



AGLT2 Site Report

AGLT2 is a Co-Production of University of Michigan
and Michigan State University








Bob Ball/University of Michigan

Shawn McKee, Chip Brock, Philippe
Laurens, Mike Nila

OSG AHM March, 2017 / San Diego



Outline

-  Site Summary
-  2016 Procurement Details
 -  2017 Retirements
-  HTCondor-CE Status
-  Site Mover Controls and Pilots
-  Networking
-  IPv6 and SL7

Site Summary

- ❏ The ATLAS Great Lake Tier-2 (AGLT2) is a distributed LHC Tier-2 for ATLAS spanning between UM/Ann Arbor and MSU/East Lansing. Roughly 50% of storage and compute at each site
 - ❏ 9408 logical cores
 - ❏ MCORE slots 950 (dynamic) + 10 (static)
 - ❏ 720 Tier-3 job slots usable by Tier-2
 - ❏ Average 10.21 HS06/slot
 - ❏ 6.85 Petabytes of storage
 - ❏ Total of 96.1 kHS06
 - ❏ Most Tier-2 services virtualized in VMware
- ❏ 2x40 Gb inter-site connectivity, UM has 100G to WAN, MSU has 10G to WAN, lots of 10Gb internal ports and 16 x 40Gb ports and a few 100Gb
- ❏ High capacity storage systems have at least 2 x 10Gb bonded links
 - ❏ Newest have 2 x 25Gb, or even 2 x 50Gb
- ❏ 40Gb link between Tier-2 and Tier-3 physical locations

Procurement Details

- ☐ All funds for 2016 and the Columbia CA purchases have been expended
 - ☐ Of order \$100 will be returned to Columbia
 - ☐ Paperwork is on track
 - ☐ Three N2048 switches yet to be brought online (March projected)
 - ☐ All other CPU and storage is online and operational
 - ☐ CA purchase summary
 - ☐ CPU: 7841 HS06 (720 logical cores)
 - ☐ 4 x Dell N2048 Gb switches
 - ☐ Dell S4048-ON 10Gb switch

- ☐ Compute and storage capacity increase December 2015 to January 2017

	2015 Actual	2016 Pledge	2016 Actual	2017 Pledge	2018 Pledge
CPU (HS06)	68506	24000	96080	57500	59289
Storage (TB)	3712	3000	6850	4242	4651

Actual as of year end except for Jan 31, 2017

2017 Retirement Details

- ☐ Potential storage retirements in 2017
 - ☐ Up to 1150TB of older storage in 60 shelves of 13 servers
 - ☐ Compression to half a rack from ~4 racks if replaced by more MD3460 with currently procured disks
- ☐ Potential CPU retirements in 2017
 - ☐ Site total 7992 HS06
 - ☐ 32U at MSU
 - ☐ 30U at UM
 - ☐ 824 Logical Cores
- ☐ Total Storage following retirements: 5.7PB (2017 pledge is 4.242PB)
- ☐ Total CPU following retirements: 88.1 kHS06 (2017 pledge is 57.5 kHS06)
- ☐ Total Logical Cores following retirements: 8576
- ☐ Future FY2017 purchase schedule not yet determined
- ☐ LOCALGROUPDISK
 - ☐ capacity 445TB
 - ☐ Usage 346TB

2016 Performance

- Overall site availability and reliability metric was 96%-100% all year
- Stage-in rate limitations likely affected cpu/wall time ratio throughout the year
 - Much improved for 2017
- CPU and Wall time below are in hours
 - Single Core includes LargeMEMory jobs

Task	Num Jobs	Cpu-time	Wall-time	Cpu/wall
Analysis	1435679	1644606	2345423	0.702
All Jobs	6945552	38446849	7812634	4.921
Score Prod	4488418	12163929	1433707	0.848
Mcore Prod	1021455	24638314	4033504	6.108

HTCondor-CE Status

- Running at AGLT2 since Fall of 2015
 - GRAM is running, but will soon be totally disabled
- Just upgraded to OSG 3.3.21 on all gatekeepers, with HTCondor 8.4.11
 - WN are still on 8.4.6
- Only skeleton of [Resource Entry CHANGEME] section in 30-gip.ini is ready
 - Will configure and test this OSG Collector reporting over the next few weeks
- Question: will BNL provide a “template” JobRouter conf file for US Sites?**

SiteMover Controls and Pilots

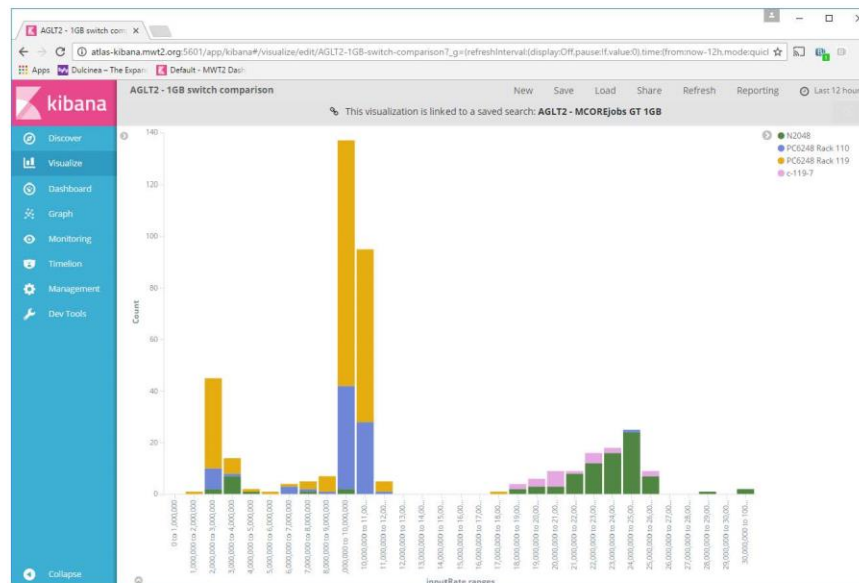
- ☐ New pilots have been running at AGLT2 for a few weeks without incident
- ☐ Using MWT2 LSM as our site mover
 - ☐ Some site-specific mods in place; reporting, locale, ...
 - ☐ Logging to syslog and UC Kibana

Networking

- ❏ No external network changes planned
- ❏ Internal network is evolving
 - ❏ Using 100Gb Mellanox SN2700 as local collector switch
 - ❏ 2x40Gb uplink to Juniper EX8200 thence to WAN
- ❏ Replacing Dell PC6248 with N2048 on public NICs for increased useful bandwidth
- ❏ Local 10Gb collector Dell S4048-ON
 - ❏ 48 10Gb and 6 40Gb ports

Networking

- Plot shows UC Kibana-collected stage-in rates
- Clear peak identified as PC6248
- Replacement N2048 has $> 3x$ better sustained rate

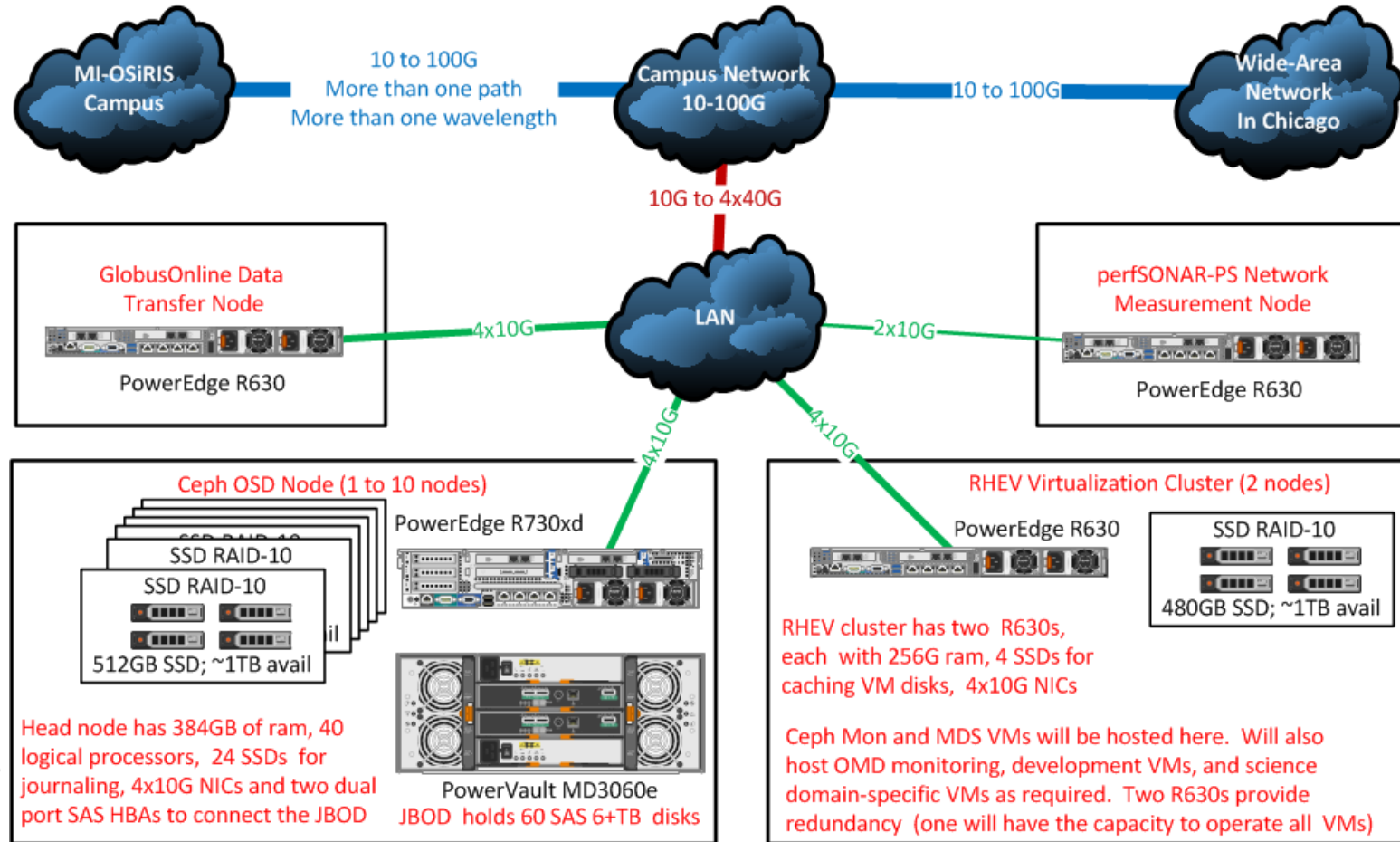


IPv6 and SL7

- Both UM and MSU sites now routing IPv6
- perfSONAR instances all now dual stacked
- Storage will be next in this regard
- No plans for SL7
 - Some servers transitioned
 - Waiting for directions on SL7 xrootd redirector
- Singularity now installed for OSG experimentation
 - Experience likely useful for SL7 transition when we are ready for that

Software-Defined Storage Research

MI-OSiRIS Data Infrastructure Building Block



See NSF announcement

http://www.nsf.gov/awardsearch/showAward?AWD_ID=1541335

NSF proposal
OSiRIS **funded**
involving UM,
MSU, WSU (and
our Tier-2)

Exploring
Ceph+SDN for
future software-
defined storage

Goal is
centralized
storage that
supports in place
access from CPUs
across multiple
institutions

Summary

- ❏ No big issues holding us back
- ❏ All funds spent out and closeout paperwork on track
- ❏ WAN operating at 2x40Gb with transparent 10Gb fallback
- ❏ IPv6 namespace in place
- ❏ No wide rollout of SL7 any time soon
 - ❏ Singularity on our cluster now that could help