U.S. Department of Energy Best Practices Workshop on

File Systems & Archives

San Francisco, CA

September 26-27, 2011

Position Paper - Business Breakout

Kim Cupps
Lawrence Livermore National Laboratory
cupps2@llnl.gov

ABSTRACT / SUMMARY

This "Business of File Systems and Archives" position paper will describe several LLNL best practices that help formulate and optimize the cost/benefit analysis all center's face in their quest to provide an ongoing, exceptional computing environment within a finite budget.

INTRODUCTION

LLNL's primary computing complex serves approximately 2800 users with access to 1.7 peak PetaFLOPs of compute, 14PB and 300GB/s of parallel file system capacity and bandwidth and 42 PB of archival storage. LLNL seeks to optimize the user experience, providing a long-lived, highly productive environment for our customers, while staying within our budget. While cost/benefit is easy to say, it's a complicated balance of usability, availability, flexibility of administration, longevity, productivity and many other factors that form the basis of our spending decisions for file systems and archives. There are several practices we use to help us make decisions on what and how much to buy, what improvements must be made and which software to develop ourselves and which to buy off the shelf. First, we place a high value on gathering direct user input via a variety of mechanisms including user meetings, surveys and customer interviews. Another best practice is to monitor and measure use of resources and plan buys "just in time" (JIT). Strong partnerships with vendors as well as hedges against the trap of becoming beholden to a single vendor or technology for file systems or archive is another best practice. Lastly, planning is crucial to any coherent business strategy. LLNL formalizes the plan for file system and archive resources by producing the "I/O Blueprint", a procurement and effort planning prioritization document.

Gathering User Feedback

Making sound business decisions requires a good understanding of the current state of affairs. It's quite easy to live in a world isolated from those who use the file systems and archives every day, just as it is common for users to work around issues and inconveniences rather than report them. We have the typical trouble ticket system user surveys to help us understand issues; this is feedback we receive on a daily basis. In addition, LLNL holds quarterly user meetings that include a user talk as well as a set of talks on relevant center activities. These meetings include a general feedback session. Most important, LLNL rotates through "Science Team Interviews" so that we meet with teams every two years or so to elicit actionable feedback. A team of center personnel representing management, platform, file system, archive and user services personnel goes out to the customer work area and asks

pointed questions about the compute environment. In general, we find that problem areas spring out of discussion and are not the sorts of issues that people call and report, often they don't even write down the issue in advance of the meeting. For example, the development of HTAR, a multi-threaded file packaging and transfer mechanism from the local file system to the HPSS archive, was the direct result of complaints received from users regarding slow transfers of small files to HPSS. The development of Lorenz, our user dashboard that shows file system usage, NFS quotas and many other customizable fields was also the result of strong user collaboration and input. All of these user feedback mechanisms serve to inform the center about where our customers feel we have the most room for improvement – this is critical to our planning.

Measuring and Metrics, JIT

Another way we identify areas for improvement is by collecting data on various aspects of the production environment. We gather data on everything from component failure rates, to sizes of files stored in the file system, to file system specific and center-wide uptime percentages for both classified and unclassified file systems and archives.

Unclassified Lustre File System Availability Statistics 8/1/11 - 9/1/11					
	Unplanned		Planned		Uptime %
File system	Impaired	Down	Impaired	Down	
Center Wide	0.83	3.42	0	0	99.86%
lscratch a	0	1.58	0	0	99.79%
lscratch b	0	0	0	0	100%
lscratch c	0.08	1.75	0	0	99.75%
lscratch d	0.75	0.08	0	0	99.89%

Tracking isn't limited to analysis of failures and availability, it's also crucial to our "just in time" purchase strategy for archive media and tape drives. Cartridges are used up by both a steady stream of new data being written to tape as well as an ongoing repack from soon-to-be-retired media of about 500 cartridges a month. Careful tracking of cartridges insures that tape buys are done on-time, but not so far in advance that the tape is never used. Just-in-time has been shown to be the most cost efficient way to purchase consumables, and with tape densities doubling (recently quintupling) every year and a half, the value of this best practice is clear.

Partnerships Coupled with In-House Software Expertise

Strong vendor partnerships are crucial to successful operation of a center. Changing vendors incurs added costs such as retraining staff, forming new relationships and learning new support processes. However, becoming a strictly one vendor operation significantly increases risk. These risks include the company going out of business, dropping support for the product line or unreasonably raising prices. A best practice at LLNL that reduces the inherent risk of the strong vendor partnership, is a staff of in-house software developers forboth open source projects (Lustre, SLURM, RHEL) and joint development contracts (HPSS). The software developers provide key value by:

1) Solving production problems immediately (increasing system uptime and thereby user productivity);
2) Providing a strong voice for DOE HPC requirements
3) Providing a mechanism for the center to remain technically competent and engaged in leading edge technology

Other important components of strong vendor partnerships include membership on vendor customer advisory boards, leadership of product user groups and regular attendance at vendor executive level roadmap briefings.

Advanced Technology and Testbeds

The Hyperion testbed at LLNL includes an 1152 node QDR IB interconnected commodity Linux cluster with two SANs (GE, IB) connected to multiple vendor storage subsystems. The testbed serves multiple purposes. It is a partnership that allows vendor partners to test their software and hardware at scale. It is a platform that allows LLNL to investigate interesting technologies

(NAND Flash, tiered storage, NFS accelerators…) as they become available. WhamCloud performs Lustre testing at scale. Mellanox tests new cards and drivers. DDN and Netapp test new controller technologies and LLNL investigates all of this new technology before it comes to market.

The Lustre testing at scale on Hyperion, and LLNL's participation with others in the concept of creating "Lustre Centers of Excellence" is an excellent example of how investment in strong vendor partnerships and testbeds can significantly impact the quality of a product.

Planning

LLNL produces a yearly planning document called the I/O Blueprint. The goal of the Blueprint is to achieve a balanced infrastructure to support the Center's compute platforms. The Blueprint documents planned purchases in global parallel file systems, NAS, visualization, network and archive areas. It also discusses area specific Center issues and plans for remediation.

The FY05 Blueprint is the document that called for converting from dedicated filesystems to a site-wide global parallel file system. During the era of local file systems, each platform purchase required that dedicated file system hardware be bought for use solely by that platform. Without global file systems, a platform was only able to leverage the speed and capacity that it came with and data needed to be moved or copied to each platform when required. Today, global file systems allow new platforms to leverage existing disk resources and allow existing platforms to take advantage of global resources added over time. As a result we are able to enhance file system resource utilization, eliminate the copying of data, ensure that file system hardware is best-of-breed rather than that available from a particular platform vendor, and focus on center-wide I/O requirements rather than that of individual machines. In short, there is a clear cost/benefit win.

Calculating bandwidth and capacity requirements for archive, file system and networks is not an exact science, and we have used different "rules of thumb" over time to plan purchases. For example, directly copied from the FY08 I/O Blueprint: *"The rule of thumb used in the past for capability platforms was that the file system should provide between 100MB/s and 1GB/s of bandwidth for every TeraFLOP. Dawn file system bandwidth requirements are projected to be 200MB/s per TeraFLIN and Sequoia is projected to be 100MB/s per TeraFLIN leading to 100GB/s and 500GB/s estimates for delivered SWGFS bandwidths for these machines."*

From our FY11 Blueprint: *"For many years now, bandwidth has been the basis for our file system procurements. We believe that our bandwidth requirement is a function of platform memory. Typically our ratio of file system bandwidth to platform memory (GB/s per TB of memory) has varied from 0.6 to 0.8. Currently that ratio is at 0.5GB/s/TB in the OCF and 0.6GB/s/TB in the SCF. Note that the Sequoia file system is currently 0.4GB/s/TB."* As they should, requirements definition methodologies have changed as architectures evolve and lessons are learned.

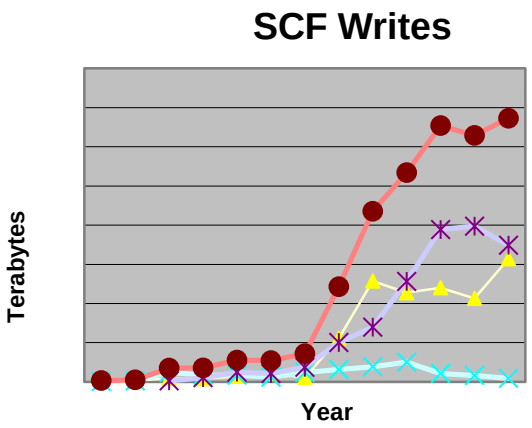
SCF Max Lustre Bandwidths

	lscratch1	lscratch2	lscratch3	lscratch4	lscratch5
Coastal	15GB/s	15GB/s	15GB/s	40GB/s	40GB/s
CSLIC	1.25GB/s	1.25GB/s	1.25GB/s	1.25GB/s	1.25GB/s
Eos	10GB/s	10GB/s	10GB/s	10GB/s	10GB/s
Gauss	15GB/s	15GB/s	15GB/s	15GB/s	15GB/s
Graph	15GB/s	20GB/s	15GB/s	15GB/s	15GB/s
Inca	125MB/s	125MB/s	125MB/s	125MB/s	125MB/s
Juno	15GB/s	15GB/s	15GB/s	40GB/s	40GB/s
Minos	15GB/s	15GB/s	15GB/s	30GB/s	30GB/s
Muir	15GB/s	15GB/s	15GB/s	40GB/s	40GB/s
Rhea	15GB/s	15GB/s	15GB/s	20GB/s	20GB/s

- File System Disk Limited
- Network/Lustre Router Limited
- Node limited (i.e. Single node job scheduling)

The Blueprint is also the document where LLNL outlined an initial plan to address exponential archive growth and associated unmanageable out-year costs. The LLNL Archive Quota implementation was planned as a first mitigation. Archive advisory quotas were implemented in

December of 2010. Initial talks with users were held beginning in August of 2009. Archive growth rates have slowed. We conjecture that this slow down is due to a number of factors, not just the quota implementation. First, simply communicating the cost of storing a particular user's data resulted in that user deleting over 1PB of data in the archive. Raising awareness of costs is a best practice that we have used with very good results. At LLNL, nothing is archived automatically; all transfers are initiated by users. Activities that increase storage to the archive include aggressive global parallel file system purge policies, file system retirements and planned file system down times. A reduction in one causes a reduction in the other. Finally, the biggest factor in archive growth is platform memory capacity. As new large platforms are added, archive growth increases. We expect substantial archive growth with Sequoia and we expect the archive advisory quota implementation to help contain the rate of growth over time. While we expect the Quota implementation to help, it's too early to claim it as a best practice.



CONCLUSIONS

Providing a balanced infrastructure to optimize user productivity, while minimizing costs, requires attention and focus in a number of areas. The cycle of events includes formalized planning, which is informed by regular collection of data and metrics, user feedback, advanced technology investigations and testbed evaluations. Strong vendor partnerships and in-house software expertise are key enablers to quickly moving forward and providing the best environment possible.