

CHPC WLCG Tier2 Facility and SA Grid Operations

Sean Murray

ALICE

CHPC

CSIR

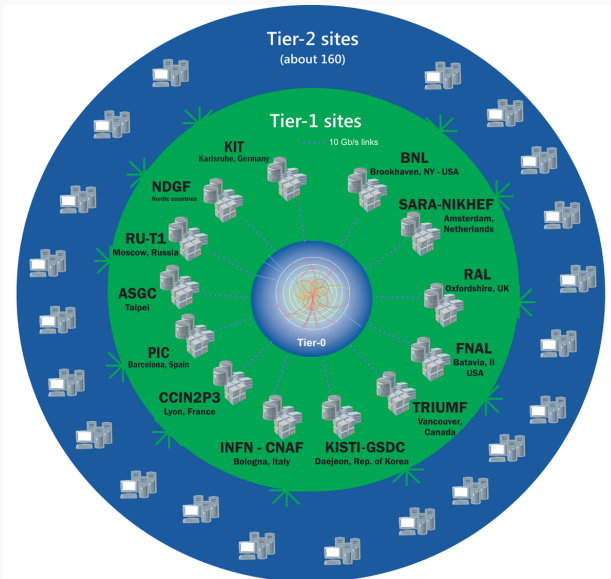
October 6, 2016

WLCG

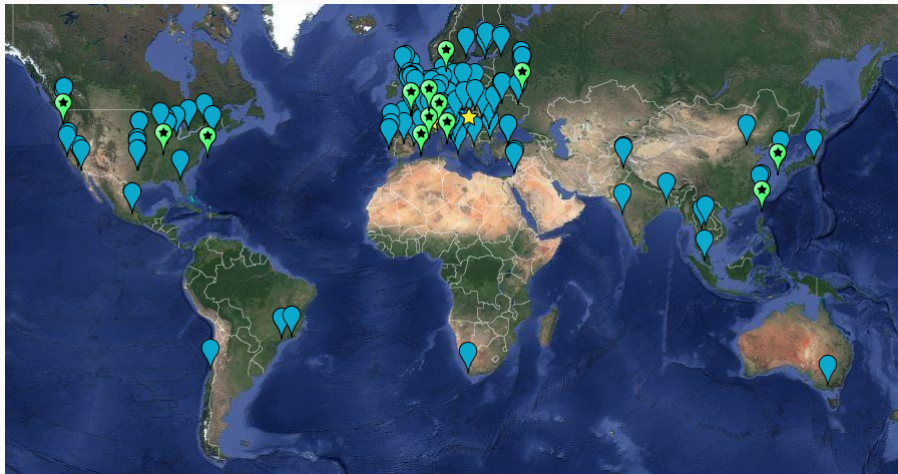
SAGrid, user analysis

Backup Slides

What is the WLCG



WLCG Map of Sites



WLCG MOU Signed



28 April 2015

Commitments

According to Tender :

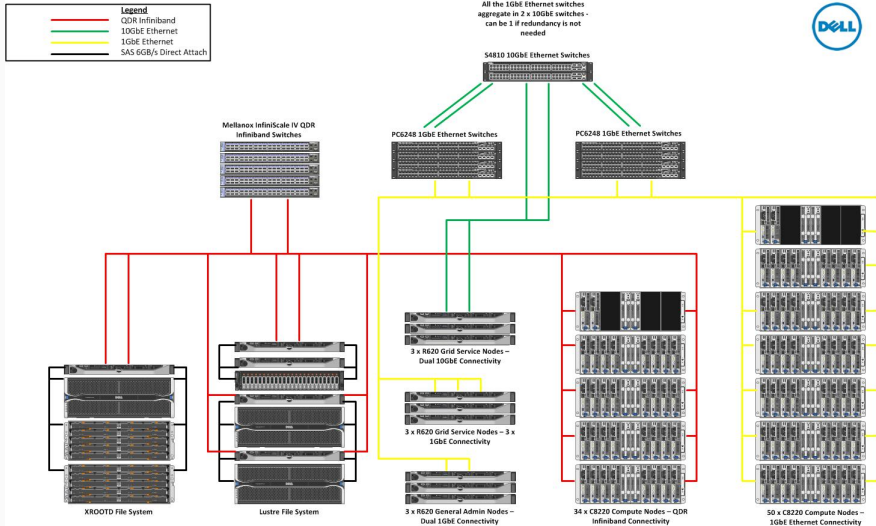
- ALICE 600 cores
- ATLAS 600 cores
- ALICE 400TB
- ATLAS 400TB

According to :

<https://wlcg-rebus.cern.ch/apps/pledges/resources/>

- 6000 HEPSPEC06 cores (c. 560 of our cores)
- 100TB storage
- All ALICE.
- now equal ALICE and ATLAS (?)

Computing Infrastructure



Current hardware

- 50 nodes of 48 cores 192GB RAM and 2x800TB of SSD, 1G ethernet
- 34 nodes of 48 cores 96GB RAM and 1TB SATA, FDR infiniband, 6 “stolen”
- c.100TB of Lustre on the 34 nodes with FDR infiniband.
- 9 management servers, lower spec
 - compute element (head node,ce),
 - storage element 2 redirectors, 2 storage nodes with direct attached multipath storage
 - authentication, user interface (gone), monitoring, provisioning.

Current Storage

- 383TB EOS for ALICE, down from 440TB
- 252 TB EOS for ATLAS, down from 400TB
- 107 TB lustre for 34 nodes.
- 104 TB EMC for ATLAS, not going to be used.
- 400TB ALICE
- 400TB ATLAS
- 400TB General EOS

Reduction in data sizes is due to reorganisation for reliability.

Current Performance

- 465k ALICE jobs in last year
- Avg concurrent jobs 1600, limited due to network..
- Data traffic consumed 70TB in and 60TB out in the last 3 months

Availability / Reliability

So who monitors us :

- WLCG
- EGI
- ALICE
- ATLAS
- me via zabbix/grafana and racktables

There are a lot of eyes, ignoring the ones in this room.

Function	July	August	September
Availability	100	100	100
Reliability	100	100	100

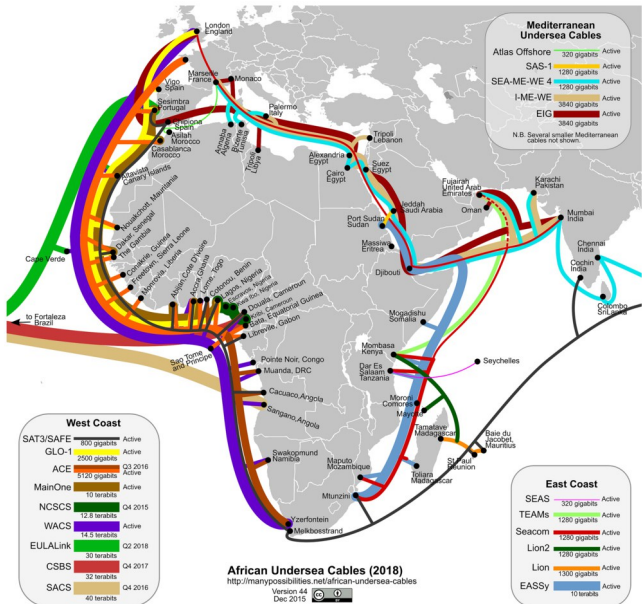
I am soooo going to regret this slide !

Grafana ALICE



This is currently being expanded to pull in more from MonaLisa, and more experiment and storage specific metrics to help to trivially diagnose and forewarn problems.

Our Data's Scenic Tour



Tier2 Processing

- ATLAS processing direct std batch jobs.
- ALICE processing, job agents and leave it to central services, vobox.
- ops jobs fail due to starvation from alice. currently a dedicated machine
- all software is distributed via cvmfs, both experiments.
- ALICE has something similar to code-rade, an entire ecosystem of modules, versions, builds, etc.
- still no official validation on CC7.

Thankfully both experiments use xrootd.

- ALICE has its own authentication system, controlled centrally from CERN.
- ALICE has one big block of storage.
- ATLAS uses grid pool accounts, more standard grid.
- ATLAS has various storage's on site, cut up locally.
- ATLAS on puppet, ALICE will go on puppet when upgraded.

- Keep It Simple Simpleton.
- Only use liberally licensed software.
- Stand on whom evers shoulders I can.
- I dont want to do sys admin, I dont enjoy it.
- Abstract sys admin to a Software Engineering problem.
- Let me get back to the things that I enjoy ALICE O2 and HLT which I enjoy, computational physics..

Simple Solution

Switch everything off for 4 or 5 months while i redo **everything**

- Keep site up and processing maintaining our Tier 2 status.
- incrementally replace systems.
- the return of 28 grid nodes, on infiniband, a place to test configurations when batch is “free”.

Improvements

- Fixed ALICE error rates, mostly.
- ALICE concurrent jobs from 700 to 1800.
- max out bandwidth now regularly.
- ATLAS running pilot jobs.
- Storage cleaned up and reduced.
- New storage quoting
- Plan in place to overhaul, and being tested.
- Work defined to Redo entire internal network.

upgrades

- attempts to delay till CC7 validation have failed.
- Puppify, to auto site deployment, r10k an issue.
- Transition to foreman from xcat.
- upgrade monitoring server to zabbix3, via docker this time.
- Rewire whole network, re-power to monitored pdu, and monitor all.
- Add inherent redundancy into 10G interfaces on ce,se,se2.
- Reinstall while TRYING to keep A/R. problems are vbox and ce.
- Storage, we need an additional 750TB for ALICE and ATLAS just to maintain Tier2.

Historically this has not been great, so . . .

- Federated logins to zabbix and grafana, i.e. You
- All code on github in line with AAROC
- All issues PUBLIC on github, this has been a serious uhming and ahring session.
- Still have GGUS for normal tickets.
- Some training on the user analysis facility (hopefully online)
- A couple of things remain private like network diagrams, obviously, and passwords. github encrypted repos (todo)
- AAROC slac channels.



WLCG

SAGrid, user analysis

Backup Slides

July 28 of the 34 nodes came back for SAGrid and HEP user analysis.

- Go back to SAGrid to support anybody on SAGrid VO.
- HEP user analysis, based on federated identities, no user account admin.
- either HEP user supplies their own eco system with job or build their custom ecosystem via code-rade.
- code based on CODE-RADE, or LHC experiments.
- HLT replica for me and others. (no gpu)
- O2 dev for O2 development with dds/mesos/pbs. (no gpu/fpga)
- Local Storage for users, eos and lustre.
- 24/4 sl6/centos7 in prep for container migration.

Foreman ...

1. deployment
2. configuration, puppet, chef, salt, ansible.
3. monitoring.

Not using the fully integrated puppet/foreman as i prefer to be tool agnostic, or at least strive for it.

A method/system to maintain a system in a known state.

hieradata, a key value store used by puppet.

We store all site specific information in hieradata, tightly integrated into puppet..

- r10k branches are mapped to environments (dev, testing, production).
- Cant be under AAROC, so chpctier2, production branch will be synced to aaroc/devops/chpc.
- 3 branches, dev, testing and production, currently only use production.

Devops requires some form of automated testing. Unit testing via rspec. Function testing. Fact testing. System testing.

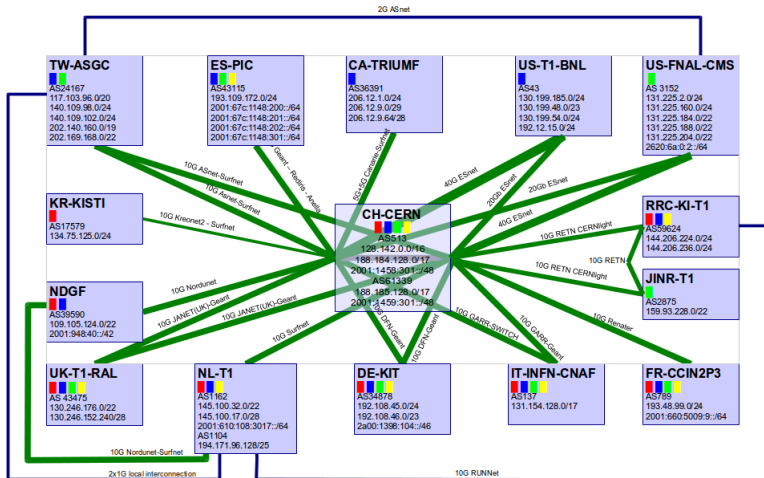
put in pics from other doc of old and new. ipmi interfaces in, monitor all including pdu's direct dual 10G links to interwebs.

explain test and results.

WLCG

SAGrid, user analysis

Backup Slides



three or more other institutes providing amongst them 52-week coverage).

The following parameters define the minimum levels of service. They will be reviewed by the operational boards of the WLCG Collaboration.

Service	Maximum delay in responding to operational problems			Average availability ⁶ measured on an annual basis	
	Service interruption	Degradation of the capacity of the service by more than 50%	Degradation of the capacity of the service by more than 20%	During accelerator operation	At all other times
Raw data recording	4 hours	6 hours	6 hours	99%	n/a
Event reconstruction or distribution of data to Tier-1 Centres during	6 hours	6 hours	12 hours	99%	n/a

accelerator operation					
Networking service to Tier-1 Centres during accelerator operation	6 hours	6 hours	12 hours	99%	n/a
All other Tier-0 services	12 hours	24 hours	48 hours	98%	98%
All other services ⁷ – prime service hours ⁸	1 hour	1 hour	4 hours	98%	98%
All other services ⁷ – outwith prime service hours ⁸	12 hours	24 hours	48 hours	97%	97%

Annex 3.2. Tier-1 Services

Each Tier1 Centre⁹ forms an integral part of the central data handling service of the LHC Experiments. It is thus essential that each such centre undertakes to provide its services on a long-term basis (initially at least 5 years) and to make its best efforts to upgrade its installations steadily in order to keep pace with the expected growth of

reviewed by the operational boards of the WLCG Collaboration.

Service	Maximum delay in responding to operational problems			Average availability ⁶ measured on an annual basis	
	Service interruption	Degradation of the capacity of the service by more than 50%	Degradation of the capacity of the service by more than 20%	During accelerator operation	At all other times
Acceptance of data from the Tier-0 Centre during accelerator operation	12 hours	12 hours	24 hours	99%	n/a
Networking service to the Tier-0 Centre during accelerator operation	12 hours	24 hours	48 hours	98%	n/a
Data-intensive analysis services, including networking to Tier-0, Tier-1 Centres outwith accelerator operation	24 hours	48 hours	48 hours	n/a	98%
All other services ⁷ – prime service hours ¹⁰	2 hour	2 hour	4 hours	98%	98%
All other services ⁷ – outwith prime service hours ¹⁰	24 hours	48 hours	48 hours	97%	97%

The response times in the above table refer only to the maximum delay before action