Lawrence Livermore National Laboratory

Level-2 Milestone 5589: Modernization and Expansion of LLNL Archive Disk Cache

Milestone Report for NNSA HQ

Prepared by Jerry Shoopman

February 4, 2016

LLNL-TR-682048

Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344.

Disclaimer

Contents

Contents

Introduction

This report documents Livermore Computing (LC) activities in support of ASC L2 milestone 5589: Modernization and Expansion of LLNL Archive Disk Cache, due March 31, 2016. The full text of the milestone is included in Attachment 1. The description of the milestone is:

> ***Description****: Configuration of archival disk cache systems will be modernized to reduce fragmentation, and new, higher capacity disk subsystems will be deployed. This will enhance archival disk cache capability for ASC archive users, enabling files written to the archives to remain resident on disk for many (6–12) months, regardless of file size.*

The milestone was completed in three phases. On August 26, 2015 subsystems with 6PB of disk cache were deployed for production use in LLNL's unclassified HPSS environment. Following that, on September 23, 2015 subsystems with 9 PB of disk cache were deployed for production use in LLNL's classified HPSS environment. On January 31, 2016, the milestone was fully satisfied when the legacy Data Direct Networks (DDN) archive disk cache subsystems were fully retired from production use in both LLNL's unclassified and classified HPSS environments, and only the newly deployed systems were in use.

LLNL Archive Disk Cache System Design

The massive scale of recent LC platforms and file system expansions required the acquisition, integration, tuning, and production deployment of correspondingly capable archive disk cache systems in both the unclassified and classified environments. Successfully integrating these systems with the existing LC environment was critical to the ASC missions that these archival disk cache resources support. In particular, the target disk residency for files written to new archive disk caches was six months or more.

The new archive disk cache subsystems, subsequently named Yuban in the unclassified and Osprey in the classified, were the product of multiple procurements. Seven separate procurements were completed for the initial disk hardware, subsequent disk expansions, HPSS Mover nodes and node expansions, as well as 40GbE Network hardware.

The hardware selected for the Yuban and Osprey archive disk cache systems consists of a total 80 NetApp E5500 60-disk enclosures with 40 NetApp E5500 controller pairs and a total of 4,800 individual 4TB hard drives, and 68 Haswell server nodes with quad-port Qlogic FC8 Host Bus Adapters (HBAs) and Intel 40GbE Network Interface Cards (NICs). The server nodes, disk controllers/enclosures and disks are installed in 10 Rack

Scalable Storage Units (RSSUs). The 4 RSSU Yuban unclassified system is pictured in Figure 1 below. Cisco 40GbE switch hardware was selected to provide high-bandwidth network connectivity between HPSS Mover nodes and LC platforms. The switches are in a single rack collocated with each of the two Yuban and Osprey clusters. The Yuban unclassified 40GbE switch rack is pictured to the right of Yuban in Figure 1 below.

Figure 1: Yuban Archive Disk Rack Storage Scalable Units (four RSSUs) plus 40GbE switch on the right (not pictured: Osprey 6 RSSUs plus switch rack)

The capacity and performance requirements of Yuban and Osprey dwarfed existing LC archive disk cache systems (see Figure 2 below). Early in the deployment cycle, it was clear that the largest challenge for Yuban and Osprey would be the availability, reliability, and performance of the Haswell servers configured as HPSS Mover nodes. In fact, several major challenges were met with these servers. Several minor outstanding issues – with no significant impact to users – continue to be investigated.

Archive Disk Cache System	Capacity (PB)	Avg. Disk Residency for Files
OCF Sumatra - orig.	1.1PB	~30 days
OCF Yuban - new	6PB	230+ days and climbing (as of 2/3/2016)
SCF Martin - orig.	1.1PB	<30 days
SCF Osprey - new	9PB	134+ days and climbing (as of 2/3/16)

Figure 2: Yuban and Osprey Capacity and Average Residency vs. Legacy Systems

A multi-year effort was launched to enable both existing and new disk cache systems to be used beneath HPSS until the new disk cache systems could be fully integrated. At milestone completion the aging, legacy DDN disk subsystems were retired. LC chose NetApp E5500 disk systems for high capacity, reliability, and availability and performance reasons. In addition, the RSSU approach enabled the Data Storage Group (DSG) to – for the first time – install HPSS Mover server nodes in the exact same rack as the attached disk subsystem, enabling simpler direct-attach FibreChannel (FC) connections and removing the need for and complexity of additional FC trunks from servers and disk controllers between FC patch panels. This standard RSSU approach also simplifies day-to-day support procedures for LC Operations Staff, who are experts with the NetApp disk hardware which is deployed widely across LC for Lustre file systems. Equally important, the disk procurements leveraged LC negotiating strengths with the disk vendor and provided significant, long-term price-performance benefits.

LLNL Archive Disk Cache High-Level Design

Archive Disk Cache hardware deployment occurred through much of the second half of FY15 and finished in the first half of FY16. The early selection and purchase of standard RSSU hardware components enabled this process to proceed smoothly. Disk hardware performance matched vendor specifications.

Although server subsystem reliability didn't initially meet original expectations, early critical problems have been worked to resolution. The most systemic problem with high impact was that the server nodes were shipped with hard disks that required 4K block boundary Operating System (OS) installation. Determination that the default OS install procedure didn't account for this, and that nodes were installed with 512 byte boundaries was time consuming. After this discrepancy was identified, OS software was reinstalled across both the Yuban and Osprey clusters, and the vast majority of the performance issues with the servers were resolved. The sophisticated design that allows for HPSS Mover failover enables the server nodes to be "failed out" of production use for reinstallation and/or service actions without user interruption or archive service degradation (see Figure 3 below).

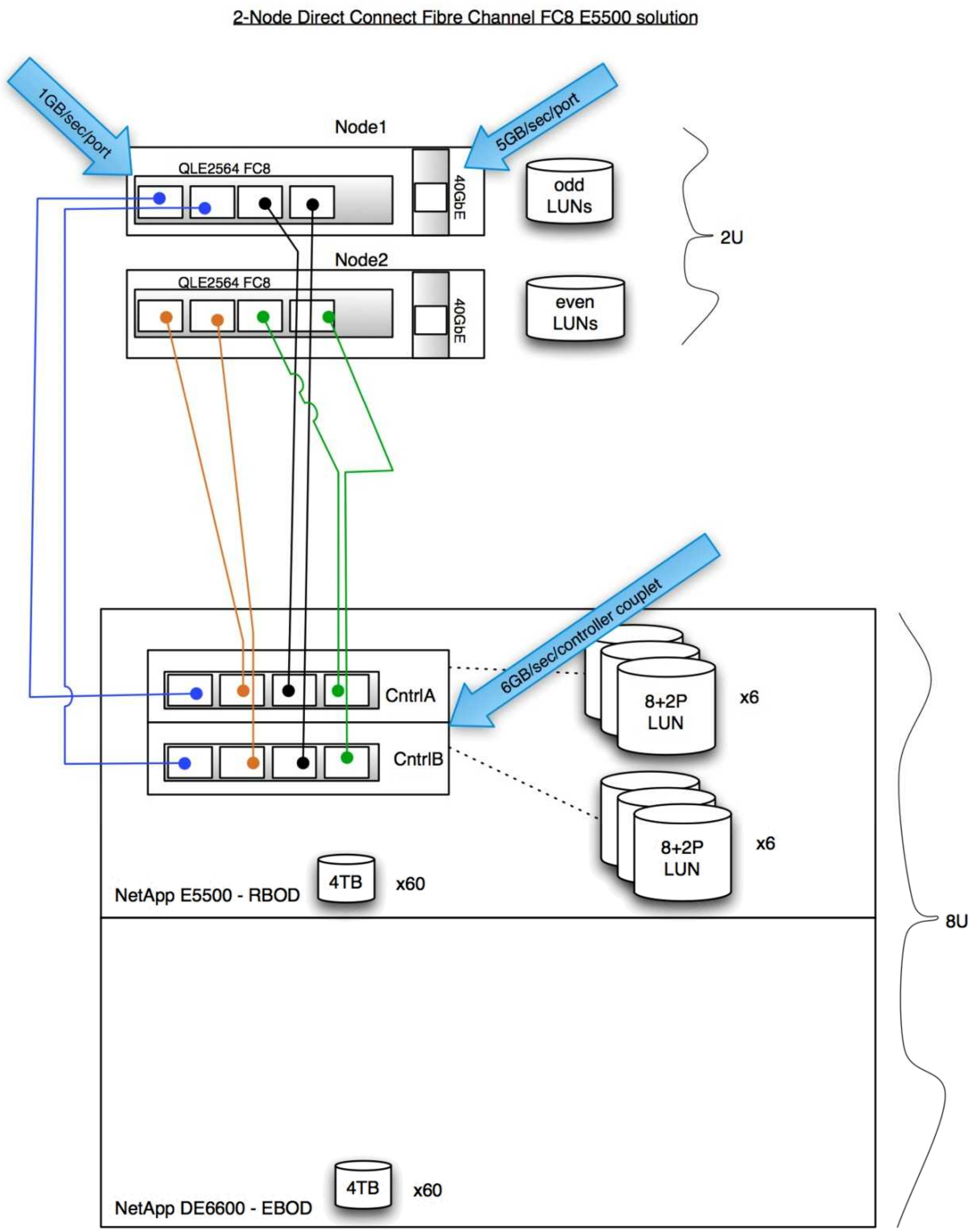

Figure 3: Dual-node Direct-attach FC8 NetApp E5500 Solution

LLNL Archive Disk Cache Detailed Configuration

Throughout the deployment period, system administrators and vendor personnel refined administration and monitoring aspects of the system. This work provided valuable insight to administrators and network engineers who, for example, had yet to deploy 40GbE into production environments. This allowed administrators and vendors to improve their tools to accommodate installations in the LC environment. This time was also spent refining and tuning the HPSS environment to fully optimize the cache subsystems. In one example, performance testing and monitoring demonstrated that the Storage to Lustre Interface Clusters (SLICs) were subject to an aggregate bandwidth that is significantly less than it was with the 10GbE subsystems, due to the 40GbE Yuban and Osprey clusters being connected to a different core switch than the SLICs. The SLICs are connected to the previous HPSS 10GbE core switch. Only a limited number of 10GbE connections exist between this switch and the newly deployed HPSS 40GbE switch. This artificially limits total bandwidth to the Yuban and Osprey systems from LC platforms to the bandwidth capability of the inter-switch trunks. LC Network Engineers have scheduled installation of new fiber trunks from the SLICs that can be directly attached to the 40GbE HPSS switches in both the unclassified and classified environments, in order to increase bandwidth to expected 40GbE backplane speeds.

Two outstanding HPSS Mover node issues are currently being investigated, and are expected to be of minimal impact to the overall service due to the failover design:

1) A small number of server nodes sporadically reboot with no apparent hardware failure associated with the reboot recorded in server diagnostic output. Collaboratively, the LC local Platform Architect and DSG members are testing a theory that the C-States (OS and motherboard CPU power management) logic is at fault. The working theory is that C-States erroneously detect a non-issue which subsequently triggers an unnecessary reboot. The C-States feature has been disabled in a spare, non-production system, and local testing is underway to deterministically qualify the root cause defect. Configuration changes to production clusters will follow if the defect is confirmed.

2) The 40GbE Network Interface Cards (NICs) often demonstrate packet loss after a server node is rebooted, often following the above-stated issue. This 40GbE failure mode presents as data transfer errors to the particular HPSS Mover node. The DSG has developed a Host Monitoring (HM) utility to proactively detect and monitor this failure mode, and an operational workaround is in place. When HM displays a 40GbE failure, LC Operations Staff are trained to reseat the 40GbE NIC optic which resolves the problem. The root cause of this problem is also collaboratively being worked in coordination with LC Platform Architect, the LC System Administration Group, LC Network Engineering Team and multiple vendors. Currently, a more recent 40GbE NIC driver is being tested on two different production nodes. It's anticipated that new driver will be deployed to production systems by the end of February, 2016.

It has been many years since DSG analyzed the distribution of file sizes across the legacy disk cache architecture and configuration of HPSS COSs and SCs. A thorough analysis of the smallest files was of the utmost importance, as they are the most problematic. Figure 4 and Figure 5 below illustrate the distribution of HPSS file sizes in the 0-4MB range.

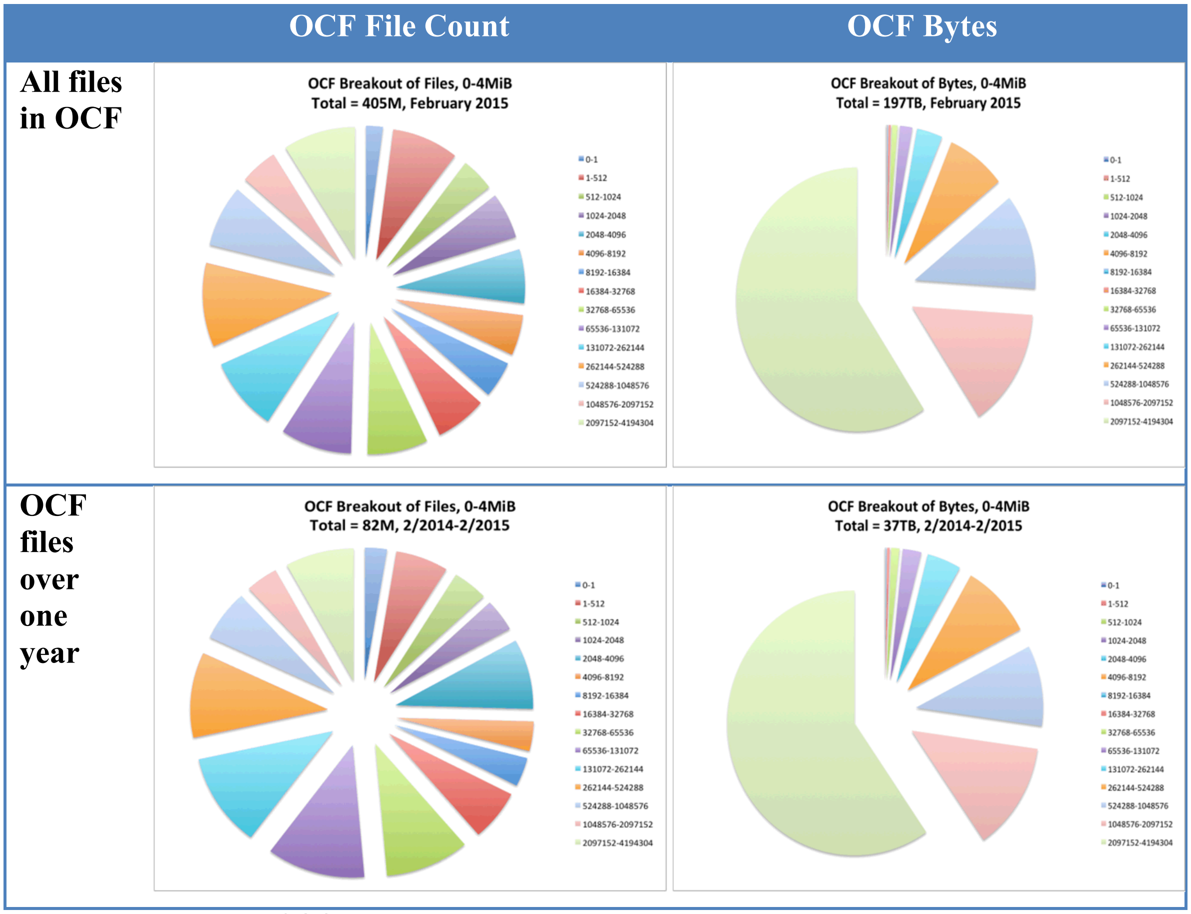

Figure 4: Breakout of OCF Files in the 0-4MB Range

For each production system, two sets of graphs are shown (in two columns): The leftmost column graphs files by size in the archive and the rightmost column graphs breakout by bytes created within the last year. The graphs in the rightmost column help determine if recent user behavior had changed any significant element of the overall data profile. For each set of graphs, two rows of metrics are shown. The graphs in the topmost row illustrate the breakout based on all files and bytes in each archive and the bottom row illustrates breakout based on one year of archive create activity.

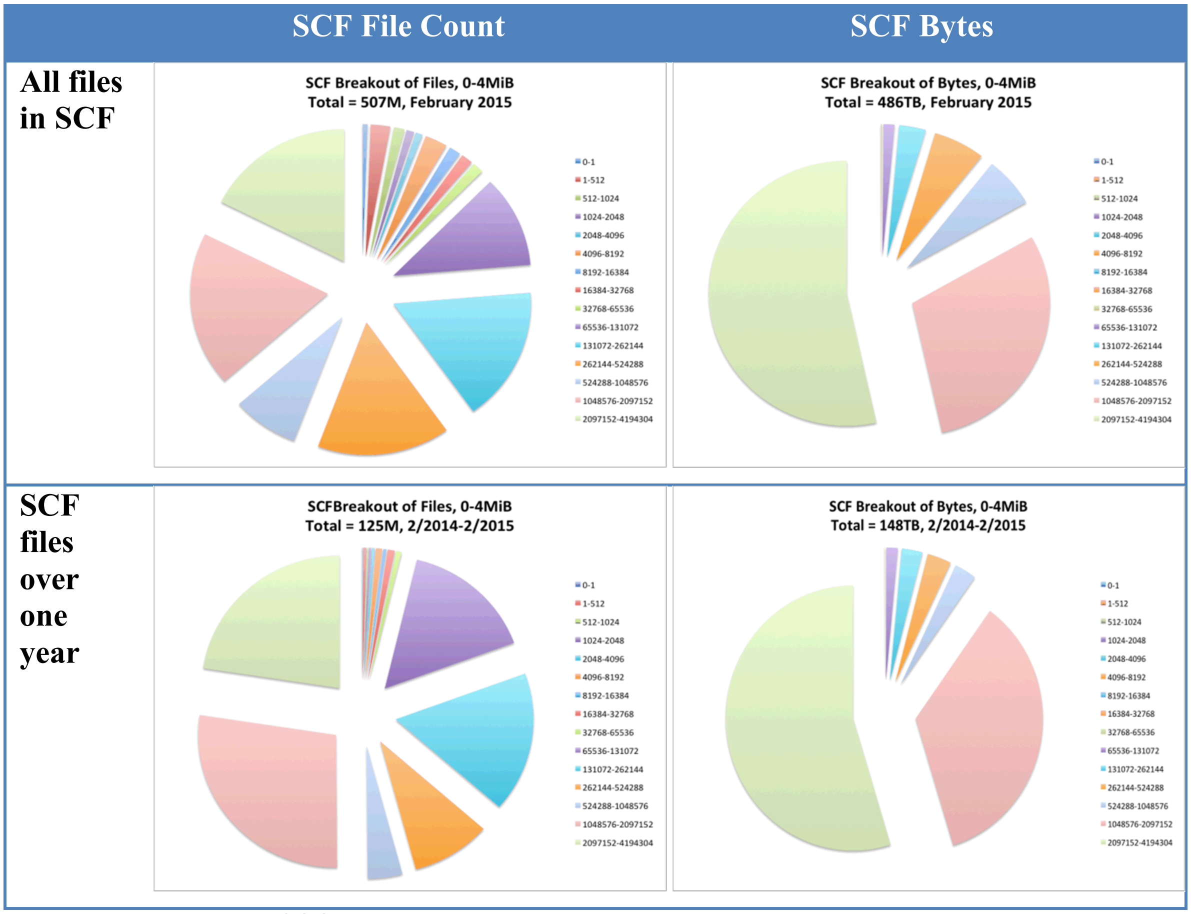

Figure 5: Breakout of SCF Files in the 0-4MB Range

Takeaway observations:

- OCF has a reasonable dispersion of files across the 0-4MB range
- SCF clearly has a higher count of the larger end of file sizes in the 0-4MB range
- Both OCF & SCF files in the 2-4MB range consume ~54-59% of the space
- SCF has significantly more files in the 1-4MB range
- In terms of change over the last year, OCF file size profile remains the same
- In terms of change over the last year, SCF file sizes grew in the 64KB-4MB range

In terms of efficiency, the primary goal with this deployment is to retain the maximum number files on disk across all disk storage classes for a longer period of time. This improves time to first byte for archive users, improves user experience and time-to-compute for archive read-intensive compute jobs. Along with purchasing additional disk capacity, DSG performed a thorough examination of how efficiently the legacy HPSS configuration utilizes the existing DDN disk. Analysis confirmed that legacy HPSS settings include a minimum segment size of 1MB for files in the size range of 0-4MB in the Small Class of Service and Storage Class (COS/SC110). Therefore, even relatively tiny files (e.g. 1KB) consume 1MB of the HPSS Storage Server's disk allocation. Looking closer at metadata consumed in the charts above, extrapolation of the current minimum segment size of 1MB based on legacy file size breakouts (assuming all files are disk resident) is illustrated in Figure 6:

	HPSS Bitfile Bytes	Storage Server Allocated Bytes	Overhead
OCF Production	**486TB**	**816TB**	**40%**
SCF Production	**197TB**	**551TB**	**64%**

Figure 6: Efficiency and Overhead

In the face of modernizing hardware, a new HPSS COS/SC configuration was warranted. Improvements in efficiency could be realized by reducing the minimum segment size. Performance could also be preserved (or even enhanced) by clearly understanding the new, underlying disk hardware and the size of I/O that performs well with that hardware. The chart in Figure 7 below illustrates the wasted space that would exist across varying HPSS minimum segment size configurations.

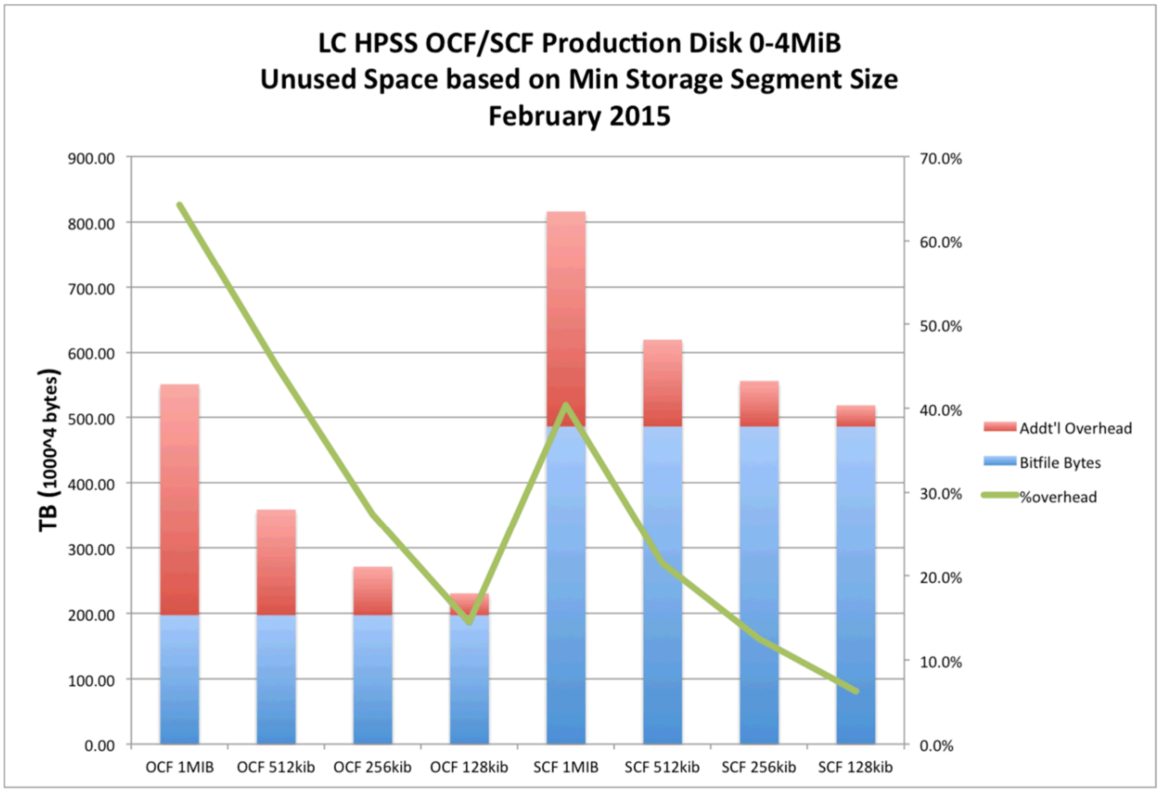

Figure 7: Various Configuration Options Considered

The underlying disk devices have some bearing on configuration choice from a performance perspective, as to their preferred I/O size. Namely, the data length size of the read() or write() system call run against them by the HPSS application. DSG performed extensive testing against the purchased NetApp E5500 systems. The goal was to see which LUN stripe size performed optimally across varying block sizes during the I/O. DSG performed both direct (unbuffered) and non-direct (buffered) I/O to a series of NetApp LUNs created using predetermined stripe sizes for the array groups in the NetApp E5500 to which these LUNs belonged. HPSS developers indicated that HPSS I/O would most likely be approximated by direct (O_DIRECT) I/O.

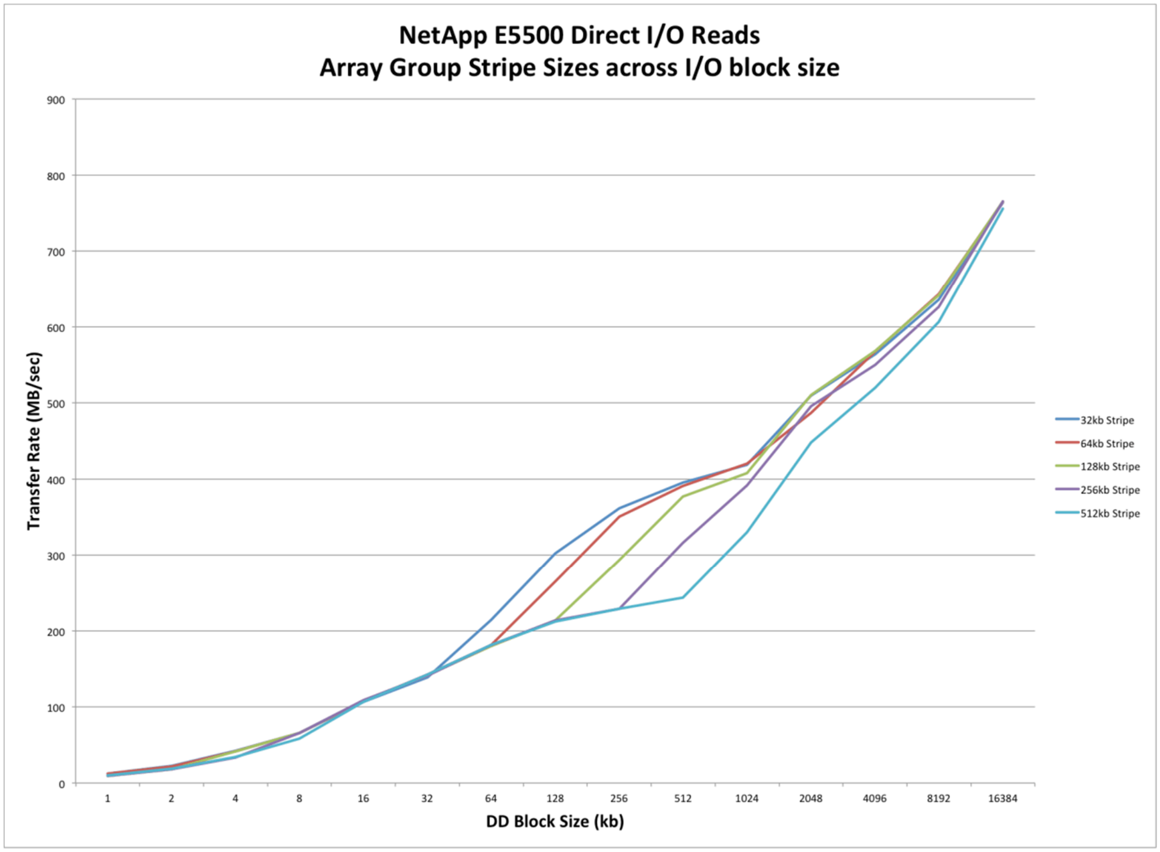

Figure 8: Direct I/O Reads

Figure 8 above illustrates the direct I/O read performance. As the I/O block size increases, performance increases. It is of interest that the various stripe sizes for the arrays ramp up in performance in a staggered fashion in the middle of the chart. This is due to the block size overcoming the penalty incurred by the read-modify-write RAID stripe length problem.

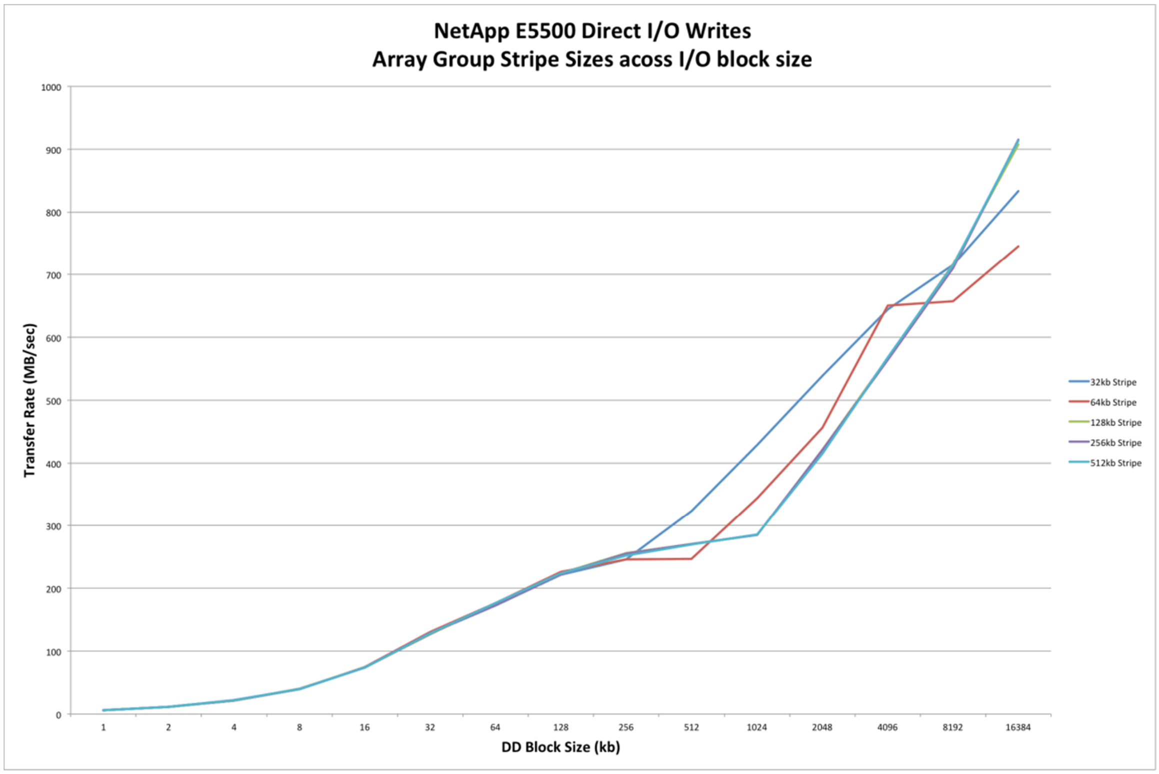

Figure 9: Direct I/O Writes

Figure 9 above illustrates the direct I/O write performance. Direct writes respond the same way to increasing the block size of the I/O—that is, smaller array group stripe sets respond well as block size increases (again, overcoming the read-modify-write penalty).

To identify the optimal Small COS configuration, many tests were performed. In the final analysis, DSG created 10,000 files from 1KB to 1MB in size, charted in Figure 10 below. These files first existed in Vulcan's lscratchv Lustre file system for the two leftmost tests, but that file system proved to be highly variable in performance and availability, so DSG moved the source tree to a RAM file system (/dev/shm). This introduced far more stability in the test results and guaranteed that the source file system would not be the gating factor.

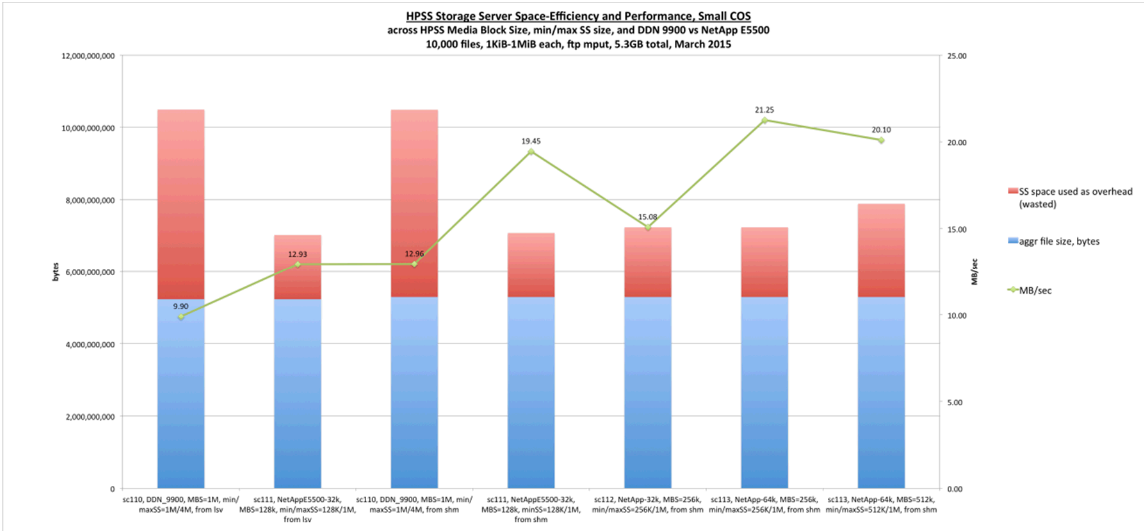

Figure 10: HPSS Storage Server Efficiency and Performance test, Small COS

The resulting dual-axis chart in Figure 10 above illustrates both efficiency in terms of wasted space for the given size of user bytes stored (left vertical axis) and average performance for the write of those 10,000 files (right vertical axis). Each bar is an average of 4-6 test runs.

The best performing configuration for files 1KB-1MB is test SC113 configured against NetApp 64k striped LUNs with an HPSS Media Block Size of 256KB. This configuration results in an average of 21.25MB/sec bandwidth. The same configuration using 512KB HPSS mover media block size and HPSS Minimum Storage Segment size performed at a slower 20.1MB/sec. The most efficient in terms of disk usage was SC111, which has a wasted-space overhead of 25% for this set of files. The fastest performing configuration (SC113) has a wasted-space overhead of 32%. The DDN 9900 production disk SC110 has a wasted-space overhead of 50% for this same set of files.

To compare performance on the Medium and Large COSs, DSG used two different sets of source files: 1,000 files of random sizes from 1MB to 10MB (5.7GB total) and 219 files of the same 256MB size (58.8GB total). Results are illustrated in Figure 11 below.

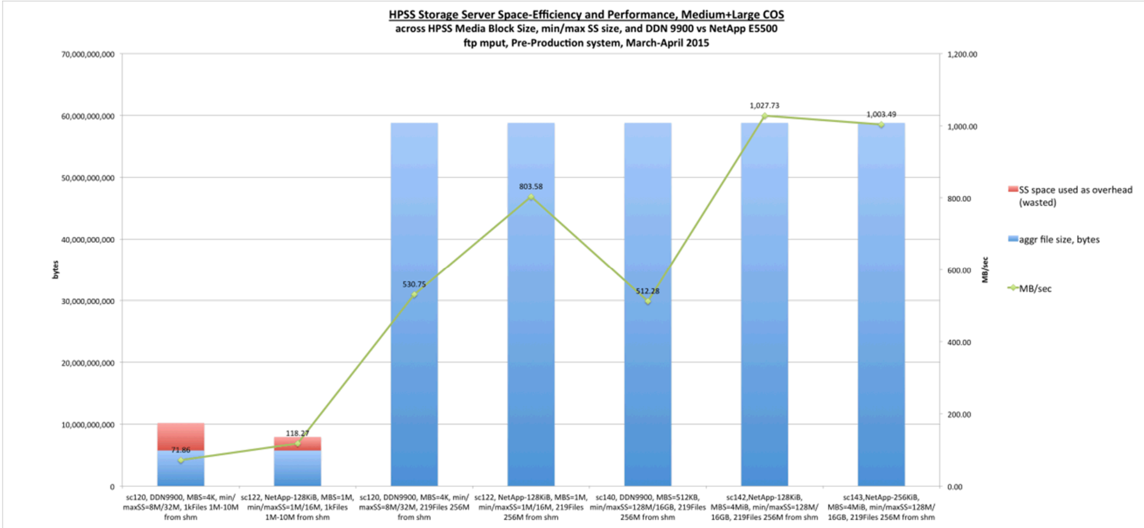

Figure 11: HPSS Storage Server Efficiency and Performance test, Medium and Large

The x-axis columns (marked either SC120 or SC122) are Medium COS/SC sized files. The columns marked SC140, SC142, or SC143 are Large COS/SC sized files.

Tests demonstrate that SC122 outperforms the current DDN 9900 SC120 configuration.

Writing 256MB files to the Medium COS outperforms current 4-wide Large COS/SC140 (804MB/sec vs. 512MB/sec).

Lastly, the NetApp E5500 disk configured with 128KB stripe segments outperforms the same otherwise-configured HPSS storage classes, but with 256KB stripe segment sizes.

The initial resulting new COS/SC configuration is illustrated in Figure 12 below.

COS/SC	Allocation Method	COS Min. File Size	COS Max. File Size	SC Media Block Size	SC Virtual Volume Block Size	SC Average Number Segments	COS/SC Optimum Access Size	COS Transfer Rate	NetApp LUN Segment Size	SC Min SS Size	SC Max SS Size
Small	Fixed Length, Classic Style	0 bytes	1 MB	256 KB	256KB	1	512KB	50000	64 KB	256KB	1 MB
Medium	Fixed Length, Classic Style	1 MB	256 MB	1MB	1 MB	1	16MB	100000	128 KB	1MB	64 MB
Large	Fixed Length, Classic Style	256 MB	100 TB	4 MB	64 MB	2	256MB	400000	128 KB	256 MB	16 GB
Mission Critical	Fixed Length, Classic Style	1 byte	100TB	-	-	-	256MB	400000	-	-	-

Figure 12: NetApp HPSS COS Configuration

The new HPSS COS/SC layout exacerbated a problem already understood by DSG. This is the issue of large files written from desktops and other non-LC systems that do not have LC-deployed transfer utilities (e.g. pftp, hsi, htar, etc.). LC-deployed clients intelligently interact with HPSS to indicate the size of an incoming file. Consequently, non-LC client files are written to the default COS, which was COS120 - Medium Dual Copy. As a result, those files are subject to the maximum segment count of 10,000 at the minimum segment size of 1MB. This means newly created files written from such hosts could not be larger than 10GB. With the previous minimum segment size of 8MB, the maximum size file was 80GB—still far less than the LLNL HPSS system's configured maximum file size of 100TB when the client knows how to intelligently interface to HPSS.

DSG explored other HPSS configuration options including the Variable Length Storage Segment (VLSS) feature to create a new COS to handle non-LC client transfers. After testing, it was determined that VLSS could be deployed and used for the Default COS. The VLSS feature allows HPSS to recognize when files are consuming segments at the minimum segment size, and immediately increase the size of subsequent segments (by powers of two) up to the maximum segment size (64 MB for the new default COS/SC). With the VLSS COS, users outside LC can now write files up to 600 GB in size without using any special commands, and avoiding the unintended limitation of 10GB maximum file sizes.

The final HPSS COS configuration is illustrated in Figure 13 below.

COS	Name	Hierarchy	Default	Allocation Method	File Min/Max	Truncate Final Segment	Enforce Max File Size	Force Select	Auto-chcos on Migration
110	Small Dual Copy	110 - Small Dual Copy	No	Fixed Length, Classic Style	0/1MB	Yes	Yes	No	No
120	Medium Dual Copy	120 - Medium Dual Copy	No	Fixed Length, Classic Sytle	1048577b/256MB	Yes	Yes	No	**Yes - files larger than 256MiB get moved to COS 160
125	**VLSS Catch-All**	**120 - Medium Dual Copy**	**Yes**	**Variable Length**	**0/600GB**	**Yes**	**Yes**	**No**	****Yes - files larger than 256MiB get moved to COS 160**
160	Large Single Copy	160 - Large Single Copy	No	Fixed Length, Classic Style	268435457b/100TB	Yes	Yes	No	No

Figure 13: Final HPSS COS Configuration with VLSS

Milestone Completion

On January 31st, 2016, legacy archive disk cache subsystems were fully retired from production use in both LLNL's unclassified and classified HPSS environments, and only the new NetApp Yuban and Osprey disk cache systems deployed on September 29, 2015 were in production use, satisfying this L2 milestone.

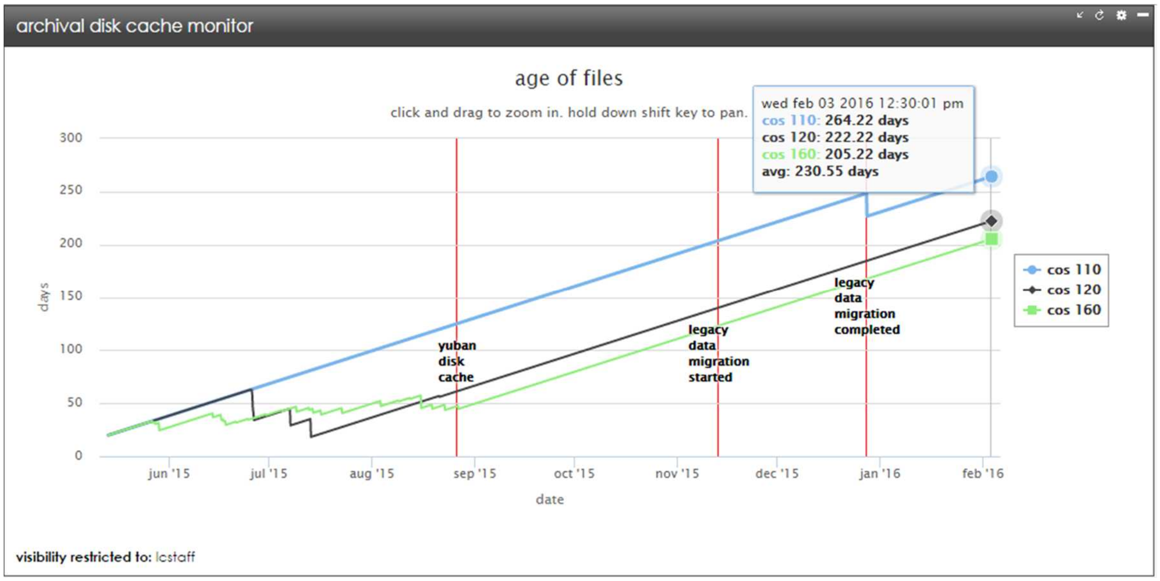

Figure 14: OCF Archival Disk Cache Monitor

The chart in Figure 14 was developed by DSG and is available to LC Staff on MyLC in both the unclassified and classified environments. This chart demonstrates OCF files written to Yuban in the last 230+ days remain resident on disk, regardless of size. The current SCF average is 134+ days, which is approximately the length of time the Osprey system has been in production service.

At this point in time, user files written to the Yuban disk cache system in the unclassified archive environment are staying disk resident for well over six months and that duration continues to climb. Files written to the Osprey disk cache system in the classified archive environment are currently staying resident over four months on average, even with the higher demand from classified platforms. These average residency periods fluctuate relative to user write activity. When fewer files are being written over a given period those files will stay disk resident for a longer duration. As the number of files written increases, their disk life residency tends to decrease.

On February 4th, 2016 a formal L2 milestone handoff letter to ASC Program Director Michel McCoy was signed by Jerry Shoopman LLNL (see Attachment 2).

Attachment 1: Milestone Definition Text

Milestone (5589): Modernization and Expansion of LLNL Archive Disk Cache		
Level: 2	**Fiscal Year**: FY16	**DOE Area/Campaign**: ASC
Completion Date: 3/31/16		
ASC nWBS Subprogram: CSSE		
Participating Sites: LLNL		
Participating Programs/Campaigns: ASC		
Description: Configuration of archival disk cache systems will be modernized to reduce fragmentation, and new, higher capacity disk subsystems will be deployed. This will enhance archival disk cache capability for ASC archive users, enabling files written to the archives to remain resident on disk for many (6–12) months, regardless of file size.		
Completion Criteria: Archival disk cache systems are modernized, and new, higher capacity disk subsystems are deployed.		
Customer: ASC		
Milestone Certification Method: Professional documentation, such as a report or a set of viewgraphs with a written summary, is prepared as a record of milestone completion. The "handoff" of the developed capability (product) to a nuclear weapons stockpile customer is documented.		
Supporting Resources: LLNL FOUS staff		

Attachment 2: Handoff Letter

ADVANCED SIMULATION & COMPUTING™

February 4, 2016
ICCD16-033

Memorandum

To: Michel McCoy, ASC Program Leader

From: Jerry Shoopman

Subject: Completion of LLNL ASC Level 2 Milestone 5589

I certify that the LLNL ASC Level 2 Milestone 5589 "Modernization and Expansion of LLNL Archive Disk Cache" was completed on January 31st, 2016.

Following extensive procurement, development, deployment, integration and testing efforts the archive disk cache subsystems, known as Yuban and Osprey, were made available for unclassified and classified utilization. On January 31st, 2016, after legacy disk cache systems were fully retired from HPSS production archives, Yuban and Osprey were the only archive disk cache systems on the unclassified and classified HPSS environments, thereby satisfying the requirements of the L2 milestone.

The details of the efforts required to meet this milestone are documented in "Level-2 Milestone 5589: Modernization and Expansion of LLNL Archive Disk Cache" (February 4th, 2016).

Jerry Shoopman
Data Storage Group Leader, LCD

Cc: File

Lawrence Livermore National Laboratory