

CA T1&T2 Status Report, Q&A

Contributions & Inputs
from all CA operations personnel

ATLAS Sites Jamboree
March 2018

Landscape Evolving Rapidly (major changes)

- CA cloud: Canada + Australia
- Changes primarily driven by funding mechanisms and political aspects
- **Tier-1:** shifting resources to a new location (see Jamboree 01/17 and GDB 05/17 talks). Process well underway and started since early 2017
 - Currently running a distributed Tier-1 across 2 locations (cpu, disk, tape); primary services still hosted at TRIUMF location.
 - Full transition to be completed for the summer (cash flow delayed).
- **Tier-2:** consolidating and shifting resources to fewer number of sites
 - Canada: started with 5 Tier-2s (prior to 2013), then 4 sites; further consolidation onto 2 sites. Moving to a national team support model.
 - Australia: current funding round (Centre of Excellence) ending this year. Application submitted for 2020. In the meantime, resources are strained (both hardware & personnel).

Tier-2 Sites and Resources for 2018 and beyond

- Decommissioned CA-MCGILL-CLUMEQ-T2 (Jan 2018) and CA-SCINET-T2 (Apr 2018).
- Commissioned new CA-SFU-T2 (Jul 2017) and CA-WATERLOO-T2 (Apr 2018)
 - each site : ~2.3k cores, 4GB/core, Broadwell, E5-2683 v4 2.1 GHz (37 KHSPEC06); and 2.3 PB
- CA-VICTORIA-WESTGRID-T2 becomes opportunistic with virtualized batch system and cloud compute, \geq 1k cores (semi-elastic) and 0.5 PB.
- Australia: expanding, but post-2018 will probably see shrinkage/status quo due to limited funding.
 - extensive work happening now to move compute to local HPC farm and Openstack cloud. Storage is not an option at the moment (not fit for purpose and too expensive). University of Melbourne has a new project looking at computing needs for next 10 years, and have been asked to add Tier-2 requirements to it.

Containerization and Migration to CentOS7 / SL7

- TRIUMF Tier-1 new resources (@ new data centre) are containerized
 - Not using Kubernetes, instead Ansible and ad-hoc management are used
 - SL6 static docker containers (in-house built) on SL7 WNs for pilots to run
 - only for production because of network pipe limitation during the transition between the two sites, and no analysis queue defined yet.
- CA-SFU-T2 resources for ATLAS are using SLC6 Singularity containers
 - All nodes are stateless with CentOS7 and run SLC6 Singularity containers provided in CVMFS
- New CA-WATERLOO-T2 site will use Singularity containers as well once put into production later in the spring.
- Native SL7 WNs tested only at TRIUMF Tier-1 (@ old data centre)
 - Working fine and passing validation/HC tests, but never been put into production (continuous testing using 1 WN for a while, ~1+ year)

Unified Panda Queues

- All production Panda queues have been unified for 2 sites:
 - TRIUMF Tier-1 new resources (TRIUMF_DOCKER_UCORE),
 - Node assignment between score and mcore dynamically controlled by a Cron at batch system level
 - SFU Tier-2 (CA-SFU-T2_UCORE).
- No avenue for Analysis queues (why not since sites are resource providers ?)
- Setup is simple (must be ARC-CE)
 - AGIS: Clone existing mcore queue and append ucore to name (make it primary queue),
 - AGIS: Change Pilot Manager to aCT,
 - AGIS: Change catchall to unifiedPandaQueue,
 - AGIS: Select PFT, PFT_MCORE HC suites (and check that they run later in HC)
 - Alessandro: Clone the validations from existing primary queue,
 - Disable (set brokeroff/offline) and drain old production queues, stop pilots there.
- Nota Bene: aCT sends resource requirements in JDL
 - Jobs exceeding this may fail
 - Local scripts may have to provision in excess of these requirements.

PANDA Site Movers (with priority if more than one)

- Pretty heterogeneous configuration; mostly driven by sites design/preferences and evolution

CA Site	lcgcp mover	xrdcp mover	gfalcopy mover	rucio mover
TRIUMF (old resources)	log, read (0), write	read (1)		
TRIUMF (new resources)			log, write	read
Australia	log, read (0), write	read (1)		
IAAS (cloud)	log (1), read (1), write (1)		log (0), read (0), write (0)	
SCINET	log, read (1), write	read (0)		
SFU	log, read (1), write	read (0)		
Victoria	log, read (1), write	read (0)		

Tier-2 Networking, Canada (1)

- Networking done by site sys. admins in conjunction with campus IT and local ORANs plus CANARIE/AARNET as partners.
- Regular meetings and long-term planning discussions at national level (including Tier-1) coordinated by HEPNet Canada.
- Newer edge switch technologies at sites could allow for SDN but limited applicability at the moment as fairly static configurations and very little opportunistic resources in Canada. Have discussed QoS aspects due to eventual competing non-grid traffic.
- IPv6 been discussed and on roadmap but O(year) away - not a priority.
 - Have /26 or /27 IPv4 at most sites for grid computing

Tier-2 Networking, Canada (2)

- Sites are upgrading to 100G R&E but in stages (most at 20-40G now). (Victoria T2 was already at 100G for sometime)
 - LHCONE connections still at 10G (CANARIE backbone could be 100G already)
 - Occasional saturation but still rather rare
 - Likely can talk about 20-40G into LHCONE for T2 sites in next 1-2 years (important following sites consolidations)
 - Storage devices have not been the bottleneck usually (few x 10G to compute farm at least)
- Been running PerfSONAR at all sites for many years
 - Have Canada mesh including Compute Canada sites (non-T2) and grid sites (T2)
 - Used mainly for troubleshooting. No alerts or systematic monitoring yet
- Usage monitoring exists but not exposed publicly at most sites
- Some monitoring from CANARIE for LHCONE endpoints

Tier-2 Networking, Australia

- Currently 40Gb WAN connection for U. of Melbourne
 - Separate 2x10Gb for LHCONE
 - 100Gb from AU-US
- AARNET currently deploying SDN-capable AARNET-X network
 - Will require Unis to pay for it, and deploy at their discretion
- AARNET is fully IPv6 compliant, but nature of U. of Melbourne network is that currently we can't deploy IPv6 just to Research hosts - would require Enterprise network to also get it, and they are very hesitant
- Project underway to completely separate Research and Enterprise, which would allow deployment of IPv6 to Grid network
- No IPv4 exhaustion problems
 - 2 /24 networks, with opportunity for more

Tier-1 Networking (1)

- Robust and reliable networking infrastructure with multiple paths
 - We use Juniper 9214 chassis switch with multiple routing engines and multiple power supplies to provide reliability on the core level of Tier-1 data centre. The Tier-1's core switch is connected to TRIUMF's core switches (stack) via multiple 10G links to provide a solid connection between the TRIUMF site core switch and TRIUMF Tier-1 centre. Tier-1 access level switches have multiple LAGs to make full mesh connections with Tier-1 core.
- TRIUMF Tier-1 team manages the Tier-1 network and has full access to site-wide core infrastructure that is primarily managed by the IST team, and we are responsible for the ATLAS VRFs.
- Have been running perfSONAR instances for several years
 - Used for debugging, especially when involved in tickets related to transfer issues
 - We have an in-house Nagios plugin/alert to monitor packets losses between CERN and TRIUMF (LHCOPN), and could be extended to LHCONE and ORAN but too many links...
 - No perfSONAR instance yet at the new data centre location (soon)

Tier-1 Networking Q&A (2)

- Is your site peered with LHCONE? If not, why not?
 - Yes, we are peered with LHCONE.
- What is the overall network capacity (in Gbps) into your site available to ATLAS?
 - Our capacity is 10 Gbit/s on LHCOPN and 20 Gbit/s on LHCONE and 10 Gbit/s on ORAN dedicated to ATLAS.
- What is the overall % utilization (during peak periods)?
 - LHCOPN utilization often reaches 9 Gbit/s and LHCONE/ORAN can go up to 19 Gbit/s.
- IPv6 transitioned yet, or plans?
 - Yes, we do support IPV6. Core level is configured with IPV6 on LHCOPN and LHCONE. We have storage test bed which is configured with IPV6.
- Do you have a monitor URL (Ganglia, Grafana) that you can share?
 - http://gridinfo.triumf.ca/aggr_m2/
 - <http://gridinfo.triumf.ca/network/>
- Any experience or plans for SDN?
 - Almost all of our switches are supporting SDN, but no concrete plans yet for a test bed.

Tier-2 Storage Experience

- Been relying heavily on shared DDN systems (9500, 9900, 10k) for both dCache and StoRM/GPFS
 - Little data loss due to underlying corruption - tend to be glitches in software or configurations
 - Some extended periods of data unavailability when recovering
- Sites based now mostly on DELL DSS7700 with self-management contracts
 - chose dual sled version (2x45 drives) to mitigate potential data loss (and added performance)
 - Need to build better monitoring/alerting for underlying hardware failures and repair mechanisms
- Some experience being developed with CEPH and S3 storage at sites but not for grid yet, other than University of Victoria and IAAS-CLOUD
- Some pressure to move off independent silos of storage (dCache) and move to large shared parallel filesystems (Lustre, GPFS) with StoRM or something similar

Tier-1 Disk Storage Experience

- Reliable and robust disk storage infrastructure and operations + dCache
 - Integrated management including site power, cooling, monitoring and remote access
 - Any new deployment goes through full functionality and stress testing
 - RAID6 (sw or hw) with hot spare drive, redundant access to storage unit from multiple servers if possible (for non direct-attached systems)
 - SNMP traps for hardware alerts with grouping and filtering
 - Mixed storage vendors and products: drive failure rate varies; some had almost all disk drives replaced in their life span, while some rarely had drives replaced.
 - Dedicated dCache billing/accounting server with pg cstore-fdw extension provides several years of historical data for mining.
 - Ansible playbooks for storage management and configurations
 - Nagios alerts + well defined procedures and documentation for 24x7 operations
- No clustered filesystem used
- Currently providing 7.8 PB disk capacity for ATLAS, 9 PB this summer (full refresh)

Tier-1 Tape Storage Experience

- Robust tape operations for several years
 - Reliable Hardware
 - The non-proprietary tape software developed at TRIUMF , reliable operation
 - Minimize tape mounts & maximize reads per mount, less pressure on library
 - Open format(mt-st,scsi2logic, tar), SCSI for data, tcp for control
 - No extra server/disk buffer for tape operation, it runs on HSM pool nodes
 - ENDIT HSM plugin interface minimize HSM interface load
- Recent operations
 - Migrated 2.5PB data from LTO-5 to remote site tape instance, which using LTO-7 media
 - No file lost, no media issues encountered
- Tape experience feedback to ADC (**repeated for several years...**)
 - Bulk staging is always preferred, 5k-30k is good
 - Large files are welcomed by tape, >1G is good
 - URL for staging requests list once unleashed ?
- Providing 18 PB capacity for ATLAS, will increase it to 30 PB this summer

Concerns & Issues

- **ADC is pretty aware of ~all of these but listed here for completeness...**
- CA operations team are pretty active in noticing issues and promptly reporting them back to ADC (a historically good working relationship)
- Overall utilization of pledged T1 & T2 resources (i.e. idle cycles)
- Some jobs forking out too many processes (MadEvent)
- MemoryMonitor process from some releases used too much CPU
- Overlay jobs overloading Frontier/ DB server (fixed ? , is direct access can be mitigated in sw ?)
- Tape file sizes and bulk recalls
- AGIS: we can be unaware if someone else makes changes for the cloud. Can we have a mechanism where change logs are automatically sent to cloud support ? (or leave all AGIS changes to CA unless force majeure event ?)