U.S. DEPARTMENT OF
ENERGY

UNIVERSITY OF
CALIFORNIA

BERKELEY LAB
LAWRENCE BERKELEY NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

Lustre in a Condo Computing Environment

John White

Lustre User Group 2016

April 2016

High Performance Computing Services At the Lab

- Positioned to provide HPC to research groups seeking solutions at the institutional level, just shy of national allocations
- 17 Research Clusters
- 4 Institutional Compute Pools
- Condo Offering
- ~1750 nodes within central infrastructure
- ~1.75PB capacity between NAS and Parallel

High Performance Computing Services At UC Berkeley

- Partnered with UCB Research IT (RIT) to spin up the Berkeley Research Computing project
- One year from proposal to first production user
- Spun up Savio condo infrastructure to address shortcomings in the University's approach to research computing
- Currently on 2nd generation of condo spec
- First installation of our modern Lustre spec
- ~1PB capacity between NAS and parallel

The Condo

“The model for sustaining the Condo program is premised on faculty and principal investigators using equipment purchase funds from their grants or other available funds to purchase compute nodes (individual servers) which are then added to the Lab's Lawrencium compute cluster. This allows PI-owned nodes to take advantage of the high speed Infiniband interconnect and high performance Lustre parallel file system storage associated with Lawrencium. Operating costs for managing and housing PI-owned compute nodes are waived in exchange for letting other users make use of any idle compute cycles on the PI-owned nodes.”

The Condo

Infrastructure

- High Speed Interconnect (IB)
- Provisioning (Warewulf 3, version 4 in development)
- Scheduler/Batch Queue (Slurm)
- NAS Environment (Hitachi HNAS, Netapp vFiler)
  - Customer Buy-in Option @ ~$12K/40TB chunk

The Condo

Infrastructure

- Global Parallel Scratch File System (Lustre mainly)
- Software Module Farm
- Remote data transfer Infrastructure (Globus Online member site)
  - Considering spinning up a DTN hosting offering

The Condo

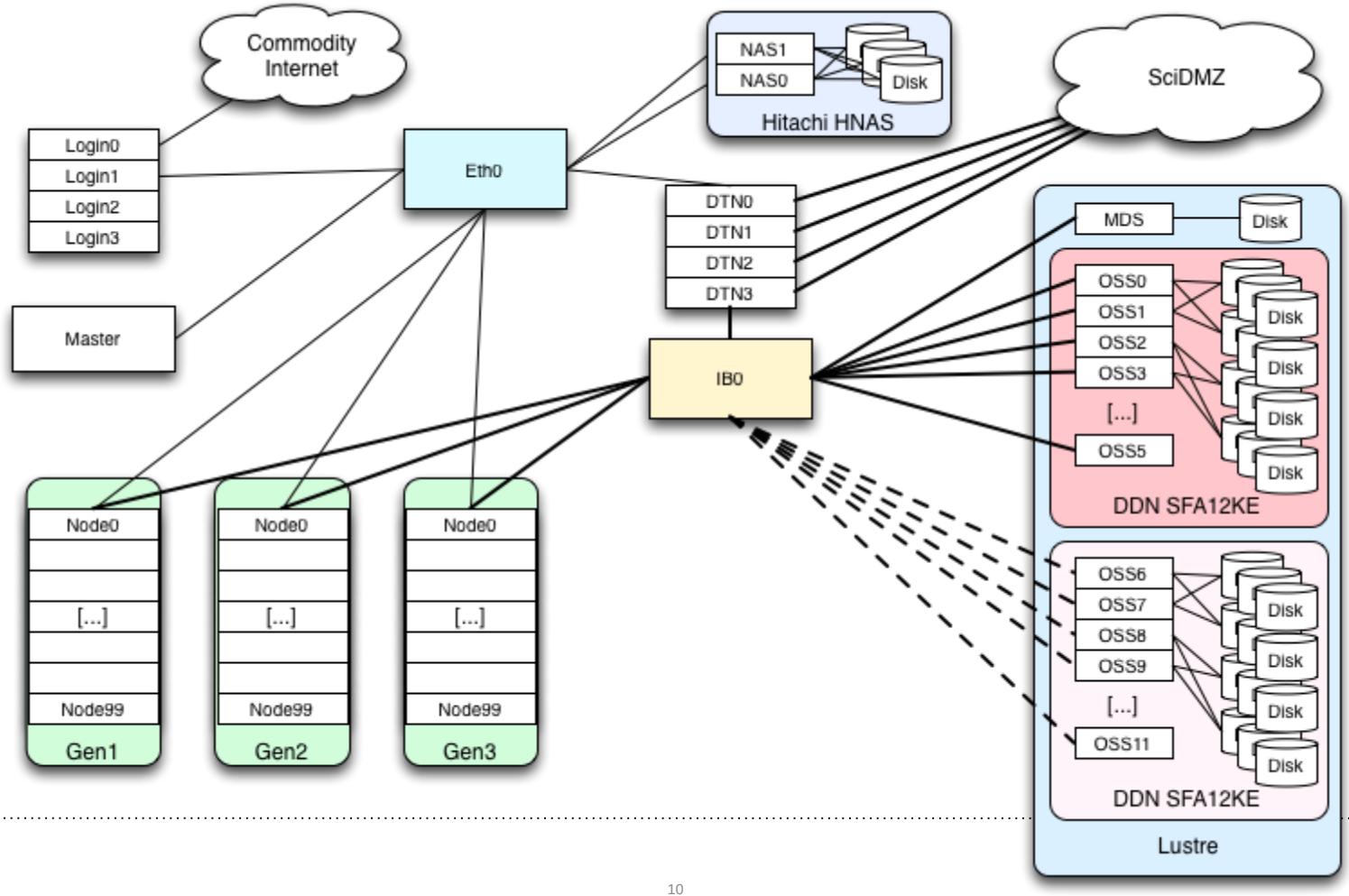
Current Compute Generation

- Dell C6320, Supermicro 2U Quad Node or Lenovo nx360m5
  - Dual Socket, 12-core 2.3Ghz E5-2670v3
  - 64GB (8x8GB) 2133Mhz DDR4 RDIMMs
  - FDR-14 (56Gb/s ConnectX3)
  - 500GB 7.2K RPM (local swap, log files)
  - 4 Year Warranty

The Condo

Current NAS Generation

- Hitachi HUS110
  - 12 x 4TB NLSAS 7.2K
  - RAID 6 (10+2P)
  - Controller Provided
  - Fronted by Dual Head HNAS 4060


Commodity Internet
Login0
Login1
Login2
Login3
Master
Eth0
NAS1
NAS0
Disk
Hitachi HNAS
SciDMZ
DTN0
DTN1
DTN2
DTN3
IB0
MDS
Disk
OSS0
OSS1
OSS2
OSS3
[...]
OSS5
DDN SFA12KE
OSS6
OSS7
OSS8
OSS9
[...]
OSS11
DDN SFA12KE
Lustre
Node0
[...]
Node99
Gen1
Gen2
Gen3

Procurements in the Condo

- Parallel storage is simply not tied to customer procurements
  - Benchmark for the lowest common denominator and hope for the best
  - Edge cases problematic and rarely do require on-off solutions
    - High IOP/Low IO size
    - Pure capacity demands
- NAS storage can be included in a procurement
  - No need to maintain stripe performance

Procurements in the Condo (cont'd)

- Procurements are incredibly simple
  - select the node count and type (highmem/gpu/vanilla)
  - TB of usable NAS capacity rounded to 12 drive increments
  - Rack, provision, accounts, create directories
  - Agreements in-place with vendors (per generation) for fixed pricing helps speed the process
    - Node and NAS

Common Lustre Environment

Previous Generation:

- DDN SFA 9900
- Dell 2950 (OSSs and MDSs)
  - 24GB RAM
  - IB-SRP Connected
  - QDR to the client / 10GbE for non-IB clients
  - DDN EF3010 MDT (FC)

Common Lustre Environment

Current

- DDN SFA 12KE20
  - 6 Embedded OSSs (5 OSTs per) - Exascalar OS
  - 40GB RAM
  - Dual Port ConnectX3
    - Split ports between IB and 10GbE
- Dell R610 MDS
  - Hitachi HUS110 (Flash) MDT

Improvements Between Generations

- Embedded
  - Huge equipment cost reductions
  - Management simplification
    - Uniformity across all installations frees up my time
    - Single point of documentation brings administration to the rest of the group
    - No more need to learn the HA du jour for each generation
    - All the scripts we would custom write already exist

Improvements Between Generations

- Counterintuitively, performance has been better than non-embedded
- See previous example of mystery performance issues
- Support Contract
  - SFAOS brings uniformity
  - Software Packaging
    - Still chasing the Kernel/OFED/Lustre compatibility chart
      - Single largest operational challenge
      - Debugging required with every new combination

A Mid-Production Predicament

- Initial SFA 12K20 install had some performance issues
  - Performance degradation when booted via PXE
  - Kept ¼ disk unprovisioned for testing
  - A year passes, file system fills
- Provisioning the remaining ¼ lead to some interesting (if predictable) performance issues of its own
- Weighted allocation algo kicked in aggressively (qos_threshold_rr and qos_prio_free)
- Performance of 22 OSTs -> 7 OSTs
  - Massive contention issues

The Storage Expansion Quandary

Post-Dec 2015 (March 7 2016)

UUID	Inodes	IUsed	IFree	IUse%	Mounted on
lr3-MDT0000_UUID	4000061520	145085606	3854975914	4%	/global/scratch[MDT:0]
lr3-OST0000_UUID	22405120	5868794	16536326	26%	/global/scratch[OST:0]
[...]					
lr3-OST0010_UUID	22405120	6448837	15956283	29%	/global/scratch[OST:16]
lr3-OST0011_UUID	22405120	5271623	17133497	24%	/global/scratch[OST:17]
lr3-OST0012_UUID	22405120	6412920	15992200	29%	/global/scratch[OST:18]
lr3-OST0013_UUID	22405120	5889487	16515633	26%	/global/scratch[OST:19]
lr3-OST0014_UUID	22405120	5704929	16700191	25%	/global/scratch[OST:20]
lr3-OST0015_UUID	22405120	6032511	16372609	27%	/global/scratch[OST:21]
lr3-OST0016_UUID	**22405120**	**1008576**	**21396544**	**5%**	**/global/scratch[OST:22]**
lr3-OST0017_UUID	**22405120**	**940563**	**21464557**	**4%**	**/global/scratch[OST:23]**
lr3-OST0018_UUID	**22405120**	**985627**	**21419493**	**4%**	**/global/scratch[OST:24]**
lr3-OST0019_UUID	**22405120**	**1011002**	**21394118**	**5%**	**/global/scratch[OST:25]**
lr3-OST001a_UUID	**22405120**	**994057**	**21411063**	**4%**	**/global/scratch[OST:26]**
lr3-OST001b_UUID	**22405120**	**1021931**	**21383189**	**5%**	**/global/scratch[OST:27]**
lr3-OST001c_UUID	**22405120**	**999257**	**21405863**	**4%**	**/global/scratch[OST:28]**

The Storage Expansion Quandary

Post-Dec 2015 (March 7 2016)

UUID	bytes	Used	Available	Use%	Mounted on
lr3-MDT0000_UUID	9.0T	321.1G	8.2T	4%	/global/scratch[MDT:0]
lr3-OST0000_UUID	21.4T	20.0T	337.7G	98%	/global/scratch[OST:0]
[...]					
lr3-OST0010_UUID	21.4T	20.0T	263.2G	99%	/global/scratch[OST:16]
lr3-OST0011_UUID	21.4T	20.0T	280.2G	99%	/global/scratch[OST:17]
lr3-OST0012_UUID	21.4T	20.0T	303.8G	99%	/global/scratch[OST:18]
lr3-OST0013_UUID	21.4T	20.0T	323.3G	98%	/global/scratch[OST:19]
lr3-OST0014_UUID	21.4T	20.0T	344.4G	98%	/global/scratch[OST:20]
lr3-OST0015_UUID	21.4T	20.0T	292.1G	99%	/global/scratch[OST:21]
lr3-OST0016_UUID	**21.4T**	**9.8T**	**10.5T**	**48%**	**/global/scratch[OST:22]**
lr3-OST0017_UUID	**21.4T**	**10.8T**	**9.5T**	**53%**	**/global/scratch[OST:23]**
lr3-OST0018_UUID	**21.4T**	**10.3T**	**10.0T**	**51%**	**/global/scratch[OST:24]**
lr3-OST0019_UUID	**21.4T**	**9.8T**	**10.5T**	**48%**	**/global/scratch[OST:25]**
lr3-OST001a_UUID	**21.4T**	**10.2T**	**10.1T**	**50%**	**/global/scratch[OST:26]**
lr3-OST001b_UUID	**21.4T**	**9.7T**	**10.6T**	**48%**	**/global/scratch[OST:27]**
lr3-OST001c_UUID	**21.4T**	**10.1T**	**10.2T**	**50%**	**/global/scratch[OST:28]**

The Storage Expansion Quandary

March 15 2016 - Need to bump qos_threshold_rr

UUID	bytes	Used	Available	Use%	Mounted on
lr3-MDT0000_UUID	9.0T	323.1G	8.2T	4%	/global/scratch[MDT:0]
lr3-OST0000_UUID	21.4T	20.2T	89.4G	100%	/global/scratch[OST:0]
[…]					
lr3-OST0010_UUID	21.4T	20.2T	82.6G	100%	/global/scratch[OST:16]
lr3-OST0011_UUID	21.4T	20.2T	69.3G	100%	/global/scratch[OST:17]
lr3-OST0012_UUID	21.4T	20.2T	108.7G	99%	/global/scratch[OST:18]
lr3-OST0013_UUID	21.4T	20.2T	107.3G	99%	/global/scratch[OST:19]
lr3-OST0014_UUID	21.4T	20.2T	110.2G	99%	/global/scratch[OST:20]
lr3-OST0015_UUID	21.4T	20.2T	90.2G	100%	/global/scratch[OST:21]
lr3-OST0016_UUID	**21.4T**	**10.7T**	**9.6T**	**52%**	**/global/scratch[OST:22]**
lr3-OST0017_UUID	**21.4T**	**11.5T**	**8.8T**	**57%**	**/global/scratch[OST:23]**
lr3-OST0018_UUID	**21.4T**	**11.1T**	**9.2T**	**54%**	**/global/scratch[OST:24]**
lr3-OST0019_UUID	**21.4T**	**10.7T**	**9.6T**	**53%**	**/global/scratch[OST:25]**
lr3-OST001a_UUID	**21.4T**	**10.9T**	**9.3T**	**54%**	**/global/scratch[OST:26]**
lr3-OST001b_UUID	**21.4T**	**10.5T**	**9.8T**	**52%**	**/global/scratch[OST:27]**
lr3-OST001c_UUID	**21.4T**	**10.8T**	**9.5T**	**53%**	**/global/scratch[OST:28]**

Expansion - One Big User

OST - IDX Count
0 - 122,771
1 - 139,788
2 - 126,491
3 - 127,313
4 - 119,314
5 - 137,899
6 - 115,083

OST - IDX Count
22 - 40,438
23 - 36,399
24 - 38,514
25 - 40,328
26 - 39,021
27 - 40,730
28 - 39,244

The Storage Expansion Quandary

- Quickly outgrowing Capacity
- Forklift upgrade unrealistic with next gen performance
- Doubling existing installation far more attractive

- The stripe problem…

The Next DDN Generation

- Less than double the current gen performance
- Mismatched disk (and LUN) size
  - Currently 3TB Spindles
- All variables lead to wide disparities in OST levels
- Tunables will clearly only get you so far

The Direct Doubling Problem

- Downtimes longer than 3 days are terribly hard to justify
- Restriping as part of a downtime is simply a non-starter
- Bolting on an identical system would leave a huge imbalance, one we're still struggling to rectify from last time

The Case for 4TB Spindles

- Currently (8*3TB)*29 = 696TB
- 4TB expansion (8*4TB)*40 = 1,280TB
- Assuming 560TB existing data:
  - Migrate existing data to new 1,280TB file system
    - ~14TB of data per OST, ~18TB free/OST
  - Integrate 3TB pools with new FS
    - Disparity between new FS and old ~10TB/OST (~3-6months to reach parity?)

Futures: Condoizing the Scratch File System

- Scale this experience into a true condo offering
- Constant state of expansion
- Unless we set storage pools per customer we either
  - Fill up oldest OSTs quick
  - Get hit with low stripe counts due to qos_threshold_rr and qos_prio_free (weighted) working aggressively
- Purchase capacity up-front and re-sell?
  - Not necessarily in our budget scope

Futures: DTN as a Service

- DTN Growth Rate is Outsized
  - Storage pools attached to detectors, etc
- Centralize Infrastructure
  - Benefit from wide striping
  - VLANs should make centralizing simple
- Funding?

Futures: Globus Online

- Drastically Reduce Duplicate Admin Time
- Google Drive Integration
- Combined With DTN as a Service
- Self-service Archive