# ZA-CHPC ALICE Computing 2019 and beyond

Sean Murray
ALICE
CHPC,CSIR
University of Cape Town

May 14, 2019
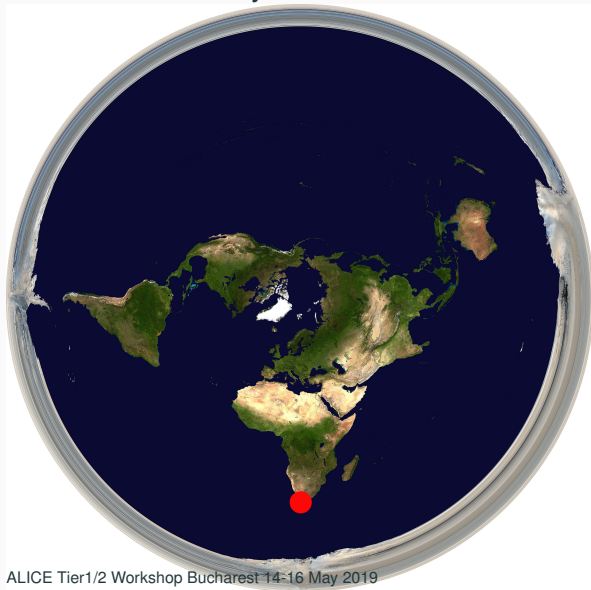
# Location
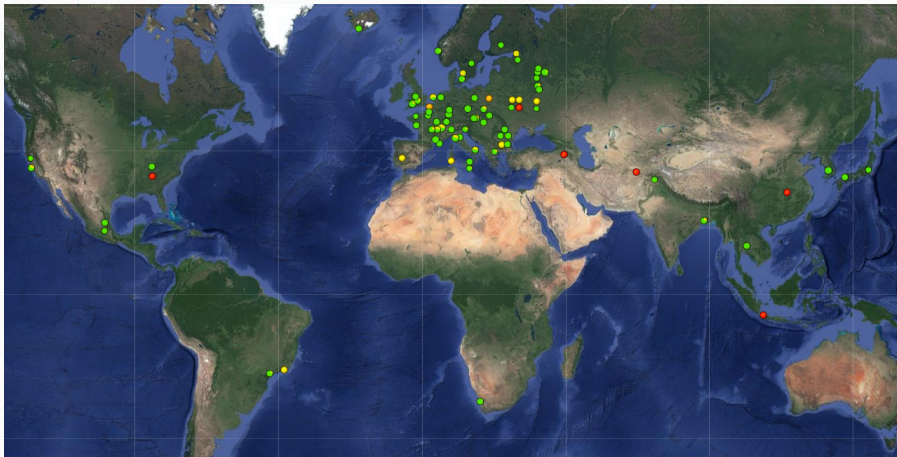
# All opinions matter these days, so to be all inclusive ...

So to not offend any flatearthers ....

## Commitments

Original hardware :

- ALICE 600 cores
- ATLAS 600 cores
- ALICE 400TB [383TB]
- ATLAS 400TB [248TB]

According to :
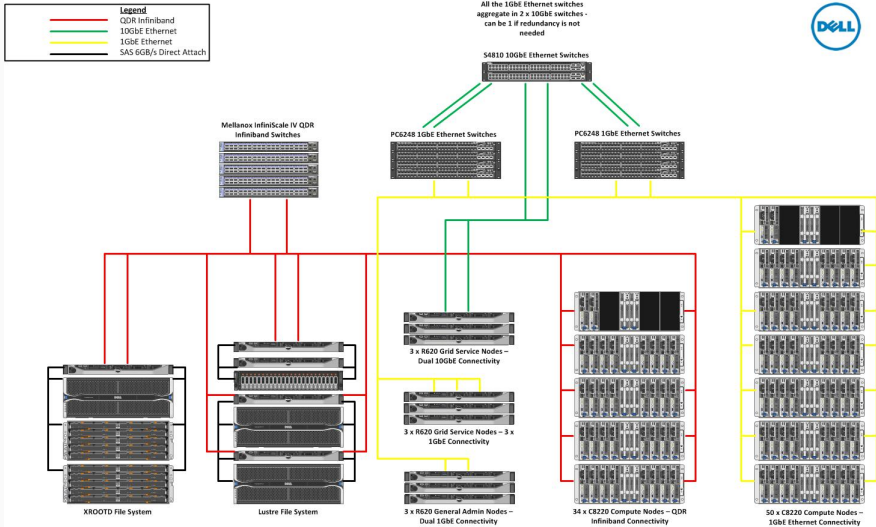https://wlcg-rebus.cern.ch/apps/pledges/resources/

- 12000 HEPSPEC06 cores
- 1.5 PB storage
- Some stuff for another experiment.

## Current hardware

- 50 nodes of 48 cores (10.6) 192GB RAM and 1.6TB of SSD, 2x bonded 1G ethernet
- 28 nodes of 48 cores (10.6) 96GB RAM and 1TB, QDR infiniband
- 9 management servers, lower spec
- A rack of stampede from TACC (part of for now)
  - compute element (head node,ce),
  - storage element 2 redirectors, 2 storage nodes with direct attached multipath storage
  - spare, monitoring, provisioning.

- 383TB EOS for ALICE
- 202 TB EOS for ATLAS, down from 252TB

**Legend**
- QDR Infiniband
- 10GbE Ethernet
- 1GbE Ethernet
- SAS 6GB/s Direct Attach

All the 1GbE Ethernet switches aggregate in 2 x 10GbE switches - can be 1 if redundancy is not needed

S4810 10GbE Ethernet Switches

Mellanox InfiniScale IV QDR Infiniband Switches

PC6248 1GbE Ethernet Switches

PC6248 1GbE Ethernet Switches

DELL

3 x R620 Grid Service Nodes – Dual 10GbE Connectivity

3 x R620 Grid Service Nodes – 3 x 1GbE Connectivity

3 x R620 General Admin Nodes – Dual 1GbE Connectivity

XROOTD File System

Lustre File System

34 x C8220 Compute Nodes – QDR Infiniband Connectivity

50 x C8220 Compute Nodes – 1GbE Ethernet Connectivity

# Current Performance last year



- stable after going onto our 10G connection.
- Was limited to 1G gateway on CE, now 10.

## Current Performance last year

- 1.7Million, last year 890k ALICE jobs in previous year, 465 the previous year.
- 970 average concurrent jobs down from 980 due to sharing with ATLAS.
- Storage became full, now at 184TB, last year 91TB of 383TB
- 1.6PB(0.235) in 54TB(97) out. (spent a long time being full)
- Average 71(8.9)MB/s in and 3(14.3)MB/s

## Availability / Reliability TODO

| Function | Last 365 days |
|----------|--------------:|
| Availability | 98 |
| Reliability | 99 |
| Storage | 94.5 |

Storage down time is primarily due to storage test failures due to mostly io bottleneck on SAS.

## Downtimes

The big ones are, ignoring network starvation :

25 July   cream-ce disaster.

30 April  Change to network ipv4/6 and ended up EOS upgrade as
          well.

## Network maximisation

We now have our own network connection (10G) to our NREN backbone.

- Down time to migrate to new network blocks ipv4 and ipv6. No ipv6 onsite and poor institute network.
- Connect directly to NREN backbone, naked. NREN is doing our routing.
- Initial plan was to upgrade EOS after, and keep ALICE storage on ipv4.
- alias local network. Idea was seperate 10G connection for local and external (fix pending).
- struggled with policy based routing.
- perfsonar node to come, a WLCG requirement.

Network was bad and then since March the storage has made things worse.
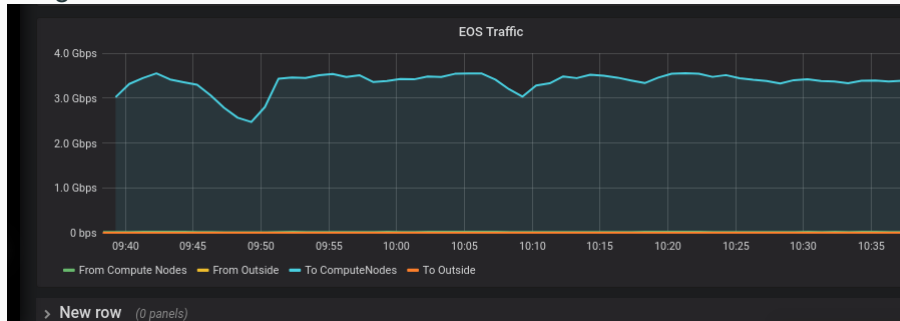
# Processing last year
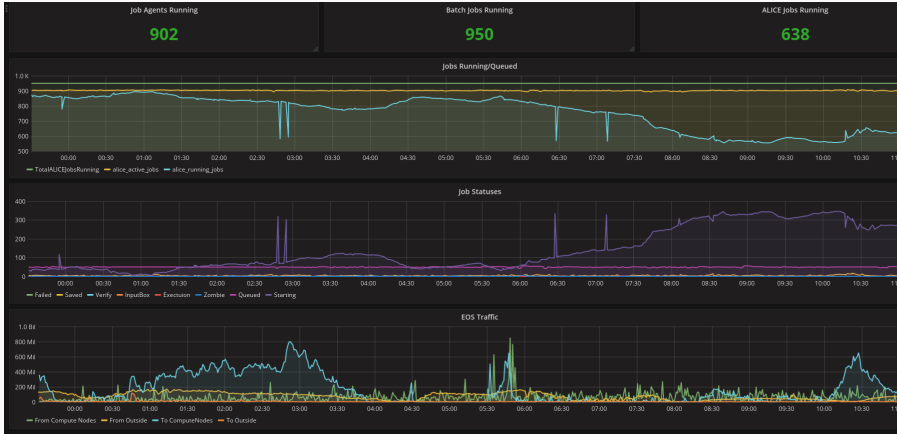
# Processing this year

**provided and (required)**

| Resource | 2017 | 2018 | 2019 | 2020 |
|----------|------|------|------|------|
| CPU kHS06 | 7.5 (7.5) | 7.5(8.09) | 12(12.35 | 22(14)) |
| Storage TB | 384 (682) | 384(990) | 384(1100) | 1000 (1300) |

- Our HS06 value is 10.6.
- Pledges are wrong, but still under.
- I am keen to stretch the network as is our NREN.

## Outline

**cpu growth**

Migrate computing into the lower spec 28 nodes(another experiment beginning with an "A")
Opennebula on Stampede rack. Shared for 3 use cases.

**Storage quotes and money.**

- Purchase pending for 5 nodes of 24 disks, 2 head nodes, 1 mellanox 40G switch.
- R2M to spend on new storage (60 disk arrays quoting), 1 arista 40G .

- 1 May (during network disaster) a ggus ticket from ATLAS requesting finalisation of Centos7 upgrade, with singularity.
- A bit of a shock, but mid June set as upgrade.
- We had intended to upgrade by end of year **including** CE replacement.
- some things not available in ansible, so work required.

Africa Undersea Cables (2018)
http://manypossibilities.net/african-undersea-cables
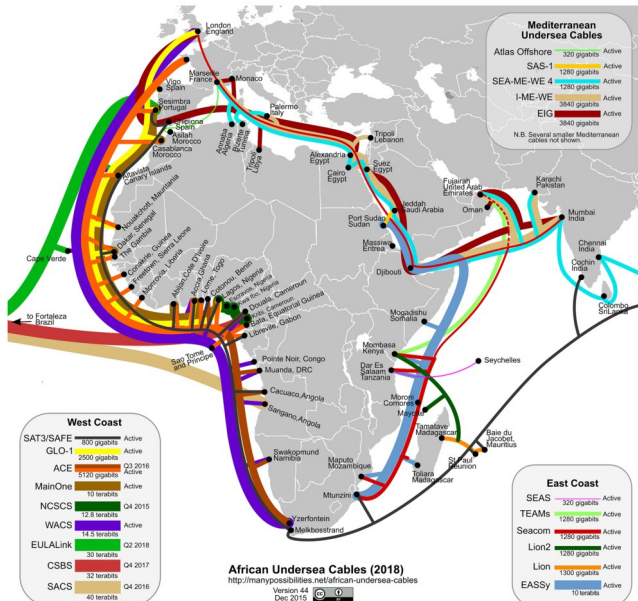
**devops (stampede)**

- migrating to complete ansible site with testing via stampede nodes.
- stampede nodes for build farm,
- devops testing (no trust after july 25th 2018)
- VMs for batch farm.
- possibility of a grid site with ignored A/R ?

## HPC Ecosystems (2018-08)

HPC Ranger Sites
★ All items

Stampede Sites
• All items

HPC Cambridge Sites
★ All items

HPC C6100 sites
★ All items

Planned sites
📍 All items

Interested sites
ⓘ All items

Update of HPC Ecosystems
sites as of 2018-08

## WCLG across Africa

A toy project for now.

- SKA project to upskill.
- Initially tried for Mauritius, for obvious reasons, then Mozambique.
- Now trying WITS (a South African university currently doing ATIAS).
- Donated hardware that is sitting idle for a large part.

## Outline

## Summary

Things are improving alot.

- Storage is shortly to increase greatly and be accessible thanks to network
- all cpu's will run thanks to network.
- Network upgraded.
- fully testable ansible deployable site (avoid july 2018 issue)
- By next meeting planning on Tier1 finalised.