Disk and Tape Storage Cost Models

Richard L. Moore, Jim D'Aoust, Robert H. McDonald and David Minor; San Diego Supercomputer Center, University of California San Diego; La Jolla, CA, USA

Abstract

The current and projected costs of storage are a critical issue as organizations face an explosive growth in data. While the cost of purchasing storage hardware is readily available from vendors, there is little published literature that describes the total cost of providing storage from an operational perspective. This paper describes current estimates of both disk and tape storage based on operational experience at the San Diego Supercomputer Center which operates a large-scale storage infrastructure. These costs include not only the storage hardware costs, but also the costs of supporting servers and related infrastructure, hardware maintenance, software licenses, floor space, utilities and labor costs. A brief discussion of projected cost trends in both disk and tape is provided, as well as a comparison to current web-based commercial storage services.

Background and Objectives

Virtually all organizations face explosive growth in their storage requirements, including exponentially growing volumes of data over increasing retention periods. Researchers at UC Berkeley estimated that 5 exabytes of data were produced in 2003 [1], while IDC recently estimated that 161 exabytes of digital information were produced in 2006 and projected nearly 1 zettabyte new data in 2010 [2]. Thus it is critical for organizations to find metrics for real cost estimates for data storage.

The San Diego Supercomputer Center (SDSC) has operated a large-scale 24*7 production data center for more than 20 years. This experience provides a comprehensive understanding of the operational costs required to provide a long-term sustainable storage infrastructure. SDSC currently operates more than 2,500 terabytes (TB) of disk storage from several different vendors including fibre-channel, SATA and MAID (Massive Array of Idle Disks) disk systems. The data volume stored in SDSC's tape-based archival system has grown exponentially with a remarkably consistent doubling rate of ~15 months, and now exceeds 5 petabytes (PB); the current capacity is 25 PB without compression.

The data infrastructure at SDSC is provided for a wide variety of applications and users, including simulation output from the national supercomputing research community, experimental and sensor data from the scientific community, and digital library collections from the Library of Congress, the National Archives and Records Administration, and others. As SDSC's storage infrastructure grows in size and evolves to support a broader set of communities and services, it is critical to develop comprehensive cost models for current and future sustainable storage.

The primary objective of this paper is to define and estimate the core elements of sustainable "bit preservation" storage costs for both disk and tape-based archival storage, based on SDSC's experience as a large-scale production facility. This includes capital investments for storage hardware and supporting infrastructure (with sustainable refresh), media (with migration), maintenance, facility costs, utilities, and the labor costs to operate large-scale disk and archival systems. This paper focuses on the "bit preservation" layer of delivering storage services; broader issues such as ingest, curation, tools, and delivery are significant, but are outside the scope of this paper.

This paper also addresses limited projections regarding future storage costs for both disk and tape; the projections are based primarily on SDSC's historical trends in the costs of capital investments and media, as well as the labor to operate the resources. Finally we briefly compare these cost estimates to the cost of storage offered by emerging commercial services.

Overview of SDSC's Production Storage Facilities

A high-level summary of the storage infrastructure at SDSC is illustrated in Figure 1 (all units in this figure represent usable space.) SDSC operates three supercomputers for the national research community (DataStar, BlueGene and an IA-64 cluster); each of these computers has a local high-performance file system built on fibre-channel disk and running the General Parallel File System (GPFS) software [3], [4]. The majority of the other disk systems are built on SATA disk. A large-scale 800 TB wide-area parallel file system (GPFS-WAN) can be mounted simultaneously by all of SDSC's supercomputers, as well as computational systems nationwide at SDSC's TeraGrid partners [5]. Nearly all of the production disks at SDSC are on a Storage Area Network (SAN) and can be accessed by multiple systems. Generally, SDSC maintains ~97-99% availability for its compute and file systems.

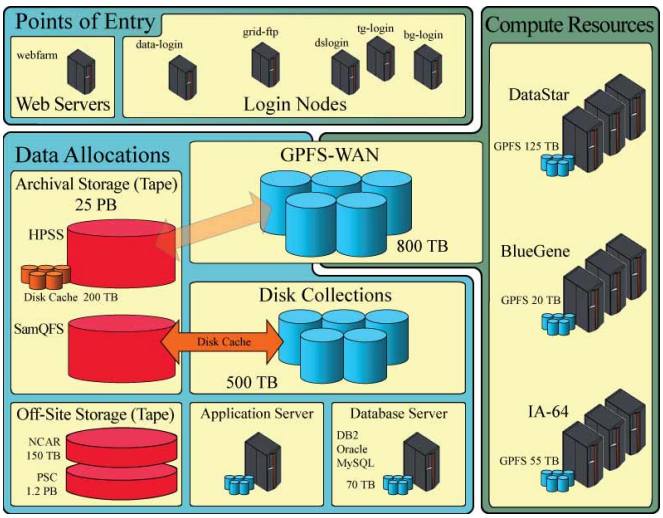

Figure 1: *Overview of Storage Infrastructure at SDSC*

SDSC's archival system consists of six silos which house ~36,000 tape cartridges. There are a total of ~110 tape drives, spanning three generations of enterprise-class tape drives. SDSC operates two archival file systems which share the same physical silos – SAM-QFS [6] and HPSS [7]. In addition to the local 25 PB capacity archive, SDSC has collaborations with the National Center for Atmospheric Research (NCAR) and the Pittsburgh Supercomputing Center (PSC) to provide archival space for geographical replication of critical files.

Data reliability requirements vary by application. Computer simulation output is valuable but generally reproducible; only single-copies of these files are normally retained. Critical experimental and collection data can be afforded higher levels of reliability via copies across different disk and archival systems as well as geographical replication at NCAR or PSC.

Cost Elements and Estimates for Disk and Tape Storage

While it is straightforward to obtain an estimate from vendors of the capital cost for purchasing a disk or tape storage system, the critical question which this paper addresses is what is the total sustainable cost of delivering that storage to users? This is a difficult question which is not typically covered in the existent literature. Most studies (for example Copeland [8], and Thompson and Best [9]) tend to focus on only one aspect of the storage system as their fulcrum for study.

The costs are based on a sustainable rate with units of $/TB/year. This cost includes amortized capital costs of the storage system itself as well as supporting infrastructure, maintenance, software licenses, facilities space, utilities, and the labor to administer and maintain the systems. Sustainability is a key issue – it is assumed that all hardware systems must be refreshed after their useful lives – *i.e.* this is not a "cold storage" model of buying disks/cartridges and simply storing the media for later use. Some users expect a $/TB cost, assuming that data can be stored once with an initial cost but virtually no ongoing costs. This is akin to a cold storage model and runs counter to the presumption of "sustainable" costs, with data being migrated to new systems/ media on an indefinite basis. As technology progresses, costs/TB/year will decline but will not be zero.

For simplicity, cost estimates in this paper are "single-copy" costs (replication would operationally be required to ensure high reliability). The disk cost estimates are based on SATA disk and the tape cost is based on the enterprise-class tape drives used at SDSC. The estimates are based on SDSC's actual aggregate costs and are normalized by the current storage that SDSC provides: ~1.8PB (raw) SATA disk deployed and ~5 PB of archival data stored. This normalization enables "cost/byte" calculations but there would be significant changes in this normalized cost if SDSC stored more or less data; the cost scaling with capacity is discussed further below.

There are several clarifications necessary in presenting data regarding the cost elements and dollar amounts. First, significant vendor discounts are often negotiated for capital purchases and maintenance; it is not unusual that the negotiated pricing is confidential and therefore some obfuscation is required. Second, there is indirect burdening included in these costs on various cost elements, and these burdens will vary by institution. Third, storage system costs are based on several large-scale purchases by SDSC over the last 18 months; there will be a wide range of system costs based on the timing, scale, and negotiations for various procurements. Fourth, there are more complex sub-issues that are not addressed in the cost estimates. For example, there are resource costs associated with each transaction (read/write); transaction costs are not differentiated here, so this averages over SDSC's usage patterns (*e.g.* much of the archival storage is "write-once-read-rarely"). In addition, the networking/bandwidth costs for users to upload/access data are not included. Transaction/access/bandwidth costs are built into cost models for many commercial services. Finally, the number/size of files has a significant impact on the infrastructure resources, but this refinement is ignored.

Table 1 lists the cost elements and associated quantitative estimates for operating SDSC's SATA disk and archival tape storage systems. The bottom lines are ~$1500/usable TB/year for disk and ~$500/TB/year for tape.

Table 1: Estimated normalized annual cost of delivering disk and tape storage at SDSC.

	SATA Disk Storage (1.8 PB)		Archival Tape Storage (5 PB)	
	$/TB/ yr	% of total	$/TB/ yr	% of total
Disk/cartridge media cost (annualized)	535	36%	100	20%
Other capital costs (annualized)	235	15%	165	33%
Maintenance & license costs	230	15%	110	22%
Facilities Costs - space, utilities	160	11%	25	5%
Maint./System Admin Labor (@ $150K/FTE)	340	23%	100	20%
Total Cost	1500		500	

In the "disk-versus-tape" debate, SDSC's integrated costs/byte are currently a factor of three lower for tape versus disk storage. (Future trends are discussed below.)

It is important to note that the raw "media" cost (spindles/arrays/controllers for disk, cartridges for archives) is only a modest percentage of the total cost of delivering sustainable storage - ~36% for disk and ~20% for tape. The other capital costs for disk include file system servers and the storage area network. For an archival system, this includes the tape libraries, tape drives (typically at least two generations for migration), archival servers and archival disk cache. Maintenance costs for all these hardware elements are included as are licensing fees for file system software. Facilities costs include an estimate of the "floor space" cost of hosting equipment in a machine room as well as the utilities for the hardware. Finally, we estimate normalized labor costs for maintaining and administering the storage systems. At present, SDSC has approximately three full-time staff dedicated to maintaining the disk file systems, and another three for the archival system (including both HPSS and SAM-QFS).

As an "at-scale" facility, SDSC has the benefits of amortizing costs over a large base and negotiating substantial vendor discounts for large volume purchases. But even for an "at-scale" facility, there are economies of scale that lower the storage cost/byte as the infrastructure grows. The key is to determine which costs scale directly with size and which are fixed or scale only weakly with size, thus producing further economies of scale. The "media" capital costs and their maintenance scale linearly with size, as do facilities costs. However the software license costs are generally fixed. Some of the supporting infrastructure (servers) could handle larger file systems, and many of the supporting infrastructure costs scale weakly with size or are fixed up until some "threshold" at which there is a discrete jump in the associated costs. Labor costs are non-linear. For a large 24*7 production facility, there is a minimal level of at least 2-3 staff that must be trained on the various disk or archival storage systems, so that there is a reasonable on-call rotation and backup. This level of staff can handle a large system but at some point, additional staff must be added. Another factor in this scaling is that for disk, the costs are normalized by the amount of disk deployed while for tape, the costs are normalized by actual data stored (there are pros and cons about which is the most appropriate normalization). But if SDSC's archival storage volume was to double in 15 months (as projected), most non-media costs would stay fairly constant and the archival storage cost/byte estimate would decline.

Projections of Future Storage Costs

While predicting the future is no small task and generally far beyond the authors' expertise, the recent Gantz study [2] reinforces that storage requirements will continue to rise exponentially; thus it is critical for vendors and storage providers to continue to reduce the cost/byte to balance the dramatic rise in volume.

There is considerable debate about whether tape storage will be (or has been!) eclipsed by disk storage. As noted above, the current estimate is that the difference in SDSC's total cost to deliver tape and disk storage is a factor of about three. The differential between the cost of tape and disk media has narrowed over the years, and is narrow enough now to merit discussions of other factors such as access latency, bandwidth and operational factors, even for applications which do not require immediate on-line access (e.g. backups). To evaluate trends, SDSC has analyzed the cost/byte from its tape media purchases over the last 20 years and the long-term trend is an exponential reduction with about a three-year halving time. The corresponding data are unfortunately not as readily available for disk purchases, but the cost/byte halving time is certainly shorter for disk media than for tape. Vendors need to provide projections of future technologies and costs, but historical projections would indicate a cross-over in media cost/byte in the foreseeable future.

But how will the technology-driven reduction in media cost/byte impact the total operational cost of delivering that storage? This question applies to both tape and disk storage, and is critical to the total storage costs institutions face as data volumes grow exponentially. For example, as shown above, the cost of the disk media represents only ~35% of SDSC's cost for delivering disk storage, and the percentage is only ~20% for tape. Fortunately most of the cost elements contained within the total operational costs scale in some fashion with the media cost and the near-term costs/byte are expected to roughly follow the trends in media costs. For example, with Moore's law the annualized cost of supporting servers for either disk or tape systems stays roughly constant (including maintenance); similarly the annualized cost of archival silos and tape drives stays fixed in real dollars, and therefore the cost/byte follows media trends. Utility and machine room costs are roughly fixed for archival silos and tape drives as technology progresses; similarly most of the advances in disk media cost/byte are achieved with increasing spindle capacity, yet spindles have roughly constant power and space requirements. In addition, vendors are improving the spindle density within racks and also improving the power consumption per spindle, including significant reductions using MAID technology. SDSC's experience is that there are not immediately noticeable increases in the hardware system administration labor costs as subsequent generations of disk or tape systems are deployed; if this continues, then labor costs will also scale with media costs/byte. (While the hardware administration time may scale, a more likely limiting factor, not considered in this paper, would be the labor time required to effectively manage and utilize the exponentially growing volume of data and sheer number of files!)

If media costs continue to drop exponentially, but some other operational cost element does not scale as quickly, that cost element will quickly dominate the cost/byte for delivering storage. Whether this is the labor cost, utilities or some other factor has yet to be determined but that will become the key element for which to focus efficiency improvements.

In Morris and Truskowski's [10] evolution of storage systems one can see that it is still too early to write the comprehensive history of what we know as our current storage methodologies. And there are certain to be disruptive storage technologies (*e.g.* holographic storage) that will change the landscape and enable continued advances in the cost of storage.

It is clear to us that more information and studies about the total cost of delivering storage, especially from at-scale data centers that operate by and for the public good, are critical to improving storage practices and costs. A clear sign that progress is being made in this area of cooperative exchange of information are the recent calls by Gibson and Schroeder [11] and Weinstock [12] for more large-scale data sets on disk storage failure. SDSC would like to make a call to all interested parties to examine and share information about production level storage costs on a similar basis.

Comparison with Commercial Services

A number of commercial web-based services such as Amazon S3 (aws.amazon.com/s3), MozyPro (www.mozypro.com) and OmniDrive (www.omnidrive.com) provide distributed data storage services. It is interesting to compare the storage costs estimated here to these services, although inevitably there are apples-to-oranges elements to the comparisons, including significant variations in services and pricing models.

For example, Amazon S3 offers storage for ~$1850/TB/yr with a transmission (access) charge of $205/TB. For "write-once-read-rarely" storage, this is cost-effective storage; for data which are frequently accessed the cost can become quite high (*e.g.* $4200/TB/year for once/month access). The S3 architecture is not specified, but presumably the data are all stored on disk with some level of replication. Replication and access costs are critical in comparing these commercial services to the SDSC storage cost

estimates. Two disk copies at SDSC would be ~$3000/TB/year while one disk/one tape copy would be ~$2000/TB/year. In addition, at present the transaction/bandwidth charges are not addressed in the SDSC cost estimates. SDSC has access to cost-effective high-speed educational/research networks and bandwidth is not a significant cost element at current usage rates.

Conclusions

There are several key conclusions from this study. First, current estimates of the total "bit preservation" cost of storage are ~$1500/TB/yr for SATA disk and $500/TB/yr for enterprise-class tape archives; thus the current difference between tape and disk costs is a factor of about three. Second concentrating solely on the cost of the storage media, whether disk or tape, provides only part of the cost of delivering that storage to users. For SDSC the media accounts less than a third of the total cost of delivering storage. This is critical to consider when anticipating the true costs of building or expanding a storage facility. Third, each element in the cost/byte equation must be evaluated individually for its scaling dependencies; these scaling factors are critical in estimating both the economies of scale as a storage infrastructure expands and the cost reductions with future technology advances, particularly in media costs. Fourth, based on a projection of historical trends, the differential between the cost of delivering disk and tape storage is likely to diminish in the foreseeable future; actual trends will depend on vendor technology roadmaps and costs. Finally, the cost estimates here are "in the ballpark" with web-based distributed commercial storage services currently being offered, although there is a wide range of services and cost models amongst these services and SDSC's cost estimates.

References

[1] P. Lyman and H.R. Varian, "How Much Information," Retrieved from http://www.sims.berkeley.edu/how-much-info-2003 on March 11, 2007. (2003).

[2] J.F. Gantz et. al., "The Expanding Digital Universe: A Forecast of Worldwide Information Growth through 2010," IDC Whitepaper, Retrieved from http://www.emc.com/about/destination/digital_universe/pdf/Expanding_Digital_Universe_IDC_WhitePaper_022507.pdf on March 11, 2007. (2007).

[3] F. Schmuck and R. Haskin, "GPFS: A Shared-Disk File System for Large Computing Clusters," Proc. of the 1st USENIX Conf. on File and Storage Technologies, pg. 19. (2002).

[4] C. Jordan, "Lessons Learned from the Terragrid, Part 2: Managing a Big Data Set on a Big Grid," IBM Developer Works, Retrieved from http://www-128.ibm.com/developerworks/grid/library/gr-teragrid2/index.html on March 11, 2007. (2006).

[5] P. Andrews, P. Kovatch, and C. Jordan, "Massive High Performance Global File Systems for Grid Computing," Proc. of the 2005 ACM/IEEE Conf. on Supercomputing, pg. 53. (2005).

[6] B. Banister and J. Bowen, "High Performance QFS with Solid State Disk Metadata Storage," Whitepaper, Retrieved from http://www.texmemsys.com/files/f000206.pdf on March 11, 2007. (2006).

[7] P. Andrews, T. Sherwin, and B. Banister, "Large Scale Flexible Storage with SAN Technology," Proc. of the 18th IEEE Symp. on Mass Storage Systems and Technology, pg. 291. (2001).

[8] G. Copeland, "What If Mass Storage Were Free?" Proceedings of the Fifth Workshop on Computer Architecture For Non-Numeric Processing, Retrieved from http://doi.acm.org/10.1145/800083.802685 on March 11, 2007. (1980).

[9] D.A. Thompson and J.S. Best, "The Future of Magnetic Data Storage Technology," IBM Journal of Research and Development. 44(3). Retrieved from http://www.research.ibm.com/journal/rd/443/thompson.html on March 11, 2007. (2000).

[10] R.J.T. Morris and B. Truskowski, "The Evolution of Storage Systems," IBM Systems Journal 42(2), Retrieved from http://www.research.ibm.com/journal/sj/422/morris.html on March 11, 2007. (2003).

[11] B. Schroeder and G. Gibson, "The Computer Failure Data Repository (CDFS)," Proc. of the Symp. on Reliability Analysis of System Failure Data, Retrieved from http://www.deeds.informatik.tu-darmstadt.de/RAF07/papers/bianca_schroeder.pdf on March 11, 2007. (2007).

[12] C. Weinstock, Z. Dubrow, and R. Stoddard, "Development and Analysis of a Software Failure Data Repository," Proc. of the Symp. on Reliability Analysis of System Failure Data, Retrieved from http://www.deeds.informatik.tu-darmstadt.de/RAF07/papers/charles_weinstock.pdf on March 11, 2007. (2007).

Author Biographies

Richard Moore received his BS degree from the University of Michigan in 1975 and his PhD in Astronomy from the University of Arizona in 1980. After a postdoctoral position at Caltech, he led aerospace research teams at the Aerospace Corporation and then Photon Research Associates. He joined the San Diego Supercomputer Center in 2002 and is the Director of the Production Systems Division, which operates the center's supercomputers, storage and networking systems.

Jim D'Aoust received his BS degree in Nuclear Engineering from the University of California Santa Barbara, and an MS degree in Engineering from Stanford University. He is currently a Project Manager at SDSC and recently implemented major procurements to upgrade SDSC's disk and archival storage infrastructure.

Robert H. McDonald received his BM and MM degrees from the University of Georgia and his MLIS from the University of South Carolina. Prior to his position as Project Manager of the Chronopolis Digital Preservation Repository at SDSC he led technology programs in the research libraries of Auburn University and Florida State University.

David Minor received his BA degree from Carleton College and his MLIS from the University of Wisconsin-Madison. He is currently the Project Manager for the Library of Congress Pilot Project with SDSC. He has worked previously for libraries at the Universities of New Mexico and Wisconsin as well as Penn State University. He has also held several systems administration positions in the commercial sector.