LUSTRE USER GROUP
TACC SITE UPDATE

Tommy Minyard, Ph.D.
Director of Advanced Computing Systems
Frontera Co-PI and Chief Systems Engineer
PRESENTATION OUTLINE

- Intro to TACC
- Frontera Lustre filesystems
- Stampede2 Lustre filesystems
- Stockyard sitewide filesystem
- Lustre operational challenges
- Testing of flash-based parallel filesystems
TACC AT A GLANCE - 2022

**Personnel**
185 Staff (~70 PhD)

**Facilities**
12 MW Data center capacity  
Two office buildings, Two Datacenters, two visualization facilities, and a chilling plant.

**Systems and Services**
> Seven Billion compute hours per year  
> 5 Billion files, >100 Petabytes of Data, NSF Frontera (Track 1), Stampede2 (XSEDE Flagship), Lonestar6, Chameleon (Cloud Testbed) system

**Usage**
> 15,000 direct users in >4,000 projects,  
> 50,000 web/portal users, User demand 8x available system time. Thousands of training/outreach participants annually
FRONTERA HARDWARE SUMMARY

- **Compute nodes:**
  - 8,008 Dell C6420 servers, dual-socket Intel 8280 28-core processors, 192GB, HDR100 IB
  - 396 Dell R640 servers, dual-socket Intel 8280 28-core processors, 192GB, HDR100 IB
  - 16 Dell R840 servers, quad-socket Intel 8280 28-core processors, 384GB, HDR100 IB, 6TB NVDIMM
  - 90 GRC GPU oil-immersion servers, four NVidia RTX5000 cards per node, FDR IB

- **Mellanox HDR InfiniBand interconnect** (200Gbps core, 100Gbps to nodes)

- **Storage subsystems:**
  - Four DataDirect Networks 18K Exascaler storage arrays, 56PB storage, 300 GB/s bandwidth
  - 72 DDN IME flash servers, 3PB storage, 1.5TB/s bandwidth
  - Stockyard (/work) sitewide filesystem, 6.2PB, 80 GB/s bandwidth
  - Ranch archival subsystem, 100PB+ of tape capacity
STORAGE SUBSYSTEM

- DataDirect Networks (DDN) Spinning Storage, 56PB capacity, 300GB/s bandwidth
  - Four ES18KX EXAscaler Lustre storage appliances
    - Pair of controllers configured for high availability with active/active failover
    - Eight Lustre object storage servers, capable of delivering 75GB/s
    - 1,440 10TB drives in declustered RAID in sixteen SS9012 90-drive SAS enclosures
- Six Metadata Server Units
  - Pairs of Lustre meta-data servers with active/active failover and DNE
  - Two DDN SFA200NV arrays to support up to 6 billion inodes
- DDN Infinite Memory Engine (IME) flash storage, 3 PB capacity, 1.5 TB/s bandwidth
  - 72 DDN IME240 appliances with 21 2TB NVMe drives each and dual IB HDR100 cards
  - Ten racks with one GigE and one IB switch per pair of racks
ADDITIONAL STORAGE OPTIONS

- Stockyard2 provides /work site-wide filesystem via Lustre routers
  - DataDirect Networks 10PB Lustre filesystem, upgraded last year
  - Sixteen LNet routers to provide 80GB/s of bandwidth
  - Mounted on all logins and computes
- Ranch archival system
  - Quantum tape library with 30PB of DDN spinning storage, can hold 100PB+ of tape capacity, currently LTO-8
  - Frontera projects account for about 50% of archived data so far...
FRONTERA FILESYSTEMS

- Opted for multiple scratch filesystems
  - /scratch1 – general purpose where most users work out of, purged
  - /scratch2 – filesystem for problematic IO codes and IME backend
  - /scratch3 – most capable and limited to those requiring most bandwidth/capacity, many groups have 1PB+ of storage here
- Due to client count, also use Lustre for home filesystems
  - /home1, /home2 – identical configurations, rsynced copy of /home1 on /home2 just in case…
SYSTEM MONITORING – FILESYSTEM AND IB FABRIC

- Cluster monitoring to ensure efficient operation
  - TACC stats: performance monitor on nodes to collect statistics on per job basis
  - lltop and xltop: Lustre monitoring utilities to correlate jobs to I/O activity
  - ibtop: InfiniBand monitoring to correlate jobs to interconnect traffic
- Very proactive monitoring of Lustre meta-data servers and aggressive intervention, we’ve learned its best to get ahead of potential user-induced filesystem issues
- Frequent scans and logging of IB port errors, link drops, or cables degrading in speed
SCRATCH FILESYSTEM PURGING

- No purges of /scratch2 or /scratch3 yet, users cleaning up themselves
- /scratch1 filesystem purging tries to keep filesystem below 75% available capacity
- Current purge time at 50 days (since last file access)
- Purging statistics July 2021 – June 2022:
  - More than 1.6 billion files removed (1.4B during year 2)
  - ~10.6PB of space reclaimed on filesystem (~12PB during year 2)
  - Max overall usage peaked at 94%
- Currently using about 75% of available space on all three /scratch filesystems
  - almost 35PB in almost 2 billion inodes (29PB year 2)
STAMPEDE2 OVERVIEW

- **COMPUTE**
  - 3700 Dell C6320P/Intel S7200 Xeon Phi compute nodes, four per chassis, “Knights Landing” 7250SE CPU
  - 1736 Dell C6420 Xeon server nodes, four per chassis, dual socket Intel “Skylake” 8160 CPUs
  - 224 Dell R650 nodes, dual socket, Intel 8380 “Ice Lake” CPUs

- **NETWORK**
  - Intel Omni-Path

- **STORAGE**
  - Xyratex/Seagate/Cray/HPE ClusterStor 300
  - 18P usable /scratch filesystem
STORAGE SUBSYSTEM

- Xyratex/Seagate/Cray/HPE ClusterStor 300
  - 35 Scalable Storage Units (SSU)
    - Pair of servers configured for high availability with active/active failover
    - 84 10TB drives, 41 drives per LUN in declustered parity (GridRAID), two drives act as filesystem external journal
    - Each SSU designed to provide ~10GB/s of performance
  - 3 Metadata Management Units (MMU)
    - Pair of Lustre meta-data servers with active/active failover
    - Disk to support up to 4 billion inodes per MMU
  - 2 System Management Units (SMU)
    - Pair of management servers, primary and secondary
    - Used to configure and manage the filesystems
  - 6 I/O racks, each with two GigE and two OPA switches
STORAGE RACK LAYOUT
STORAGE FILESYSTEMS

- Cray/HPE CS300 storage provides two Lustre filesystems
  - Home: 2 SSUs, 1 MMU, 1 SMU; quota and backed up to archive
  - Scratch: 33 SSUs, 2 MMUs, 1SMU; no quota but purged, designed for >300GB/s bandwidth, 18PB usable capacity
- Stockyard storage system provides /work site-wide filesystem, updated earlier this year with new hardware
  - DataDirect Networks 8PB Lustre filesystem
  - Lustre LNet routers between OPA and IB fabrics
- Ranch archival tape archive, 30PB DDN spinning storage, Quantum tape library with 100PB capacity
STAMPEDE2 FILESYSTEM EVENTS

- Primarily drive failures:
  - 2023 - 31 drive replacements
  - 2022 - 113 drive replacements

- User caused events:
  - 2023 (9 events - so far!):
    - Too many metadata Req/s: 5
    - Writing to single target from many threads: 3
    - Sundry: 1
  - 2022 (23 events):
    - Too many metadata Req/s - 8
    - Writing to single target from many threads: 3
    - Sundry: 12

- Unfortunately (or fortunately?) for a LUG presentation our Stampede2 filesystem events are not lustre related

- Version changes:
  - Server side (lustre version... ? lost):
    - 3.0 SU16
    - 3.1-022
    - 3.4-040
    - 3.5-040.98
  - Client side:
    - lustre-client-2.7.19-12
    - lustre-client-2.7.22-1
    - lustre-client-2.9
    - lustre-client-2.10.1
    - lustre-client-2.11.0-200
    - lustre-client-2.12.8
STAMPEDE2 TRIPLE DRIVE FAILURE

- Two events:
  - August 2022
  - October 2022
- Both recoverable despite outage
  - No data-loss

- Personalities: [raid1] [raid6] [raid5] [raid4]
  - 303163760640 blocks super 1.2 level 6, 128k chunk, algorithm 1003 [41/39/2]
    - in: 105782353 reads, 3269265 writes; out: 2389519913 reads, 299179001 writes, 1504145 zwrites
    - recovery = 0.7% (267818432/37895470080) finish=10705.1min speed=58581K/sec
    - bitmap: 175/2259 pages [700KB], 8192KB chunk, file: /WIBS/homen04:md0/WIB_homen04:md0
OUR STATUS AS STAMPEDE2 APPROACHES EOL

- We’ve experienced the famous: “With disk capacity ever-increasing, RAID recovery based on disk rebuild becomes more costly, and more likely to trigger an additional failure”
  - Declustered parity was still not enough in the case of S2’s events
  - Proof-of-concept with GRAID – doesn’t integrate with our controllers-based model, very good if you’re DIY’ing it.
  - Possible early adopter issue – we’re unsure
  - “HDD average life span misses 3-year mark in study of 2,007 defective drives” - arstechnica

- Proof-of-concept project investigating storage-solutions on Flash
  - Lustre (on IME hardware)
  - Beegfs (on IME hardware)
  - WekaIO (on IME hardware)
  - VAST (on POC hardware)
THANKS!
ANY QUESTIONS?