Dell XL Fall 2013 Meeting - GW Site Report

October 22, 2013

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

GW | Division of Information Technology

Columbian College of Arts & Sciences
THE GEORGE WASHINGTON UNIVERSITY
Office of Technology Services

Colonial One Background

- Colonial One is the new shared HPC cluster at GW
- Columbian College of Arts and Sciences lead the design and initial implementation
- Shared across colleges and departments, and supported by the Division of Information Technology
- “Pay to play” - groups who contribute resources have priority in the scheduling system
  - Not a condo model, resources are allocated exclusively through priority scheduling model

Colonial One - Current System

- Dell C8220 cluster, 96 nodes
  - 32x GPU nodes, each with dual NVIDIA K20 GPUs
  - 64x CPU nodes, each with dual 2.6GHz 8-core Intel Xeon CPUs, and 64/128/256GB of RAM
- Mellanox FDR Infiniband fabric
  - 6x 36-port switches as leaf currently (adding more)
  - 4x 36-port switches as the spine
- Two primary filesystems
  - 150 TB NFS fileserver (Dell NSS) for /home and /groups (MD3260)
  - 300 TB Lustre (Terascala) filesystem for high-speed scratch (MD3220 + 2x MD3260)

Colonial One - Hardware

- 4x PowerEdge r720 login nodes
  - each with a 20 Gbps connection to campus core
  - login / file transfer / compilation / …
- 2x PowerEdge r720 management servers, shared PowerVault MD3620i iSCSI array
  - running Debian / Xen 4.1
  - redundant Slurm controllers, license servers, network, cluster manager, web frontends for software …

Colonial One - Current System

Colonial One - Implementation

- What Went Well
  - A group that had never run a cluster was up and going in 2-weeks
  - Hardworking, accommodating Dell team (rack changes, IB airflow directions)
- Room for Improvement
  - Our facilities readiness (racks, cooling, cable management)
  - On-site BMC cable installation
  - IB/Ethernet cable lengths

Colonial One - Support Model

Bootstrapping a University HPC resource

New positions created to support this environment

- Senior HPC Systems Administrator - Tim Wickberg
- HPC Specialists - provide direct support for researchers, with domain-specific backgrounds
  - Physical Sciences - Glen MacLachlan (CCAS)
  - Genomics - (starting soon)

Colonial One - Software

- Bright Cluster Manager (6.0)
  - node images, system monitoring
- Slurm Resource Manager (2.6.3)
  - running independently of Bright, wanted latest features
  - tier-3 support contract through SchedM.D. in progress
- Globus Online (Globus Connect Multiuser endpoint)
  - easy + fast file transfer to XSEDE, other sites
  - encouraging users to run on their own systems, but limited adoption so far

CAAREN

- Capital Area Advanced Research and Education Network - CAAREN
- New high performance research network in the DC and Northern Virginia Area
- 100 Gbps backbone between Virginia and DC campuses, Internet2, and beyond
- High speed access to Colonial One, better network access for campus researchers

Four months on…

- We’re quite happy with the cluster
- Users love the stability and performance

“Colonial One is the easiest to use and most stable of all the GPU clusters I've used; while I'm wrestling trying to get things working on some other hellish machines, C1 is sitting there cranking out results. You and the rest of the administration team deserve a pat on the back!”

Four months on…

- Higher failure rate with nodes than expected
  - > 1 major repair / week, for 96 nodes currently
  - How does this compare to others?
  - Main culprits:
  - Power distribution board - TX8WP (~ 10, most from GPU trays)
  - Motherboards (~ 10, with another 5 suspect)
  - GPUs (~ 3)

Four months on…

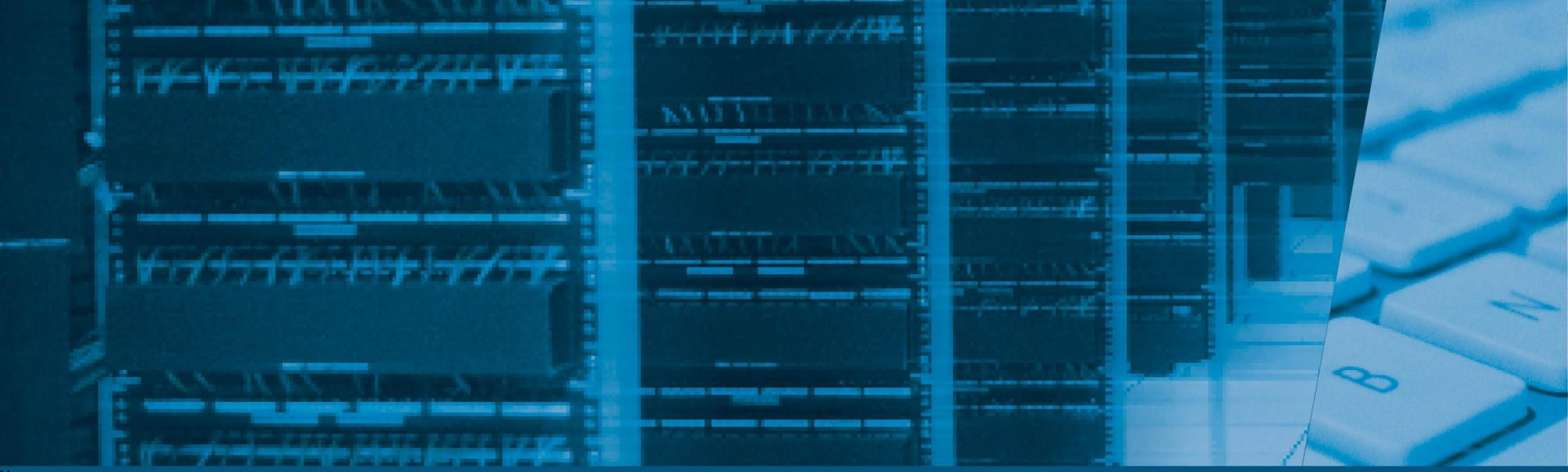
- Cooling is a challenge
- Get your datacenter folks thinking about this early on if you're going into an existing space
- 20 - 25 kW per rack
  - 4x denser than anything else in the room
  - 6" raised floor for cold air supply is insufficient
  - Literally sucking in air from the entire room
  - Cold air supply is 50 F
  - Chimneys don't work correctly at this density
- OpenGate Chimney Fans are helping considerably
- Long-term - looking at rear-door heat exchangers as an alternative strategy

Four months on…

- Odd issues with GPU cards
  - Card reset occasionally hangs, node has to be restarted to clear
  - ~0.5 - 1% of card resets…
  - but we currently reset after each job, and user jobs are O(3 hours)
    - so roughly one node hangs each day
  - Not sure how to get support involved on this - problem is intermittent, and usually a different node each time

Expansion plans

- Growing the system in the next 3 months, target is ~250 nodes
- Getting additional colleges and departments in brings new challenges
  - Scheduling model gets complicated - mix of long-running and HTC-style work is hard to address
  - Looking at enabling almost all Slurm priority / QOS features to mitigate



For More Information

Colonial One overview:

http://it.gwu.edu/colonialone-high-performance-computing

User documentation:

http://colonialone.gwu.edu

Email:

Tim Wickberg - wickberg@gwu.edu

Warren Santner - wsantner@gwu.edu