N94-33814

Mass Storage - The Key to Success in High Performance Computing

Richard R. Lee

Data Storage Technologies, Inc.
Post Office Box 1293
Ridgewood, New Jersey USA 07451-1293
Phone: (201)-670-6620
Fax: (201)-670-7814
rrl@dst.com

Abstract: There are numerous High Performance Computing & Communications Initiatives in the world today. All are determined to help solve some "Grand Challenges" [1] type of problem, but each appears to be dominated by the pursuit of higher and higher levels of CPU performance and interconnection bandwidth as the approach to success, without any regard to the impact of Mass Storage. My colleagues and I at Data Storage Technologies believe that all will have their performance against their goals ultimately measured by their ability to efficiently store and retrieve the "deluge of data" created by end-users who will be using these systems to solve Scientific Grand Challenges problems, and that the issue of Mass Storage will become then the determinant of success or failure in achieving each projects goals.

In today's world of High Performance Computing and Communications (HPCC), the critical path to success in solving problems can only be traveled by designing and implementing Mass Storage Systems capable of storing and manipulating the truly "massive" amounts of data associated with solving these challenges. Within my presentation I will explore this critical issue and hypothesize solutions to this problem.

Topics to be discussed: To properly lay the foundation for this paper I must briefly discuss the history of Mass Storage in respect to high performance computing. Once these background materials are discussed, I will then focus the body of the presentation; "Current and Future HPCC Initiatives and the Impact of Mass Storage on Their Success or Failure"

The areas of background to be discussed are as follows;

i.- Basic Definitions and Key Underlying Factors
ii.- The Early Days of Mass Storage and its Role in Advancing the art of Computing
iii.- The current Role of Mass Storage in High Performance Computing & Communications

Basic Definitions and Key Underlying Factors

Mass Storage per my definition is: "Any type of Storage System exceeding 100 GB in total size (not off-line), and operating under the Control of a Centralized or Distributed File Management Scheme.

CPU Power has Increased at a rate of 25% per year (CAGR) for the Past 10+ years, While I/O Bandwidth/Rates have Remained Constant. 1 MIP of CPU power should correspond to 1 MB/s, of I/O bandwidth performance. This fundamental relationship has not been adhered to since the early days of the mainframe, and is not found anywhere today in HPC.[1]

[1] The following are a partial listing of the HPCC Coordinating Offices "Grand Challenges" research teams projects; Computational Quantum Materials, High Resolution Operational Weather Forecasting, Numerical Tokamak, Multidiscipline Simulation of High Speed Civil Transport and Performance Aircraft, etc.

I/O and Network Bottlenecks, along with OS and other software inefficiencies are crippling all types of computing systems today, and jot just those utilized in the world of HPC.

There are no panaceas to solve the "Mass Storage Crisis" found in HPC today. A new paradigm in Systems Architecture and Design Philosophy is required to meet the requirements of future HPC environments. [4], [5], [6], [7]

As the "deluge of data" continues to grow (25+ CAGR)in the HPC data center, many end-users will be faced with the dilemma of not being able to store Critical Data due to increasing economic constraints. Not only is the Cost-per-MB of On-Line and Secondary storage too high, but the CPU cycles required for off-loading and accessing large files (Multi-GB) is quickly becoming unaffordable! [2], [5]

The Early Days of Mass Storage and its Role in Advancing the Art of Computing

Early Mass Storage systems consisted of removable hard disk packs, and magnetic tape drives-freestanding or serving large off-line round tape repositories. These early peripheral based systems were augmented by unique, proprietary storage systems such as the IBM 1360 photo-store, the IBM 3850 helical scan tape library, the Ampex terabit memory , the Braegen automated tape library, and others from CDC, Remington Rand, etc. Although these early systems offered increased capacities over stand alone peripherals, none were commercially successful and most were sold into US Government labs or to the Intelligence Agencies.

Surprisingly thought, these early systems were much better matched to their accompanying CPU's I/O bandwidth than that found today and they truly did provide very good performance and value to the customer during their heyday, given the lack of practical alternatives.

The Current Role of Mass Storage in High Performance Computing

Mass Storage systems today range in size from 100 GB to 30+TB, with all under the control of some type of dedicated File Server CPU. Most of these systems are; slow in performance, woefully under powered in terms of I/O Bandwidth, and utilize very immature Hierarchical File Management software schemes. These systems provide cost reductions in terms of storing a variety of bitfile data set types, but do very little to actually improve the performance of the overall system. This problem is further exacerbated by the divergence between CPU and Network Operating Systems (MVS, UNIX, OSI, etc.). and their fundamental differences in approach to the task at hand and the hardware interfaces supported.

All of today's' Mass Storage systems utilize dedicated, and very expensive components in order to optimize performance capabilities and most are based on technologies developed in the 1980's which are now just becoming commercialized e.g. RAID, HiPPI, FDDI, DD-2 , UniTree, etc. These systems will be the benchmark in the early '90's but will be replaced by radically new approaches scheduled to become available in the mid-'90's.[5], [9], [3]

"Current and Future HPCC Initiatives and the Impact of Mass Storage on Their Success or Failure"

1.0 The HPCC Initiatives

High Performance Computing and Communications or HPCC has become the buzzword acronym of the early 1990's. In its simplest form it refers to Public Law 102-194 1991 The High Performance Computing Act/Initiative Of 1991, signed into law by President, George Bush (12/91). It is broken down into four constituent parts;

1.- TeraFlop (now referred to as "Teraop") Computing
2.- NREN (National Research & Education Network)
3.- Advanced Software and Algorithm Development
4.- Training & Research

In its most complex form HPCC is a catchall for every advanced computing activity in the world today. It has been widely promulgated as fundamental to the Clinton administrations' endeavors to improve the US's competitiveness and productivity in respect to Japan and Europe, and is deeply mired in party politics. Many new initiatives have been tacked on to the original legislation[2] and funding is anticipated to increase in out years regardless of the wrangling by each political party that continues to go on..

2.0 Mass Storage's Role in the Success or Failure of HPCC

In spite of its politicization, HPCC has provided a focal point for addressing all issues relevant to the future of computing. In monitoring this focus, it is painfully obvious that the issue of Mass Storage has been largely ignored, with the exception of the National Storage Laboratory @ LLNL and a few other small projects spread around the HPCC community. [7], [6], [2]

When the issues of "how to achieve" the levels of performance necessary to solve "Grand Challenges" scientific computing problems are addressed by all parties involved at conferences and symposia as well as in articles and abstracts and testimony to Congress; it is painfully clear that Mass Storage is forgotten altogether or minimized in importance in the grand scheme of things This is a critical error in my opinion.

As computing moves quickly towards client-server topologies in every imaginable application, the network will essentially become the computer. Numerous heterogeneous computing resources will be linked together over "data superhighways" (multi-gigabit links) to form large on-line computing capabilities. These systems will range from clusters of high-end workstations to numerous supercomputers in many locations linked together i.e. the NSF MetaCenter. These meta-type systems are touted as having the capability to finally begin to address some of the really difficult "Grand Challenges" problems that many believed could only be solved by Teraops type machines of the future (Table Number 1 lists the capabilities of many of the network topologies being discussed to form the '"data superhighways".) This approach has been widely endorsed as of late, but within those endorsements there is no mention of how these meta-type systems will store and manage the avalanche of data created by "the system", much less how one can practically afford the cost associated with the task.[3]

The NSF MetaCenter is one of these systems and will utilize the capabilities of some 21 supercomputers (vector, scalar & parallel), linked together over an optical network (NSFNET). The amount of data to be generated by this system begins to boggle the mind, and yet is is treated as a secondary issue by many in the MetaCenter development group. What is clear is that when these types of systems are finally up and running is that they will all essentially swamp their local storage capabilities and that the data sets generated by the meta-computer will not be able to be stored and further manipulated due to cost, bandwidth and capacity constraints at every link of the "MetaCenter chain". This is a quandary not only for the meta-

---

[2] As of this writing, the following new bills and acts regarding add-ons to the original HPCC legislation are in process;

1.- "the National Information Infrastructure Act of 1993 (formerly known as "the High Performance Computing and High-Speed Networking Applications Act" - HR 1757
2.- "the National Competitiveness Act of 1993", HR 820, S.4
3.- "the Electronic Library Act of 1993", S.4 Attachment

[3] It has been said by many that current costs in the data center are split 50-50 between the CPU and the peripherals. This has been fairly accurate until recently when, scientific visualization and the use of more on-line archives has produced a phenomena where peripheral costs are now climbing to 60+% of the overall cost and we predict that in the future this may rise to almost 75% if not abated by a new paradigm in systems design.

computer types, but those involved in visualization, parallel computing and scientific activities such as CD, etc. The quandary is as follows: What makes more sense; to utilize the entirety of the data centers available resources (storage capacity and CPU cycles) to store the results of a complex computational problem, or to throw the data away and re-calculate the results on another day, often without the same results achieved or computational resources available? In its simplest form this quandary speaks to the fact that we have spent the last 20 years pursuing the Holy Grail of CPU power and speed, but cannot utilize it to its fullest capabilities, because we have nowhere to store the data!

Table 1

Emerging Networking Standards

Network:	Type:	Data Rate(s):	Data Type(s):	Max. Distance:
Fast Ethernet	TP- Cu	100 Mb/s	Digital	25m
CDDI	TP - Cu	100 Mb/s	Digital	50-100m
FDDI	Opt. Fiber	100 Mb/s	Digital	60 km
FDDI-II	Opt. Fiber	100 Mb/s	A, V & Digital	60 km
HiPPI	TP - Cu	800/1600 Mb/s	Digital	25m
Fibre Channel	Opt. Fiber	1000 Mb/s	Digital	10 km
SONET/ATM/B-ISDN	Opt. Fiber	51-2488 Mb/s	A, V & Digital	LD Network Limits

In spite of it looming over the future of HPCC, the issue of Mass Storage is not insurmountable by any means. What is needed are new approaches to the problem and new storage devices capable of storing, manipulating and retrieving vast sums of data at faster speeds, with higher volumetric efficiency and will attendant incremental reductions in cost-per-unit stored.

Many of the storage technologies shown in Table Number 2 have been around for some time now, but have been recently adapted to offer orders of magnitude increases in capacity and bandwidth, while increasing volumetric efficiency (in terms of physical space utilized) as well as having unbefore seen low costs-per-unit of data stored.

Table 2

High Performance Data Storage Devices

Name/Std.:	Storage Technology:	Data Rate(s):	Data Capacity:	Device Cost:
IBM 3490E	1/2" Longitudinal OT	4.5 MB/s	500 MB (Native)	$70K
ANSI DD-1	19mm Helical OT	15 - 45 MB/s	15, 50, 100 GB	$250K
Ampex DD-2	19mm Helical MT	15 MB/s	25, 75, 186 GB	$200K
STK DD-3	1/2" Helical MT	15 MB/s	20 GB	$65K
Metrum 2150	1/2" Helical OT	2-4 MB/s	14.5 GB	$35K
CREO 1003	35mm Optical Tape	3 MB/s	1000 MB	$250K

These devices when wedded with robotics and advanced Data Management Software schemes can begin to meet the challenge of the MetaCenter and other such initiatives. They provide almost infinite capacity, with wide bandwidth (for time is money) and extremely low cost relative to the service that they are providing.

The issue of Data Management cannot be overlooked when challenging the "deluge of data" to be found in the future. This class of software and its influence on the systems architecture cannot be relegated to the role of freeing up more DASD, and therefore temporarily abating the data centers capital problems. (see Table Number 3 for a listing of currently available File System and File Management S/W) It must instead become the central director of all activities within the network and its attached resources (CPU's, peripherals, etc.). The orderly flow of data within the hierarchy of storage devices and networks will ultimately control the overall capabilities of the entire computational system.. The need for this class of software is made self-evident by the MetaCenter concept. Much attention is currently being paid as to how to break big problems up into large parallel pieces, but this effort will be futile if not supported by the Data Management S/W mandated by this type of challenge.

Table 3

File Systems/Data Management Software

Trade Name.:	Developer(s):	Type:	OS Baseline:	OSI/IEEE Oriented:
Network File System - NFS	Sun Microsystems	F.S.	UNIX	No
Andrew File System - AFS	CMU/Transarc	F.S.	UNIX	No
OSF DCE/DFS	OS Foundation	O.S./F.S.	OSF/1	OSI Model
DataTree	LANL/DISCOS	F.M.S.	MVS	Yes (early)
EpochServ	Epoch Systems	F.M.S.	UNIX	No
DFSMS/DFDSM	IBM Corp.	F.M.S.	MVS Family	No - Proprietary
Open Vision UniTree V1.8X	LLNL/DISCOS	F.M.S.	UNIX/NFS	Yes V3.0
NSL UniTree V1.X	LLNL/IBM	F.S./F.M.S.	UNIX/AFS	Yes V5.0 Oriented

Conclusions and Recommendations

Mass Storage has become a critical path driver in the success of all HPCC Initiatives. To achieve the level of Systems Performance required to solve "Grand Challenges" computing problems, all elements of system must be optimized, with special emphasis on the role of Mass Storage in controlling the performance of the entire system.

The cost of storage will be a critical factor in determining the allocation of resources in the HPCC Initiatives. To meet the challenge, many orders of magnitude of cost reduction in -per-unit data stored must achieved. Part and parcel to these cost reductions will be increases in storage device bandwidths, volumetric efficiency and overall capacity. The hardware costs will be supported increasingly efficient Data Management S/W systems who manage and optimize the flow of data within the entire system.

I strongly advocate that Mass Storage and its attendant issues be brought to the forefront of the HPCC Initiatives. Only by applying this level of visibility and sensitivity to the issue will there be success in utilizing the HPCC Initiatives to solve "Grand Challenges" problems. Mass Storage can no longer be a secondary issue.

References

1. Lee, R. and Dan Mintz, "Grand Challenges in Mass Storage - A Systems Integrators Perspective", Second NASA Goddard Conference on Mass Storage Systems and Technologies, Greenbelt, MD, September 1992

2. Lee, R., "The Future of Mass Storage", THIC Winter Meeting, San Diego, CA, January 1993

3. Lee, R., "Interfacing 19mm Helical Scan Recording Systems to Computing Environments", THIC Spring Meeting, Annapolis, MD, March 1990

4. Kuhn, T., *"The Structure of Scientific Revolution"*, University of Chicago Press, Chicago, IL 1970

5. Lee, R., 19mm Helical Scan Recording Technology for Data Intensive Computing Environments", 10th IEEE Symposium on Mass Storage Systems (vendor poster session), Monterey, CA, May 1990

6. Coleman, S. and R.W. Watson, "The Emerging Paradigm Shift in Storage System Architectures", review copy for Proceedings of the IEEE, April 1993

7. Coyne, R. , H. Hulen and R. Watson, "Storage Systems for National Information Assets", Proceedings-Supercomputing '92, Minneapolis, MN, November 1992

8. Lee, R., "Mass Storage - the key to success in high performance computing" (early version), Convex File Server Seminars, Milan/Rome, Italy, February 1993

9. Lee, R., "19mm Data Storage Applications", THIC Fall Meeting, Annapolis, MD, October 1990