Active Archival Storage

A Cost of Ownership Analysis

By Brian Garrett
Analyst, Enterprise Strategy Group

August 2005

Executive Summary

The technology choice for implementing deep and active digital archives has historically been between tape and optical. Content addressed systems that are based on commodity disk drives have recently become available as an alternative. The Enterprise Strategy Group (ESG) was contracted by Plasmon, a leading optical archiving solutions provider, to perform a third party analysis of the cost of acquisition and ownership of each of these archival technologies – tape, optical, and disk.

This analysis is based on the actual case study of a financial services firm shopping for a 12 TB archive. The financial services firm needed a solution to archive 8 GB of new data and handle 2,500 queries daily. The cost of acquisition and ownership was measured over 3 years of operation. Only clearly quantifiable list prices were included in this study to avoid the effects of subjective interpretation.

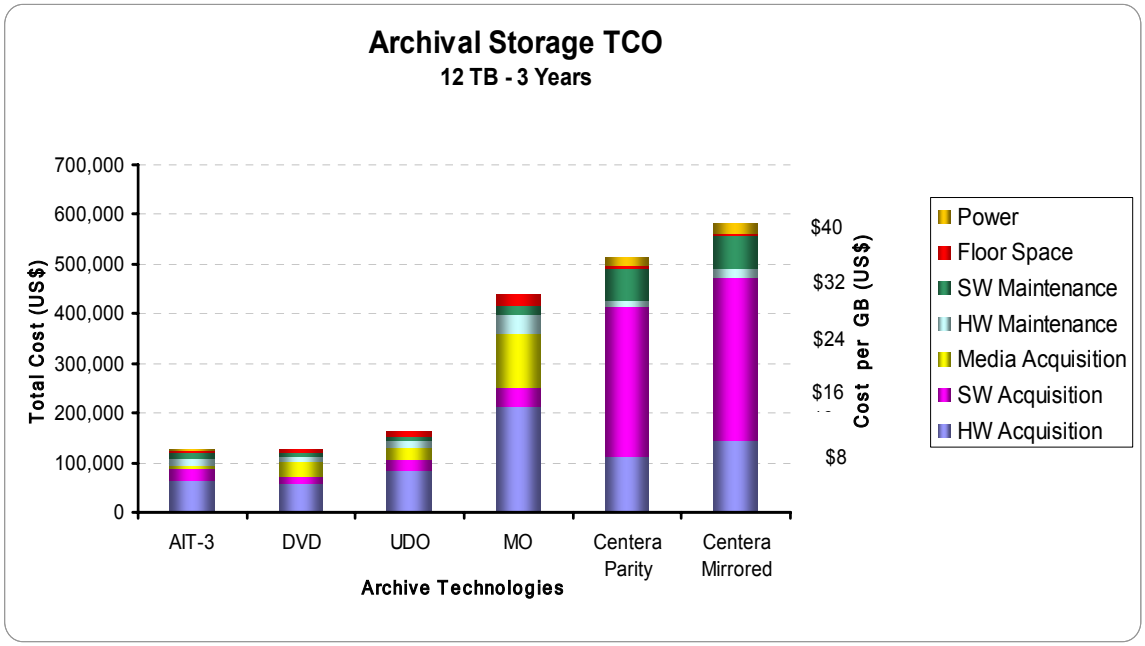

Figure One: ESG Archival Cost Analysis

The results, as shown in Figure One, *ESG Archival Cost Analysis,* clearly indicate that the cost of a Plasmon G438 library populated with Ultra Dense Optical (UDO) technology is competitive with AIT tape and DVD optical technology and represents a fraction of the cost of a Centera disk or MO optical solution[1]. **As a matter of fact, our analysis indicates that 12 TB of parity protected Centera capacity is 361% more expensive than a comparably configured automated UDO optical library.** A closer look at the results reveals that the cost of a Centera solution is inflated significantly due to the price of software acquisition and maintenance, while the cost of MO optical is burdened by high hardware acquisition and media costs. A more careful examination uncovers noticeable power consumption costs for the disk-based Centera system compared to tape and optical.

Proponents of the Centera architecture might object to the conclusion that optical technologies like UDO are significantly more cost effective than a disk-based Centera system. They would argue that the self-healing and scalable Centera architecture reduces the cost of administration when compared to

[1] The methodology and data behind this diagram are documented later in this report and in the Appendix.

multiple tape or optical libraries. That argument is explored in Figure Two, *Scalable UDO vs. Centera Analysis*.

The diagram compares a 16 node parity protected Centera with 22.4 TB of capacity to a fully loaded 19.1 TB Plasmon model G638 UDO library. Pairs of sixteen node Centera clusters and UDO libraries are then added to scale near-line capacity to 150 TB. The total cost per Gigabyte presented above was used to calculate the costs depicted on the Y-axis in millions of US dollars. The cost per Gigabyte derived during the single library G438 UDO library analysis presented earlier was re-used as a conservative approximation during this multi-library comparison based on larger capacity G638 libraries. A more rigorous examination based on the cost of fully populated G638 libraries would show the cost of UDO acquisition to be reduced by approximately 20%.

Let's focus first on the set of data points at the bottom left of the diagram, where we've already seen that an automated UDO optical library is about 1/3rd the cost of a parity protected Centera. Since all of the media for the automated UDO solution is inside the library, there is no need for a system administrator to handle removable media. The EMC Centera is known for its ease of administration. However, customers may find it difficult to justify the $698,638 difference due to reduced administration costs.[2]

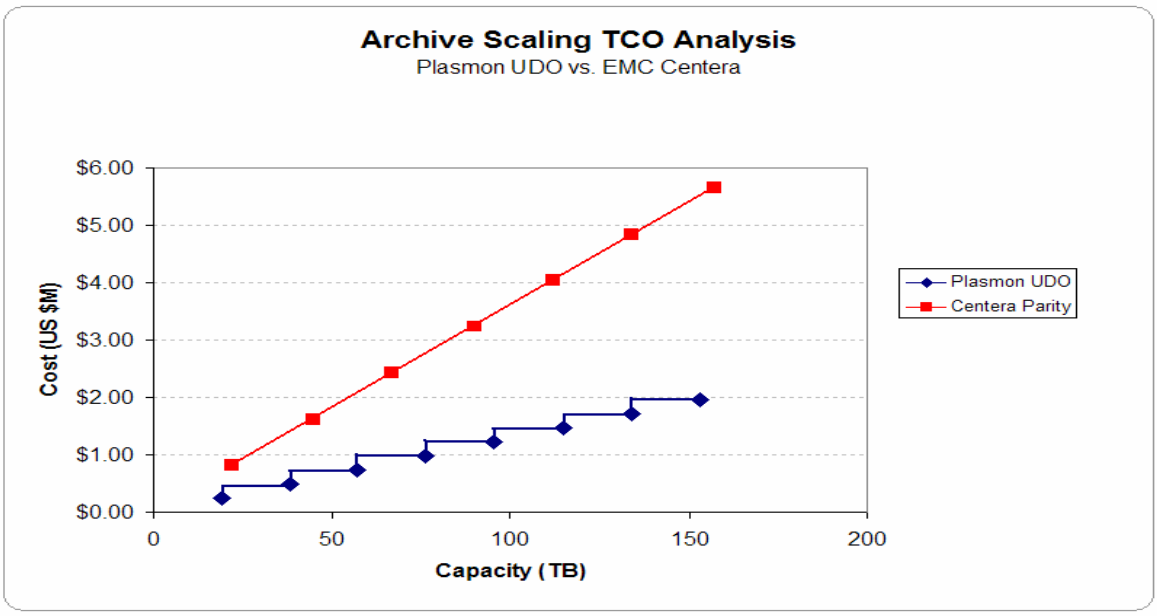

Figure Two: Scalable UDO vs. Centera Analysis

Now consider the cost of a 150 TB capacity point as shown toward the upper right of the diagram, where the cost of eight optical libraries is compared to a 224 node Centera cluster. The Centera upgrade path from 22 TB to 157 TB is depicted as a smooth red-line to depict the fact that nodes can be added to a singly managed pool of storage as needed. Although removable capacity can be added within an optical library for additional capacity as needed, we depict the optical solution as a step-wise series of fully populated library additions. The cost difference at 150 TB is $3.6M. **Conceding the fact that managing eight optical libraries is more difficult than managing a single Centera system, ESG finds it difficult to believe that the majority of end users would incur $3.6M of additional system administration costs over three years for an eight library UDO optical archiving solution.**

[2] It should be noted that while this type of analysis holds true for near-line archiving applications with all media maintained within an automated solution, it does not apply when some portion of the archive resides outside of a library on removable media. Caution should also be used when trying to apply this analysis to all optical technologies. As shown in this report, UDO and DVD are clearly cost competitive with Centera, while legacy MO optical technology is closer to cost neutral.

ESG Research

End user research performed by ESG indicates that cost is a major concern of storage professionals considering a move to disk-based archival. ESG conducted a survey of 163 North American storage professionals and IT managers in November 2004. The majority (58%) of respondents responsible for 10 TB or more of primary storage fell into a "fast follower" camp, indicating that they would consider replacing some or all of their enterprise-class tape libraries with a disk-based archive, but have yet to take any action to do so.

Nearly half of the fast followers (44%) say they would consider a disk-based solution to support on-line or near-line archiving. These fast followers who are motivated by compliance and corporate governance initiatives are looking for a new cost-effective tier of near-line storage between on-line disk and off-line tape. As shown in the following diagram, the vast majority of fast followers (77%) indicate that the cost of new disk-based solutions is the primary objection. It should also be noted that fast followers are also concerned about the reliability of ATA drives, the lack of media portability and the ability to ensure regulatory compliance due to the absence of Write Once Read Many (WORM) media – all of which are not a concern when DVD and UDO optical technology are used for near-line archival.

> "We looked at a variety of archiving technologies to replace DVD jukeboxes that were multiplying like rabbits. We found the disk-based EMC Centera to be too expensive. Our fully automated Plasmon UDO library costs three to four times less than a Centera from EMC and is eight times faster than our old DVD jukeboxes."
>
> - IT Director, Major US Newspaper

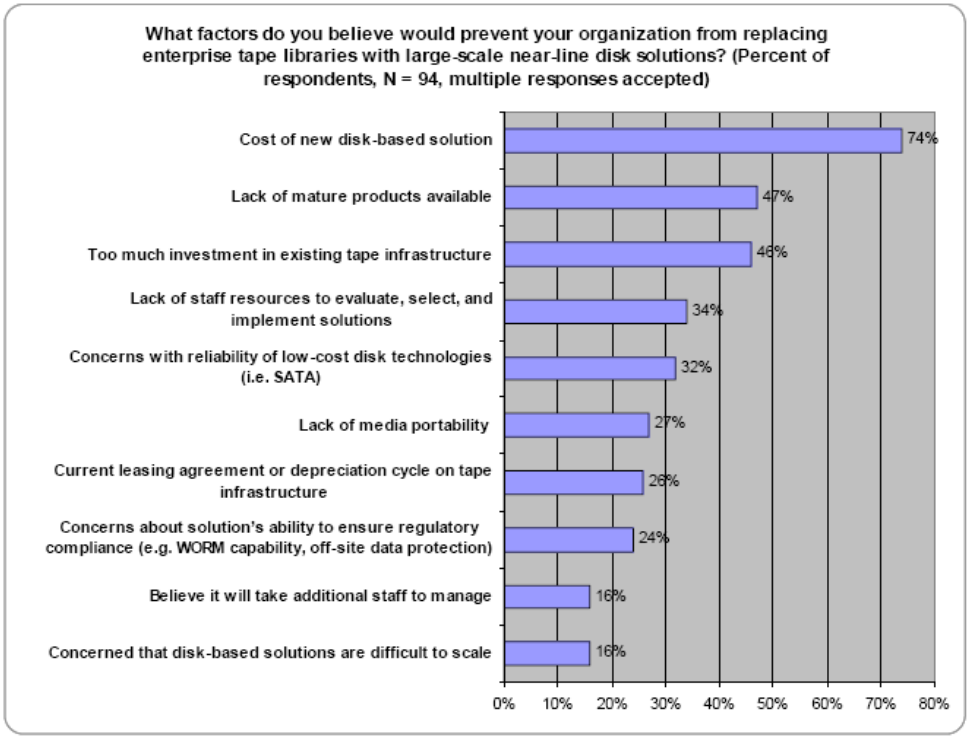

Figure Three: ESG Research

Technology and Cost

There are significant differences between the CD and DVD technology we use in our homes and the professional DVD, UDO, and MO optical technology analyzed in this report. The first and most obvious difference is automation. Professional optical libraries can be packed full of robotically controlled media, support multiple drives which act in parallel for improved performance, and come in sizes as big as a refrigerator. For example, Plasmon optical libraries (Figure Four, *Plasmon G-series Libraries*), with capacities in excess of 19 TB, support advanced features including dual pickers, barcode readers, hot/warm swappable drives, redundant power supplies, and very high duty cycles.

Figure Four: Plasmon G-series Libraries

Figure Five: UDO Cartridges

Another major difference between professional optical and the technology we use at home is how the media is packaged and handled. DVD optical discs arrive from the factory looking much like what we use in our homes and are loaded into magazines which are then inserted into a library. Magazines make media handling easier and reduce the risk of data loss due to scratches and contamination. MO and UDO optical discs are encased in a 5.25" cartridge that keeps the media individually shielded (Figure Five, *UDO Cartridges*). The cartridge packaging of MO and UDO media virtually eliminates the risk of scratches and contamination experienced with DVD.

UDO uses the latest blue laser technology which delivers 30 GB of capacity per disc, compared to legacy red laser MO with 9.1 GB, and DVD with 9.4 GB of capacity per disc. The MO optical solution as configured for this analysis is significantly more expensive than the DVD and UDO solutions. This is due to the high cost of lower capacity legacy media and a need for two libraries to get close to 12 TB of capacity (compared to one library for DVD and UDO).

MO optical can be erased and rewritten much like the CD-RW discs used in PC's to make personal backups. In contrast, UDO and DVD drives use Write-Once-Read-Many (WORM) media that can not be erased or overwritten. Although software and processes can be used to make non-WORM technologies like MO optical and hard disks behave like WORM media, UDO and DVD with native WORM capability are a better choice due to the reduced cost of the processes and controls needed to secure and audit compliance-mandated information assets. Although both DVD and UDO use a phase change approach to actually write WORM data, UDO's eight layer media stack is considered much more robust. UDO media is also available in a rewritable format, which is based upon physically different media materials and structure, and is supported in UDO drives and libraries either with or without Write Once media.

AIT tape technology has price/performance characteristics and WORM support which make it well suited for active archiving when compared to most tape technologies. AIT tapes can be unloaded without being rewound, which provides fast random access times. Fast random access times make it possible to provide predictable performance using a small number of drives servicing many archive read requests. The LTO tape format was considered and dismissed due to poor price/performance in comparison to AIT[3]. Un-compressed 100 GB AIT-3 tape cartridge capacity was used for this analysis, based on the assumption that data stored by archiving applications is typically compressed before it is written to an archive[4].

The disk technology chosen for this analysis was the EMC Centera. The Centera was chosen because it is a market leading disk-based archive system that is specifically designed to meet the WORM functionality, management, and authenticity requirements of compliance-mandated archive applications. EMC Centera software costs include mandatory CentraStor software and the CentraStor CompliancePlus Option, which is needed in mandated environments that require auditable WORM attributes. Two Centera configurations were included in the analysis, one with parity for protection from a hard disk failure (CPP) and the other with mirroring (CPM). The parity protected configuration, which is the most popular, requires fewer disk drives and therefore costs less than a mirrored configuration with less chance of data loss due to a hard drive failure.

Since each technology option could not be configured to exactly meet the 12 TB requirement, fully populated configurations that came closest to 12 TB were used. A price per GB of capacity was then calculated for each configuration and multiplied up to 12 TB for a rational cost comparison. The number of tape and optical drives were determined by matching publicly available performance specifications with the need to sustain 8 GB of write data and 2,500 requests daily[5].

The cost of QStar HSM software was included in this analysis. QStar HSM is an enterprise-class hierarchical storage management software package that supports tape, optical, and Centera hardware. Although there are a large number of software packages that can be used for a near-line archival cost analysis, the choice of archival software matters little in this analysis since the cost of archival software, and the server(s) it runs on, should be the same regardless of the hardware technology chosen for archival. Archiving software solutions that have not been ported to the Centera programming interface require the use of CUA servers and software so that a Centera can be accessed using standard network file system protocols. Because QStar HSM supports the Centera programming interface, the cost of Centera Universal Access (CUA) hardware and software was not included in this analysis.

The cost of commercial power at $0.07 per Kilowatt, floor space at $3,235 per square meter, and cooling at 40% of the cost of power were included in this analysis. These rates, which are typical for a large US or European city, would need to be adjusted higher for areas like California and New York City. The cost of power for the Centera system, which ranged between $5,000 and $6,800 a year, was noticeable compared to the automated libraries, which averaged $350 a year. The relative cost of floor space for all configurations was negligible, although a 12TB Centera configuration approaches a ton in weight, which is an order of magnitude higher than a UDO library.

Although the cost of disaster avoidance was not included in this analysis, media survivability and remote vaulting are issues that should be considered when implementing a long term digital archive. Data residing on optical media has a better chance of surviving a disaster (e.g. an earthquake or flood) than tape or disk. Removable tape and optical media are easy to replicate and transport to a remote site for safe keeping. Replicating a disk-based Centera system is also easy, but can be quite expensive due to the cost of another set of disk drives, EMC software at the remote site, and the recurring cost of WAN bandwidth.

---

[3] The list price of an ADIC LTO-2 automated tape library and media is more than twice that of the AIT-3 solution used in this analysis.

[4] A compression rate of 2:1, which would cut the media cost in half, would not noticeably change the results of this cost analysis, due to the fact that AIT media costs contribute relatively little to the overall cost of a 12 TB automated AIT-3 solution.

[5] Details are available in the Appendix.

And finally, the cost of media maintenance should be considered when deploying a digital archive in which data will reside for many years. Tape and disk technology have inherent media maintenance costs (e.g. tape re-tensioning and disk migration) that can inflate the cost of digital archiving as compared to optical media, with a typical specified shelf-life of 50 years.

The Bottom Line

Compliance regulations, corporate governance initiatives, and the explosive growth of unstructured digital content are driving the adoption of archiving strategies and technologies. Organizations supporting business needs with on-line disk technologies are evaluating the deployment of a new tier of cost effective near-line capacity for long term data archival. ESG Research has shown that the high cost of emerging near-line disk-based systems is a major concern for storage professionals and IT managers.

The true cost of acquisition and ownership of tape and optical solutions, compared to systems based on commodity disk drives, has been widely debated in the industry recently. Common mistakes made during such a comparison include a myopic focus on the cost of the raw media and an inflated estimate of system administration costs. Despite the fact that some vendors suggest that tape and library configurations require dramatically higher administration costs, the subjective costs associated with system administration were not included in this study. The Enterprise Strategy Group stands behind this assumption based on customer feedback that confirms the fact that modern tape and optical libraries can be configured with massive amounts of near-line archival capacity, which eliminates the system administration costs associated with removable media handling.

The intent of this analysis is to provide readers with a reasonable starting point for the comparison of the cost of archival technologies available on the market today. Although a variety of hard and soft costs were not addressed in this analysis, ESG believes that the methodology and results presented in this report form a valid relative comparison of the most significant costs of archival storage ownership. ESG encourages readers of this report who are considering an active archiving solution to perform their own cost of ownership analysis. We are confident that such an analysis will make a compelling case for the consideration of automated professional optical technology.

Appendix

Table One: 12 TB Adjusted Cost Summary

Archive Type	System Capacity	Actual System Cost	$/GB	Adjusted 12 TB Cost
AIT-3 Library	9.6 TB	$100,328	$10.45	$125,410
DVD Library	13.8 TB	$144,467	$10.47	$125,624
UDO Library	13.1 TB	$176,503	$13.47	$161,682
MO Library	11.6 TB	$423,518	$36.51	$438,122
Centera CPP	11.2 TB	$476,934	$42.58	$511,001
Centera CPM	12.5 TB	$605,148	$48.41	$580,942

Table Two: System Cost Breakdown

Archive Type	HW$	SW$	Media$	HW Maint$	SW Maint$[6]	Floor Space$	Power$	Total$
AIT-3 Library	49,252	19,500	5,280	11,685	10,004	3,386	1,221	$100,328
DVD Library	65,700	15,400	36,875	10,066	7,900	7,643	883	$144,467
UDO Library	91,124	22,900	26,280	15,755	11,771	7,643	1,030	$176,503
MO Library	205,720	38,200	102,080	37,382	19,597	18,479	2,060	$423,518
Centera CPP	104,700	280,500	0	12,606	59,582	5,418	14,128	$476,934
Centera CPM	150,100	341,000	0	18,909	68,529	5,418	21,192	$605,148

Table Three: Drives and Media

Media Type	Vendor	Product	Drive Count	Media/ Drive Capacity	Media Count	Usable System Capacity
Tape (AIT-3)	ADIC[3]	Scalar 100	8	100GB	96	9.6TB
Disk	EMC	Centera Parity	64	320GB	-	11.2TB
Disk	EMC	Centera Mirrored	96	320GB	-	12.5TB
DVD	Plasmon	D1525	6	9.4GB	1,475	13.8TB
MO	Plasmon	G638 x 2	4 (x2)	9.1GB	1,276	11.6TB
UDO	Plasmon	G438	4	30GB	438	13.1TB

[6] This column indicates the cost of QSTAR HSM software maintenance plus the cost of Centera software maintenance if applicable.

Table Four: Optical and Tape Drive Specifications

Drive / Library Specs.	DVD	MO, UDO	AIT-3
Load Time	15 sec	5 sec	10 sec
Unload Time	3 sec	3 sec	10 sec
Average Seek Time	200 msec	35 - 50 msec	27 sec
Average Rewind Time	0 sec	0 sec	12 sec
Media Exchange Time	6 sec	6 sec	6 sec
Average Data Access	5 sec	5 sec	5 sec
Average Access Cycle	29 sec	19 sec	70 sec
Access Cycles per Hour	124 cycles	189 cycles	51 cycles
Drive Count for Read	3 drives	2 drives	6 drives

Table Five: Optical and Tape Drive Counts

Drive Type	Read Drive Count	Write Drive Count	Spare Drive Count	Total Drive Count
AIT-3	6	1	1	8
DVD	3	1	2	6
MO	2	1	1	4
UDO	2	1	1	4

Table Six: Environmental Factors

Archive Type	Power Watts	BTU/hr	Power $/hr	Cooling $/hr	Total $/hr	Total $/yr	Total 3 years	Adjusted 3 years
AIT-3	415	1,418	0.0291	0.0174	0.0465	407	1,221	1,527
DVD	300	1,025	0.0210	0.0126	0.0336	294	883	768
UDO	350	1,196	0.0245	0.0147	0.0392	343	1,030	944
MO	700	2,392	0.0490	0.0294	0.0784	687	2,060	2,131
Centera CPP	4,800	16,400	0.3360	0.2016	0.5376	4,709	14,128	15,137
Centera CPM	7,200	24,600	0.5040	0.3024	0.8064	7,064	21,192	20,345