



Watts Up? How Much Electricity Does It Take to Backup Your Data?

Analyst: Dianne McAdam

Management Summary

The amount of data that IT organizations need to store and manage grows every year. In fact, many corporations report that they are experiencing data growth rates of 50% or more every year, and it appears that there is no end in sight.

Corporations are not only storing more data but are keeping it for longer periods of time. All of this data must be protected against disasters, viruses, and of course, human error. Backing up data to tape has been a traditional approach to protecting data. When data is lost, corrupted, or accidentally deleted, it can be restored from the most current backup tape.

Tape vendors have continued to increase the performance and capacity of their devices to keep pace with the growing amount of data that needs to be backed up. Disk vendors also have brought higher and higher capacity disk storage systems to the market. IT organizations continue to buy more storage to store primary data and its backups. That makes the storage sales representative happy but strains already limited IT budgets.

Little can be done to stop or reverse the growth of data. In fact, recent surveys have shown that this explosive growth of data (usually coupled with shrinking backup windows) is the top concern of storage administrators. Some describe this as a tsunami of data hitting the shores of the data center. It is a tsunami that never seems to end.

If budgets were unlimited, we could continue to buy more and more storage devices to store our primary and backup data. However, many data centers are facing a new problem – they can no longer bring enough power into their data centers to keep servers running, disks spinning, and tape robots active. Some companies have moved their large data centers near dams to get inexpensive electrical rates (and abundant power). Migrating data centers and relocating or hiring new talent can be impractical for many companies.

Recent surveys have shown that one of the major concerns for CIOs is the cost to power and cool their data centers. In some cases, data centers are located on electrical grids that can no longer supply the electricity to power and cool existing and future hardware purchases. CIOs should be concerned about power. It has been estimated that in a few years it will cost more to power and cool servers than to buy servers. Many customers and vendors are now focused on saving energy with servers. It makes sense that this project gets high priority. Several large data centers estimate that about 60% of the power in the data center is consumed by servers, but that percentage may soon change. If our storage needs continue to grow, soon it is likely that 60% of the power in the data center will be dedicated to powering and cooling storage devices. **We now need to think about storage in terms of the cost of energy to power and cool the devices.**

IN THIS ISSUE

➤ The New Metric for Storage	2
➤ The Size of the Backup.....	2
➤ Backing Up to Tape	2
➤ Backing Up to Disk	2
➤ VTLs - Advanced Functions.....	3
➤ Conclusion	4

The New Metric for Storage

In a previous life when I worked for a vendor, one of the questions that I was always asked was the *cost per gigabyte* of the latest storage system. It was a convenient and important metric that allowed customers to compare two different storage systems easily. In those days, I usually was only asked two questions concerning power – (1) *What kind of plug was required?* and (2) *Did the data center have enough cooling capacity left in their air conditioning system to support the new devices?* Few cared about the cost to power the devices.

Power costs are a big concern today and they should be. We need to be concerned about the cost per terabyte of storage, but **we must also be concerned about the cost of energy per terabyte of storage.** This new metric will help to determine the most energy efficient storage today.

So how much energy does it take to backup data? The answer is multi-dimensional – it depends on the storage target, and it varies a great deal. Let's examine the costs to backup data to several different devices.

The Size of the Backup

For the sake of this discussion, let's assume that an IT department schedules a full 70 TBs backup every week. During the week, they run incremental backups to capture the data that has changed daily (about 10%). They want to retain all of their full backups and 10 days of incremental backups. By the end of the fifth year, they will need to store:

- 70 TB x 52 weeks = 3,640 TBs, plus
- .7 TBs of incremental backups x 10 days = 70 TBs
- Total storage requirements is 3,710 TBs

We have an eight-hour window to complete our full 70 TB backups, so we need a solution that can backup 8.75 TBs an hour.

Backing Up to Tape

For the tape system in our analysis, we choose LTO-3 drives and cartridges installed in an automated tape library. We used current list pricing for all hardware configurations.

We made several assumptions about the automated tape configuration.

- Each LTO-3 drive can read and write data at .288 TB per hour, when it is streaming data. However, the backup environment cannot always keep high-speed tape drives streaming for optimal performance. We estimated that these drives would run at 50% of its maximum performance throughout the 8-hour window¹. This means we would need about 30 LTO-3 drives.
- We assumed that data would be compressed at 2:1 using hardware-based tape compression and the cartridges would be, on average, utilized at 80% of their capacity. LTO-3 cartridges can store up to 800 GBs of data compressed at 2:1 on a single cartridge. That means we need about 5,800 cartridges.

The list cost for this configuration, which includes a 6,000-slot library is about \$955,000 or about \$260 per TB (\$.26 per GB).

The next generation of LTO drives, LTO-4, is now available. These drives will have higher performance and capacity. However, list pricing for this newest generation of tape drives was not available at the time this paper was written. It is expected that these new drives will reduce the average cost per GB. Since the newer LTO-4 will store more data per cartridge, it is also expected that the cost per kilowatt per terabyte will also decrease.

Unlike disk drives that require power to keep them spinning, tape drives require little power when they are not reading or writing tape cartridges. **In the case of our tape library configuration, the power and cooling costs per TB per year is about \$.45 (or \$.00045 per GB).** We used a conservative rate of \$.10 per kilowatt. Your electrical rates will vary, depending on your location. Many parts of the United States would be happy if their electrical rates were only \$.10 a kilowatt-hour!

Bottom Line: The cost of energy per TB for tape library systems is about \$.45 per TB per year. Another way to look at energy costs – the annual power and cooling costs per TB of the tape system was 0.2% of the acquisition costs.

¹ LTO-3 tape drives run at 80 MB/sec or 160 MB/sec with data compressed at 2:1. We used the more conservative number of 80 MB/sec.

Backing Up to Disk

While tape is very “energy friendly” storage, many applications demand quick restores that require the randomness of disk to meet their service level agreements. Some data centers choose to back up data to native SATA disk systems.

For this example, we choose a large vendor’s SATA disk system for our backup target. Each disk system consists of one dual controller and 56 TBs of disk. The total list price for 67 disk systems to store 3,710 TBs was over \$9,000,000 or about \$2,600 per TB (\$2.60 per GB). **The power and cooling costs for this SATA disk system is about \$96 per TB per year.**

Tape is not always a suitable backup target. For example, small files that require quick restores are better suited to be backed up to disk. However, there is a cost with backing up to SATA disk. Per GB, it costs 10 times more in acquisition costs and 200 times more in power and cooling costs. The additional power and cooling costs may not be very significant to data centers located near hydroelectric dams, where electricity is relatively inexpensive. Conversely, it is extremely important to those urban data centers that cannot pull any more power off the grid.

Bottom Line: The cost of energy per TB for SATA disk systems is about \$96 per TB per year. The annual power and cooling costs per TB of the SATA disk system was 4% of the acquisition costs.

Virtual Tape Libraries

Virtual tape libraries (VTLs) have been one of the most successful disk-based backup systems today. These systems usually incorporate SATA disk drives. The system emulates tape drives (and libraries) and responds to the backup application with normal tape commands. Virtual tape libraries are appealing since they require few, if any, changes to the backup environments.

VTL vendors have incorporated numerous features and functions into their solutions. For this reason, VTLs tend to cost more than native SATA disk systems. List prices for Virtual Tape Libraries vary by vendor and by feature set. We have seen list pricing that ranges from \$4,000 to \$10,000 per TB (\$4 per GB to \$10

per GB). In general, the larger the VTL, the lower the cost per terabyte. Since most VTLs are based on SATA drives, the cost per kilowatt-hour per TB is, on average, identical to SATA disk systems.

Bottom Line: VTL systems cost about the same for power and cooling per terabyte as SATA disk systems.

VTLs – Advanced Functions

VTL vendors have been developing new technologies that address the energy crisis. Two technologies that have been on the market for several years help to lower energy costs.

Massive Arrays of Idle Disks

Much of the data stored on secondary storage, such as backups or archives, is infrequently accessed. So how do you reduce the energy costs in a disk system? Don’t power on the disk until you need to. That is the architecture behind MAID systems. MAID, or *Massive Array of Idle Disks*, systems only power up disk drives when they are accessed or when periodic maintenance is performed. The result – fewer spinning disks costs less energy. Some MAID technology is offered as a disk solution; some MAID technology emulates tape drives and libraries and is part of a virtual tape library solution. In the MAID systems that we examined, the cost to power and cool one terabyte of data was about \$19 per year – quite a bit less than constantly spinning SATA disks. However, it does take time to power up the disk drives to access the data. The energy savings for MAID systems assumes that the vast majority of the disk drives are powered off all of the time. Your results will vary, based on your workload.

Bottom Line: MAID systems may only cost about \$19 per TB to power and cool the system.

Another way to reduce energy costs is by reducing the amount of data that must be stored on disk.

Reducing Power by Reducing Data

Eliminating duplication is the next evolutionary step in backup technology. This technology is called by many names such as *data reduction*, *global compression*, *compaction*, or *commonality factoring* but is more commonly

referred to as data de-duplication, or simply, *data dedupe*.

The premise is simple – we backup the same data over, and over ... and over again. A small percentage of the data changes daily (usually about 10%) with the rest of the data remaining fairly static. Finding duplication within this data and eliminating it can dramatically reduce storage requirements.

Reducing data is not a new concept. Compression that detects and eliminates repetitive characters has been available with tape drives for many years. This technology can usually reduce storage requirements by 2:1 in open systems environments. Many VTL vendors today offer software- or hardware-based compression that can reduce storage requirements by 2:1. That also reduces the cost per TB and cost per kilowatt per TB by 2:1 as well. Of course, your mileage may vary – some data compresses very well and some does not.

Data de-duplication takes this to the next level by eliminating duplicate segments of data.² Like MAID, some data de-duplication solutions are disk solutions while others are part of a virtual tape solution. What is the effect of data de-duplication on reducing storage costs? The answer is – *it depends*.

Data de-duplication can reduce storage requirements significantly but is application (data) dependent. Take the example of the email that is sent to 25 different people within an organization. This email was sent with a very large *PowerPoint* presentation for all to review and edit. Data de-duplication will store only one instance of that message and one instance of the presentation. That results in a reduction of 25:1. Now change the title to that presentation and send it to 25 people again, and the data reduction software will only store one copy of the second email and the changed title – again there is a significant storage reduction.

What can you expect to see for data reduction ratios? The answer is dependent on your data. In general, email and office documents see very high reduction ratios and images, such as digitized MRIs, will not benefit from data reduction at all. Some vendors

claim 20:1 reduction; others claim 25:1, 40:1, 100:1, or more.

To be conservative, let's assume that you have installed a data de-duplication solution. After several weeks, the data reduction ratio is about 20:1. How does that reduce your acquisition costs?

VTL vendors may charge a separate licensing fee for data de-duplication software. Some include the cost with the VTL system; others may charge on a cost per terabyte. There are software-only solutions that will work with any vendor's disk systems. No matter what solution you choose, you will need less disk storage. On average, the savings resulting from buying less disk storage will greatly offset the cost for the deduplication license. That reduces the acquisition costs. We have seen some solutions that can reduce the cost per GB to less than that of native SATA disk systems.³

There is another reason to consider data de-duplication. It reduces energy costs. If you need only one twentieth of the disks, then your energy costs will decrease significantly. Your power savings may be reduced proportionately with the disk reduction ratio.

Bottom Line: Data de-duplication systems can significantly reduce acquisition costs and power costs. The savings is dependent on the type of data and reduction rates achieved.

Conclusion

Years ago, we had one choice to backup data – directly to tape. Today, there are several options – tape, native disk, virtual tape libraries, and systems with MAID or data de-duplication technology. In fact, many solutions combine two or more of these options. For example, VTL systems can have integrated data-deduplication software. Other VTLs automatically migrate data to physical tape when appropriate. These combined solutions can

² See the issue of *Clipper Notes* dated February 1, 2007, entitled *The Evolution of Backups – Part Two – Improving Capacity*, and available at <http://www.clipper.com/research/TCG2007016.pdf>.

³ Some vendors perform the data de-duplication process in-band while others store the entire backup on disk before the data de-duplication process starts. Those that process the backup after it is stored on disk need additional disk capacity available to store the largest backup in its entirety. If the full weekly backup is 70 TB, then an additional 70 TBs always must be available to hold that backup.

provide high performance while reducing energy costs by reducing disk capacity and exploiting the energy savings of tape.

There has been a great amount of discussion about which solution is the best. There is no simple answer to that question. Each application has its own service level agreement and its own *Recovery Time Objectives (RTOs)*. An application with stringent RTOs may need disk-based backup to meet the objectives. Email systems may benefit from data deduplication solutions. Medical images that need to be stored for a long period of time may benefit from the energy and cost savings of an automated tape solution.

One-size-fits-all no longer applies to backup solutions. Choosing a high-performance, high-cost solution for all applications wastes money and wastes energy. **For many data centers, the most cost effective solution is a multi-tiered backup solution – a blending of different disk-based and tape solutions to address service level agreements and data archive and protection objectives while keeping costs in line.**

Bottom Line: It is important to consider the cost per terabyte when evaluating a storage solution. It is equally important to consider the cost of energy per terabyte per year for each solution. Choosing the right combination of solutions (i.e., implementing a tiered-storage solution) will save energy costs and it may just make our planet a nicer place in the future.



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.