Real-Time Processing and Archiving Strategies

Isabel Schnoor
ITS Public Sector
IBM Germany
Hamburg, Germany
schnoor@de.ibm.com

Abstract— **This paper will discuss how real-time processing is affecting archiving strategies. This puts up the thesis, what type of system is needed to ensure write & retrieval of real-time data from an archive infrastructure. The paper will not focus on compliance, rather on archiving strategies in context of system requirements and availabilities for real-time data applications and hence where real-time processing is applicable.**

Keywords- Archive; Real-time; long-term archiving; HSM; HPSS; real-time compression

I. INTRODUCTION

Today due to huge data growth, more digitization needs, mobile devices producing more data, faster supercomputer power, and applications, end user expect to store and archive their data in real-time and have continuously real-time access to that data. Keeping up with this fast growing demand of user behavior of "self-service" real-time data access, common and established IT-infrastructures are put under stress. Storing everything on disks becomes to large to backup, leading to long back-up windows, and increasing energy cost for more power consumption of disks systems. Especially, if needed to store data at least for 10 years or longer.

Ideal to store large amounts of data long-term are tape technologies. Due to the availability of open standards with Linear Tape-Open (LTO) as magnetic tape data storage technology, originally developed in the late 1990s, it is accepted to archive long-term data. Tape-based archival systems suffer from poor random access performance, which prevents the use of inter-media redundancy techniques and auditing, and requires the preservation of legacy hardware [1]. Many disk-based systems are ill-suited for long-term storage because their high energy demands and management requirements make them cost-ineffective for archival purposes [1].

However, combinations of disk and tape storage infrastructures do not fulfill the requirement of fast write and retrieval access to data at adequate speed, as today end user require "real-time" processing and availability. De-duplication is reducing the amount of data even further, because de-duplication technology is now common place in the backup market [2]. But backup is not the same as archiving. An archive is a collection of computer files that have been packaged together for backup, to transport to some other location, for saving away from the computer so that more hard disk storage can be made available, or for some other purpose [3]. An archive can include a simple list of files or files organized under a directory or catalog structure (depending on how a particular program supports archiving) [3].

Section two will first define real-time processing and archives as well as the terminus data transfer rate (DTR). It will position, what possible storage solutions could be used for high volume and near real-time or real-time processing applications. As software for hierarchical storage management for very large data volumes, High Performance Storage System (HPSS) is described. Section three will evaluate how HPSS is being used for real-time applications. Section four closes with a short summary and future outlook of HPSS and other storage solution for archiving.

This paper will position archiving strategies to real-time processing needs and discuss how in a context of an extreme scale archive environment, e.g., HPSS, is applicable for real-time applications.

II. ARCHIVING STRATEGIES FOR REAL-TIME APPLICATIONS

Real-time processing is defined according to Wikipedia as follows: "Each transaction in real-time processing is unique. It is not part of a group of transactions, even though those transactions are processed in the same manner. Transactions in real-time processing are stand-alone both in the entry to the system and also in the handling of output" [4].

Archives can be differentiated by types of deployment of storage components, i.e., Disks, Tape, Hierarchical Storage Management Software, or Appliances. All of these components can be positioned in entry, midrange or enterprise storage systems, and being enriched with so-called "Storage Optimizers", i.e., SAN Volume Controller (SVC), Easy Tier, or IBM Real-Time Compression Appliances. Archiving strategies are rather defined by their usage requirements either for compliance need, regular batch processing backup and archive (i.e., for databases, applications systems like SAP, Oracle, etc.) or for long-term archiving, including clustered HSM file repositories. One of the available HSM clustered file repositories is High Performance Storage System (HPSS) for extreme scale environments.

 ISBN: 978-1-61208-161-8

Figure 1. Positioning of IBM Archiving Solutions, IBM Corp. / IBM Germany, J.A. Kerr, O.J. Knopf, and I. Schnoor

Requirement	nSeries	IA	TSM	SONAS/GPFS	HPSS
1. Customer need	NAS Appliance	Compliance Appliance	Backup & File Repository	Extrem Scale NAS Appliance	Extreme Scale HSM (incl. GPFS Support)
2. Maximum Capacity	up to 2PB	up to 304TB	up to 3PB	up to 14PB	Extreme (over 100 PB)
3. Maximum Bandwith	limited	limited	limited	Extreme	Extreme
	~ max. I/O of a single server	~ max. I/O of Server (up to 3 within appliance)	~ max. I/O of a single server	900 MB/s per node; up to 30 nodes	1000(s) MB/s per mover; over 100 movers
4. Availability					
- No Single Point of Failure	Standard	Standard	Available	Standard	Available
- High Availability	Available	Available	Available	Available	Available
- Data Mirroring	Available	Available	Available	Available	Available
5. Security					
- Authentification	yes	yes	yes	yes	UNIX, Kerberos
- Autorisation	yes	yes	yes	yes	UNIX, DES
- Compliance	SnapLock	SSAM/WORM	SSAM/WORM	no	no
6. Interfaces	CIFS, NFS	TSM-API, NFS	TSM-API	CIFS, NFS, SFTP	FTP, PFTP, HPSS-API, RHEL-VFS (CIFS, NFS, SFTP, HTTP), GPFS, div. 3rd Party
7. Storage Media	Disk	Disk, Tape, Optical	Disk, Tape, Optical	Disk, Tape and Optical via TSM only	Disk, Tape
8. Storage Hierachies	no	yes	yes	yes	yes
9. Hardware Independent	no	no	yes	no	yes
10. Storage Solutions OS	ONTAP	RHEL	multiple	RHEL	RHEL, AIX
11. Client OS	multiple	multiple	multiple	multiple	multiple

IBM has a broad range of storage systems available, addressing various needs for storage infrastructures, i.e., for open systems and optimized for z/OS and System i, for block-based or file-based storage infrastructures. The strategy of the IBM System Storage products is aligned to the need of delivering the right system by "fit for purpose" for any workload. For archiving needs, the figure 1 shows a table, outlining by eleven requirements, the possible archive solutions from IBM. The following products and solutions are: IBM NSeries, IBM Information Archive (IA), IBM Tivoli Storage Manager and Tivoli Storage Manager for Space Management (TSM/HSM), Scale-out Network Attached Storage with IBM General Parallel File System (SONAS/GPFS), and HPSS. According to type of archiving need and scenario, figure 1 describes what storage solution is applicable and possible to select.

Researchers get the brightest ideas and with better equipment to analyze, research, discover, and exploit, the possibilities of new ways of developing fundamental experiments or create new simulations is large. Due to this type of research, high amounts of volumes of data are generated and need to be stored and archived in large file systems. Real-time applications are common in weather prediction, measuring seismic activity, applications for simulations for the new development of airplanes, in air traffic control systems, or in rail way switching systems etc. Perhaps, in future due to new compression mechanism, i.e., IBM Random Access Compression Engine (RACE) can reduce the Terabytes (TB) managed or even Petabytes (PB) in archives. Hence, making it easier, that even NAS environments, become applicable for long-term archiving. But for certain industries, i.e., for basic research organizations or governmental institutions, the need is clear, according to sources, to grow in data by far more than 60 PB by 2016 [5]. Real-time compression may come in use for unstructured files and in hence in use for NAS environments. Relating this to archives, real-time compression in primary storage may help to reduce the amount of data to be archived, i.e., of databases, IBM Lotus Notes®, Text, CAD, or VMware VMDK files [6], with a compression rate up to 80 percent.

It is of interest that from high performance computing and extreme scale environments, often new and innovative results can be derived for commercial IT environments. As a recent study by the German Federal Ministry of Education and Research shows, 43 percent of businesses are participating in recent research for developments in HPC-Software for scalable parallel computing for new algorithms in Germany [7]. HPSS is at current in use by more than 30 sites worldwide, mainly in use at governmental institutions, research organisations or in defence, but rarely in place at commercial businesses.

In order to measure how fast data from applications is running, is defined by its data transfer rate to the storage infrastructure. In a world where programs and files are becoming ever-larger, the highest data transfer rate is most desirable [8]. However, as technology moves quickly to advance the data transfer rate of many components, consumers are often faced with systems that incorporate varying specifications [8]. This circumstance becomes visible in figure 1.

Figure 2 explores the simple idea, how to position each storage solution by its data volume to be archived into a single system ("high" as in greater than 15 PB, "low" as below 1 PB data volume) and by which favorable processing type ("batch" vs. "real-time"). This positioning does not

 ISBN: 978-1-61208-161-8

account the type of network used and assume a standard file-size of ~ 160 MB [9]. This is an average file size for an typical HPSS installation.

Quadrant A positions the storage systems and solutions that apply mostly for real-time processing and are able to store per single system more than 1 PB. Quadrant B positions storage systems and solutions that are rather batch oriented as processing type of data. However, newer features in metadata management of the utilized relational databases, will make these solutions more applicable for real-time processing. Quadrant C positions storage systems and solutions that have a lower possibility to store less than 1PB and are rather seen as "Storage Optimizers" and hence not for long-term archiving needs applicable storage solutions for high data volumes. Quadrant D positions furthermore NAS filer storage systems that have a high real-time processing capability, however lack the vertical scalability in storage capacity. In addition, RACE is an appliance in front of the Storage Area Network (SAN) and not used as a single storage system, rather reflecting a "Storage Optimizer" as earlier described.

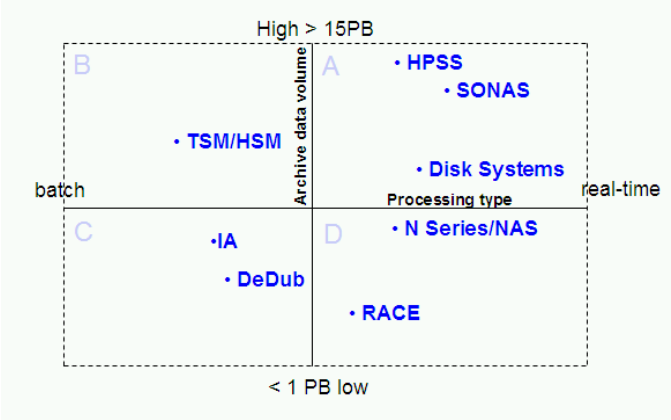

Figure 2. Positioning Archive Systems by Data Volume vs. Processing Type (batch vs. real-time)

Data transfer rate (DTR), is the speed at which data can be transmitted between devices. This is sometimes referred to as throughput [8]. The data transfer rate of a device is often expressed in kilobits or megabits per second, abbreviated as kbps and mbps respectively. It might also be expressed in kilobytes or megabytes, or KB/sec and MB/sec [8].

HPSS is software that manages PB of data on disk and robotic tape libraries. HPSS provides highly flexible and scalable hierarchical storage management that keeps recently used data on disk and less recently used data on tape. HPSS uses cluster, LAN and/or SAN technology to aggregate the capacity and performance of many computers, disks, and tape drives into a single virtual file system of exceptional size and versatility [10]. This approach enables HPSS to easily meet otherwise unachievable demands of total storage capacity, file sizes, data rates, and number of objects stored [10]. HPSS is known to be the leading archival storage software system to fulfill extreme scale requirements [11]. Speeds are limited only by the underlying computers, networks, and storage devices. HPSS can manage parallel data transfers from multiple network-connected disk arrays at hundreds of megabytes per second. These capabilities make possible new data-intensive applications such as high definition digitized video at rates sufficient to support real-time viewing [12]. At the German Climate Computing Center (DKRZ), the implemented HPSS as the data archive has available bidirectional bandwidth of 3 GigaByte/s (sustained), and 5 GigaByte/s (peak) [13]. A central technical goal of HPSS is to move large files between storage devices and parallel or clustered computers at speeds many times faster than today's commercial storage system software products [14].

HPSS supports striping to disk and tape. At 100 MB/s, it takes almost 3 hours to write a 1 TB file to a single tape. Using an 8-way tape stripe, that time is cut to less than 25 minutes [14]! Commercially available 2,5'' disks based on MLC chips, solid state storage featuring SATA (3 Gb/sec.) interface, have data transfer rates of 150 MB/sec. (read) and 90 MB/sec. (write) [15]. It must be differentiated in reading data and writing data to demonstrate real-time or near real-time data storage access. Writing data to storage is sometimes called "data ingestion" [15]. A typical day of writing data to the archive at the European Centre for Medium-range Weather Forecast (ECMWF) is ~ 42 TB [17]. Reading from the archive is ~ 19TB per day [17].

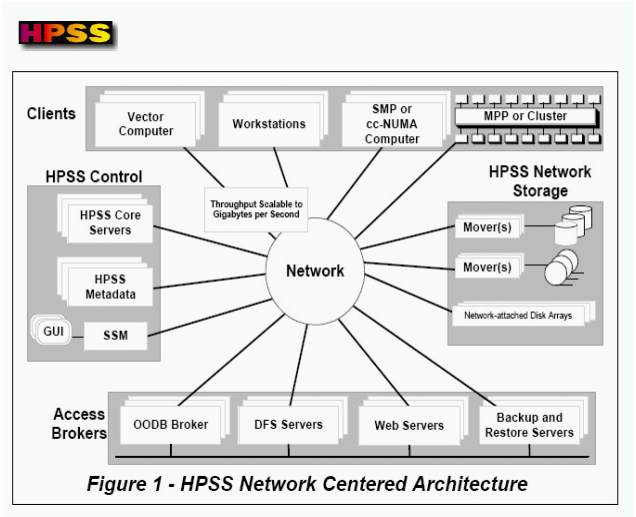

Figure 3. Basics of High Performance Storage System [12]

The focus of HPSS is the network, not a single server processor as in conventional storage systems. HPSS provides servers and movers that can be distributed across a high performance network to provide scalability and parallelism. The basis for this architecture is the IEEE Mass Storage System Reference Model, Version 5 [12], see Figure 3. Once a transfer session is established, the actual data transfer takes place directly between the client and the storage device controller [12].

HPSS achieves high data transfer rates by eliminating overhead normally associated with data transfer operations.

 ISBN: 978-1-61208-161-8

In general, HPSS servers establish transfer sessions but are not involved in actual transfer of data. For network-attached storage devices supporting IPI-3 third party transfer protocols, HPSS Movers deliver data at device speeds. For example, with a single HiPPI attached disk array supporting IPI-3 third party protocols, HPSS transfers a single data stream at over 50MB/sec [12].

The HPSS Application Program Interface (API) supports parallel or sequential access to storage devices by clients executing parallel or sequential applications. HPSS also provides a Parallel File Transfer Protocol. HPSS can even manage data transfers in a situation where the number of data sources and destination are different. Parallel data transfer is vital in situations that demand fast access to very large files [12].

Due to the fact that HPSS does not have a volume-based licensing, the invest for such a high performance archiving solution is costly for the first year, but after it operates, the pay-off is clear, as it only costs for the tape library and media have to be calculated in future years. At the UK Met Office the payoff of HPSS for research data was clear: the archive size was in 2010 approx. 3PB, but with a forecast to year 2013 by 20PB [18], the UK Met Office needed a solution that support Extreme Scale Computing, store and keep data as central asset for research; but at the same time to be Energy efficient and affordable long-term. The use of HPSS is hence technically and commercially attractive for this type of storage and archiving demands see Figure 4.

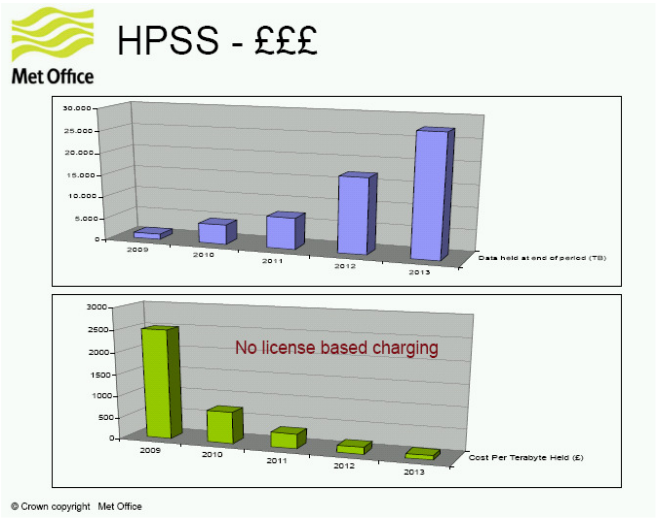

Figure 4. HPSS - £££ [18]

III. REAL-TIME PROCESSING WITH HPSS

The recommended use of HPSS is for storing data sets that require high availability and that need longer-term storage. It is not intended for real-time access or short-term storage for temporary data [18]. Therefore, the team "near real-time" for HPSS is introduced by Harry Hulen, IBM Global Services – Federal, Houston, Texas [16]. He describes the possibility of using HPSS for a near real-time application, as follows:

"A good example of near-real-time data ingestion is when a scientific experiment is running and producing data that must be captured, such as a high energy physics experiment or a satellite that is sending down data that must be ingested and put on tape. There are two cases: "bursty" data that requires high rate ingestion for a period of time, and then stops for a while. In this case a disk cache usually takes care of the real time part of the problem, and then data migrates to tape when the burst of ingestion is complete. A tougher case is when there is a steady stream of data for a long period of time. In this case the tape system must operate at close to real time for a long period. An example of this kind of data would be the satellite that continually downlinks data.

Usually the read side of the problem is defined by response time and not data rate. If data is on disk and the storage system is not suffering from queuing, then the response is quick. The best example to think of is Yahoo or Google. If data has migrated to tape and then purged from disk, then the user may have to wait for minutes instead of seconds or sub-second responses. This reading of data is real-time, or near real-time, in the sense that the answer closely follows the interrogation; however, the terms "real time" and "near real time" are used less often for this type of transaction. A distributed system like HPSS are valuable for high data rate near real time data ingestion because a cluster can keep data flowing to more disks and tapes than a monolithic, single-computer storage system can manage.

When designing a tape system for near-real-time storage, it is necessary to accommodate the non-real-time characteristics of tape, particularly what happens when a tape is full. At that point the tape cartridge must be unmounted and another tape mounted. This process can take two to three minutes in a modern robotic tape library. There must be sufficient tape drives so that other tapes are waiting and ready to go, or there must be sufficient disk to catch and hold the ingested data. In any case, it is necessary to over-provision a near real time system when compared with a system that can gracefully allow longer queues to develop.

HPSS supports near-real-time data ingestion for both bursty data and steady state data with its distributed data mover architecture and its ability to write stripes of data across multiple tape drives."

Within HPSS, file metadata are maintained in DB2, IBM's real-time database system [19]. Further, other major application domains, such as real-time data collection, also require such extreme scale storage. We believe the HPSS architecture and basic implementation built around a scalable relational database management system (IBM's DB2) make it well suited to this challenge [20].

IV. CONCLUSION

Real-time processing and long-term archiving needs will not be possible to be accommodated at the same time in just one single storage solution. Either for the main objective to support really real-time processing by high I/O systems with

their respective DTRs disks systems are applicable or for long-term archiving needs with tape infrastructures. A combination of disks and tape systems, i.e., with HPSS, will address at leas near real-time processing needs.

There is to take into consideration future developments and features, such as a) Real-time compression mechanism and b) future of HPSS:

a) IBM Real-Time Compression: IBM Random Access Compression Engine (RACE) is IBM patented technology that is the key to IBM Real-time Compression Appliances for NAS. RACE technology allows read and write operations from any location within a compressed file while avoiding the need to decompress the whole file. Additionally, the technology preserves all file metadata as it is stored on disk, making the compressed file transparent to the application. The technology enables data compression without compromising performance or data integrity. By operating in real time and reducing the amount written to disk, IBM Real-time Compression solutions enable enhanced storage performance and efficiency [21].

Ideal for seismic research, the company Halliburton has developed a storage solution called "PetroStor™," that is an integrated solution comprised of technologies from Landmark, NetApp and Storwize that can lower the cost of on-line archival storage to less than $1,000 per TB by using industry leading technology for data compression and de-duplication for oil and gas exploration organizations [22].

b) HPSS will include in future real-time applications requirements. Already in 2004, at SC'04, IBM demonstrated with HPSS performance using three computers, one each for HPSS, reading and writing. A large 128 GB file was written and read in 512 MB blocks using16-way striped SAN-attached disk files, using 8 host bus adapters on each client computer. As one computer wrote each block, it was immediately read by a second computer, thus demonstrating "read behind write" performance. The file transfers were measured at 1016 MB/s on the write side and 1008 MB/s on the read side, for an aggregate data rate of just over two GB per second [23].

For instantaneous throughput, the HPSS development aims at ~ 50GB/s for HPSS [23]. In addition, there are multiple developments underway, to how to most effectively utilize DB2 partitioning and other capabilities to support multiple dynamic Core Servers (Metadata Servers). The upcoming newest Version 8 of HPSS will include architecture of Distributed Metadata. Using the DB2's Data Partitioning Feature this will provide the necessary infrastructure to distribute HPSS metadata. The partitioning feature is based on a share nothing architecture, where each system manages the local partition, but has access to all partitions transparently [24]. DB2 is extremely well tested and supported by IBM; HPSS development and takes advantage of this mature and robust capability [24].

First prototypes at the HPSS development team have shown the new architecture provides 10x performance of HPSS V7 single metadata server architecture [24]. Therefore, HPSS will accommodate both, end user expectation to have data in real-time or at least "near real-term" access, store them long-term, and, for IT-Managers, HPSS demonstrates a viable alternative disk and tape solution scenario, to save costs for extreme scale future storage data growth needs.

As discussed in this paper, real-time processing needs will be vital to be addressed soon in data management infrastructure and technology decisions.

V. ACKNOWLEDGMENT

The author would like to thank Harry Hulen from IBM Houston, Texas, for his help in getting the examples for read and write and the real-time implication for HPSS.

References

[1] TechRepublic Pro, (2008) Source: NetApp, Pergamum: Replacing Tape With Energy Efficient, Reliable, Disk-Based Archival Storage. [Online]. Available: http://www.techrepublic.com/whitepapers/pergamum-replacing-tape-with-energy-efficient-reliable-disk-based-archival-storage/1294097, [Accessed 2011-06-17].

[2] H. Newman, "LBL HPSS Workshop for DOE/SC," CTO/Instrumental, Inc., Jul. 2009.

[3] J. Mahendra, (1998) What is archive? Definition from Whatis.com [Online]. Available: http://searchstorage.techtarget.com/definition/archive, [Accessed 2011-06-17].

[4] Wikipedia.org, (2011) Transaction processing system [Online]. Available: http://en.wikipedia.org/wiki/Transaction_processing_system, [Accessed 2011-06-17].

[5] H. Weber, Deutscher Wetterdienst (DWD), HPSS Reference and Booth Demo at official presentation at CeBIT 2011 trade fair, Hannover, Germany 2011.

[6] IBM Systems and Technology, (2010), Optimize storage capacity with IBM Real-time Compression, [Online]. Available: ftp://public.dhe.ibm.com/common/ssi/ecm/en/tsb03020usen/TSB03020USEN.PDF, [Accessed 2011-06-18].

[7] Bundesministerium für Bildung und Forschung (BMBF), (2009), Ergebnisse der ersten HPC- Fördermaßnahme, „HPC-Software für skalierbare Parallelrechner", Bundesministerium für Bildung und Forschung, [Online]. Available: http://www.pt-it.pt-dlr.de/_media/HPC_Infoblatt.pdf, [Accessed 2011-06-23].

[8] wiseGEEK, (2003-2011), What is Data Transfer Rate? [Online]. Available: http://www.wisegeek.com/what-is-data-transfer-rate.htm, [Accessed 2011-06-23].

[9] H. Hulen and G. Jaquette, "Operational concepts and methods for using RAIT in high availability tape archives", [Online]. Available: http://www.storageconference.org/2011/Papers/MSST.Hulen.pdf, [Accessed 2011-06-08].

[10] HPSS Collaboration.org, (2010), What is High Performance Storage System? [Online]. Available: http://www.hpss-collaboration.org/index.shtml, [Accessed 2011-06-18].

[11] National Energy Research Scientific Computing (NERSC) Facility, Extreme Scale Workshop, (2009), HPSS in the Extreme Scale Era, [Online]. Available: http://www.nersc.gov/assets/HPC-Requirements-for-Science/HPSSExtremeScaleFINALpublic.pdf, [Accessed 2011-06-18].

[12] H. Hulen, Basics of the High Performance Storage System, [Online]. Available: http://www.isi.edu/~annc/classes/grid/papers/HPSS-Basics.pdf, [Accessed 2011-06-18].

 ISBN: 978-1-61208-161-8

[13] Deutsches Klimarechenzentrum (DKRZ), (2011), Data Archive, [Online]. Available: http://www.dkrz.de/Klimarechner-en/datenarchiv, [Accessed 2011-06-18].

[14] J. A. Gerry, H. Hulen, P. Schaefer, and B. Coyne, (2009), High Performance Storage System Overview, Slide Number 10, [Online]. Available: http://www.hpss-collaboration.org/documents/HPSSIntroduction2009.pdf, [Accessed 2011-06-18].

[15] Logic Supply, (2011) Transcend Commercial 2.5" SATA SSD, 64 GB , [Online]. Available: http://www.logicsupply.com/products/64gssd25s_m, [Accessed 2011-07-03].

[16] H. Hulen, personal communication per email, Jun. 17, 2011.

[17] S. Richards and F. Dequenne, "HPSS at ECMWF", High Performance Storage System User Forum (HUF) 2010 presentation, Sept. 28, 2010, Hamburg, Germany.

[18] M. Francis, UK Met Office, "HPSS User Forum", High Performance Storage User Forum (HUF) 2010 presentation, Sept. 28, 2010, Hamburg, Germany.

[19] University Corporation for Atmospheric Research (UCAR), (2011), CISL HPSS, [Online]. Available: http://www2.cisl.ucar.edu/book/export/html/964, [Accessed 2011-06-18].

[20] HPSS Collaboration.org, (2010), Learn about HPSS, The HPSS Collaboration, [Online]. Available: http://www.hpss-collaboration.org/hpss_collaboration.shtml, [Accessed 2011-06-18].

[21] IBM Systems and Technology, (2011), FAQ: IBM Real-time Compression [Online]. Available: http://www.realtime compression.com/library_brochures.asp, [Accessed 2011-06-18].

[22] Halliburton, (2009, Feb.), Press Release, Landmark introduces its PetroStor™ Cost-Competitive, Disk-Based tape Replacement Storage Solution for Oil and Gas Environments, [Online]. Available: http://www.realtimecompression.com/content/press_releases/PetroStor_Storage_Solution_Press_Release.pdf, [Accessed 2011-06-26].

[23] D. Watson, (2007, Oct.), Yes, Virginia, There is an HPSS in Your Future, [Online]. Available: http://www.hpss-collaboration.org/documents/WatsonSalishan2007.pdf, [Accessed 2011-06-26].

[24] D. Boomer, K. Broussard, and M. Meseke, (2011), HPSS 8 Metadata Services Evolution to Meet the Demands of Extreme Scale Computing, [Online]. Available: http://www.pdsi-scidac.org/events/PDSW10/resources/posters/HPSS8.pdf, [Accessed 2011-06-26].

 ISBN: 978-1-61208-161-8