many, many years to come.

Another area of significant opportunity is in the utilization of the existing tape drives. It is difficult to always make the most out of every tape drive since the environment needs to be sized to accommodate the maximum data transfer requirements. Generally, the average data requirements are much less, leaving systems sitting idle. While the tape system might have been in balance when it was originally purchased, it is likely that after multiple changes that occur in any data center, the utilization of the tape drives is no longer efficient.

Just like a new car requires a tune up after it is used for a while, the utilization of tape drives also requires tuning. In fact, in many

“AFTER ALL, DATA KEEPS GROWING REGARDLESS OF THE ECONOMY AND STORAGE ADMINISTRATORS ARE MEASURED ACCORDING TO THEIR ABILITY TO MANAGE AND RECOVER ALL DATA.”

environments tape drives seem to get lost, never being used for months at a time. A major telecommunications company had six drives (out of 47) that had no I/O (write or read) operations in the five months that was monitored. This either represents savings in the form of ongoing maintenance and power/cooling requirements, or the opportunity to rebalance and improve the backup window.

The opposite problem also occurs where certain tape drives are over utilized. This comes in the form of constant data transfer and/or excessive tape load/unloads. Every drive type has a duty cycle rating which defines how often the drive should be used (typically in a 24-hour period) as well as the average and total load/unloads it is rated for. There are environments which can exceed these rates, in some cases quite significantly, such that the tape drives will experience premature failure. From a user's perspective, this appears to be a problem with the overall quality of tape systems, when in fact the issue was caused by inappropriate use of the devices.

The data throughput performance of the tape drives is another area of concern and required focus. Most users purchase the latest and greatest tape drive for its throughput (typically doubled from the last drive) and tape capacity. The problem is that most environments struggle to meet the minimum required throughput to keep the latest tape technologies streaming data to the tape medium. Every drive comes advertising the native data rate and the 2:1 compressed

rate. However, there is a critical number, called the streaming rate, which isn't advertised as prominently. This is the rate at which data must be consumed by the tape drive to keep the tape media streaming past the write head.

When data falls below the streaming rate, the drive must stop writing, back-hitch (reverse the tape), wait for enough data to fill its buffers and then reengage to start writing again. This is a time consuming operation that causes the overall performance to drop further below the streaming rate and can also potentially cause physical damage to the media that won't show up until the next time a tape is read.

As an organization looks to improve the backup window or add more data to the existing backup process, many times the only option available is to look at purchasing newer, faster tape devices. However, if the environment cannot keep the existing technology streaming, adding next generation drives will do little to solve the problem. The analogy is being stuck on the highway in a traffic jam. Buying a Ferrari won't get the driver home any faster. However, if the I/O throughput can be improved, an environment that is below streaming rate could see a 100 percent or greater improvement in performance. This is by far the most significant variable that should be attacked versus buying latest generation systems that might at best see a 5 to 10 percent improvement.

Finally, organizations should implement a proactive process by which existing media is verified. This can be done offline, not requiring

the use of existing servers. It will give the IT staff the ability to identify defective media before it is required by the business. This not only solves existing compliance and governance requirements, but gives the storage administrator the ability to solve the problem in a timelier and cost-effective manner.

IN CONCLUSION

Organizations are finding they can make informed decisions rather than making costly, untimely or completely unneeded purchases. Proactive management and accurate data turns a largely unscientific guessing game into an intelligent process that both IT and corporate executives can understand and appreciate with real-time information that can truly assist in proactively managing an environment.

To address cost cuts and headaches with tape efficiency, advanced tape storage management is more than marketing hype – it's a dollar-driven, time-tested approach to an otherwise costly alternative. It is part of the larger strategy that many of the world's largest companies have doggedly stuck to in recent months: leverage existing tools and stop buying more systems than are really needed.

Much like that of a carpenter drilling it into the mind of his young apprentice to “measure twice, cut once,” this approach will undoubtedly be implemented by a growing number of organizations, asking their data storage teams to take a deeper, more accurate look at their true needs. **BTQ**



ROB SIMS is President and CEO of Crossroads Systems (www.crossroads.com), a global provider of solutions that connect, protect, secure and restore data. An industry veteran and pioneering voice in data storage and protection strategies, Mr. Sims brings more than 20 years of hands-on experience in IT design, development, manufacturing and engineering. In 2008, he was recognized with the industry's Most Valuable Performers Award.