Lustre deployment and early experiences

Florent Parent
Coordinator, Québec site
Outline

• What is CLUMEQ?
• CLUMEQ's new HPC cluster: *Colosse*
• Lustre experience
What is CLUMEQ?

- Consortium of 11 universities in the province of Québec, Canada
- Part of the *Compute Canada* national platform
- Two HPC sites:
  - Montréal
  - Québec City
Who is CLUMEQ?
CLUMEQ's mission

- To serve the HPC needs of its member institutions in all fields of research
- To outreach non traditional and emerging HPC fields
- To train "highly qualified personel" (HQP)
Enabling Canadian research excellence through high performance computing

Favoriser l’excellence en recherche au Canada avec le calcul de haute performance

PLANNED SYSTEMS
- Capability
- Capacity
- Vector
- Major Storage

Canadian Advanced Data Network
International Links

COMPUTE / CALCUL CANADA
Outline

- What is CLUMEQ?
- CLUMEQ's new HPC cluster: *Colosse*
- Lustre experience
CLUMEQ Colosse

- Sun constellation system
  - 10 fully loaded Sun Blade 6048, with X6275 modules (double Nehalem EP blade, 2.8GHz, 24GB of RAM)
  - full-bisection IB-QDR interconnect (2xM9 switches)
  - 1 PB of Lustre storage in a high availability configuration, using 2 MDS and 9x2 OSS
  - Sun J44000 storage arrays

- 86 Tflops peak
  - 77 Tflops max (preliminary run)
  - ---> 80 Tflops ?
Infiniband Architecture

- 24 nodes per shelf
- QNEM IB leaf switch

- 24 nodes per shelf
- QNEM IB leaf switch

- 24 nodes per shelf
- QNEM IB leaf switch

- 40 shelves in 10 racks
- 960 2-sockets nodes

- M9 648-port core switch
- 24 links

- M9 648-port core switch
- 12 links

- M2 leaf sw. 36 ports
- 21 links

- M2 leaf sw. 36 ports

- 5 Lustre OSS HA pairs

- 4 Lustre OSS HA pairs

- 11 infrastructure nodes

- 5 Lustre OSS HA pairs

- 11 infrastructure nodes

- 1 Lustre MDS

- 11 infrastructure nodes
First floor contains file system & infrastructure nodes

Second floor contains all compute racks + core networking switches

Racks aligned in a circle around a central hot core; outside ring is a cold aisle
Street view...
Satellite view...
cold air plenum (32 m²)

hot air core (25 m²)
View inside hot air core
Main specifications

- Rack capacity: 56
- Cooling capacity: ~1.5 MW
- Electrical capacity: 1.1 MW (1.6 MW)
- Blowing capacity: 132,500 CFM
- Maximum air velocity: 2.4 m/s
- Floor loading capacity: 940 lb/ft$^2$
Outline

- What is CLUMEQ?
- CLUMEQ's new HPC cluster: Colossus
- Lustre experience
Timeline

• New staff hired in Apr 2009 and July 2009
• Nov 12: System acceptance signed
• Dec 16: First “Beta users” on machine
• Nov - now: Learning, debugging, patching, tuning, helping users
Experience so far...

- Many technologies to get up to speed on installing, monitoring, debugging, tuning ...
  - Lustre
  - Infiniband
  - Grid Engine

- So far, pretty much everything is learned as we go
  - asking questions to SMEs
  - reading documentation, mailing list discussions
CLUMEQ Lustre deployment

- At acceptance:
  - ✓ CentOS 5.3 (2.6.18-128.2.1.el5)
  - ✓ OFED 1.4.1, Lustre 1.8.1.1

- Now
  - ✓ CentOS 5.3 (2.6.18-164.11.1.el5_lustre.1.8.2)
  - ✓ OFED 1.4.1(?) stock from RH
  - ✓ Lustre 1.8.2 + patch
Lustre fixes

• OSS crash during heartbeat/failover
  ✓ “CPU hog/soft lockups” bug (21612, 19557)
  ✓ Patched in 1.8.2

• 1.8.2 installed when GA
  ✓ Started to see high load on MDT, Lustre hanging on clients
  ✓ inode link count fix (22177)
  ✓ Installed patched version
Filesystem structure

- Lustre is the only FS on Colosse
  - ✓ /home for user accounts
  - ✓ /rap for group-shared space
  - ✓ /scratch for temporary data

- Lustre striping
  - ✓ /home and /rap use striping of 1 (typically small files)
  - ✓ /scratch uses striping of 72 (parallel IO performance)
OSS HA and RAID setup

- OSS HA pairs
  ✓ 4 OST per OSS
  ✓ Linux heartbeat used to signal failure
  ✓ Have 8 OST on OSS in failure mode

- RAID 6
  ✓ 10 1TB disk per OST
  ✓ Software raid (md)
HA and heartbeat

- Linux heartbeat/HA
  - ✔ Does not always work. Seen issues w.r.t. node not able to “kill” its neighbor (IPMI issues, investigating)
  - ✔ Currently working in manual HA mode

- Observing that it takes a long time for clients to use the new OSS taking over OSTs.
  - ✔ Not sure yet why. Needs investigating. todo++
Finding a failed disk

- “Interesting” experience
- Device name to physical location is not reliable
  ✓ need to double check with md* commands
- Then came “blinkenlights”
  ✓ http://wikis.sun.com/display/HPCSoftware/JBOD+Troubleshooting+Utilities
  ✓ REALLY useful!
Performance

- IOR used for parallel I/O measurements
  - IOR read performance = 33.6 GB/s
  - IOR write performance = 17.3 GB/s
  - all over IB

- Performance monitoring
  - Is there a BCP out there?
Conclusion

- Learned quite a lot in the past few months
- Lustre support team is key
- Looking to share experience