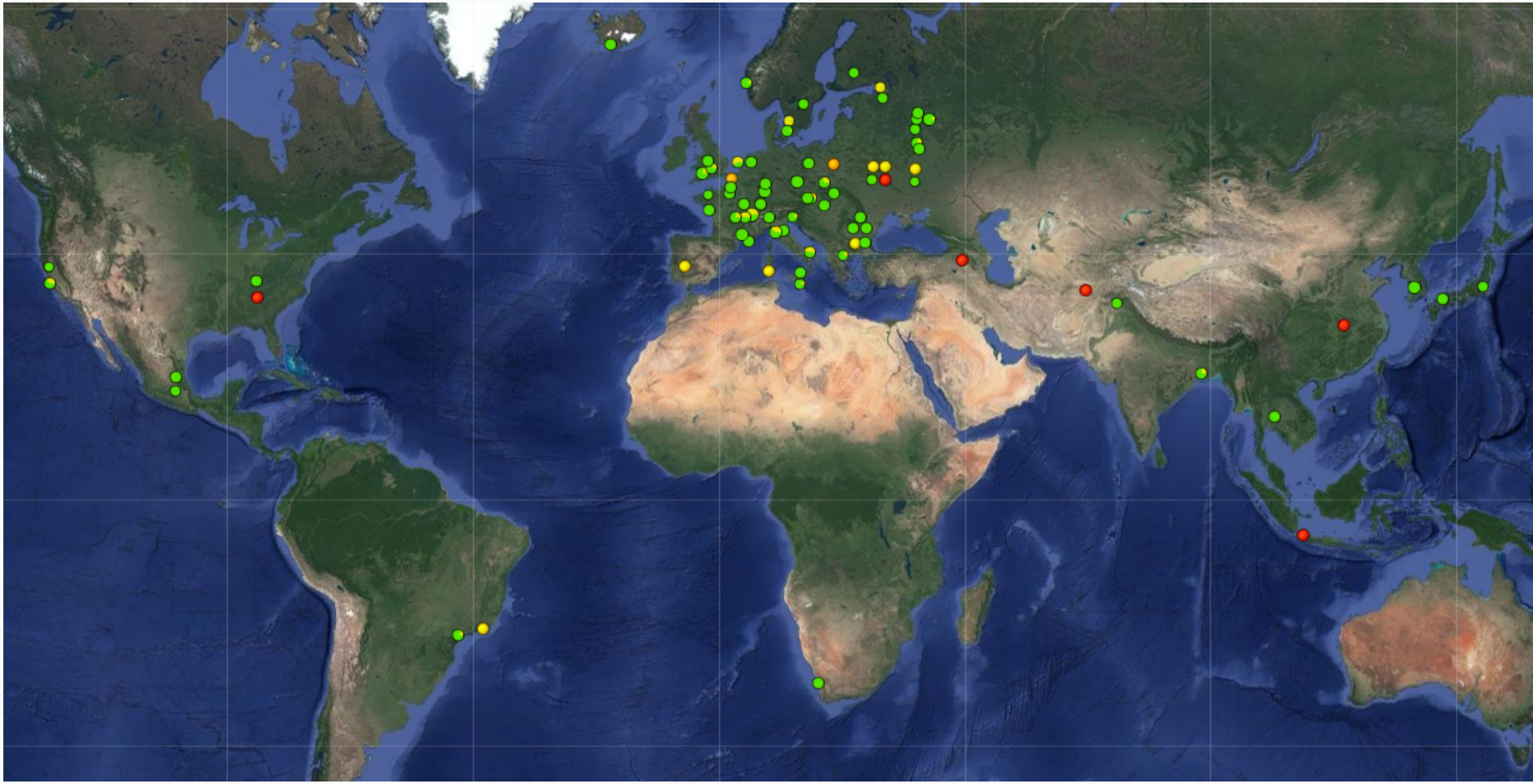
South Africa ALICE Computing

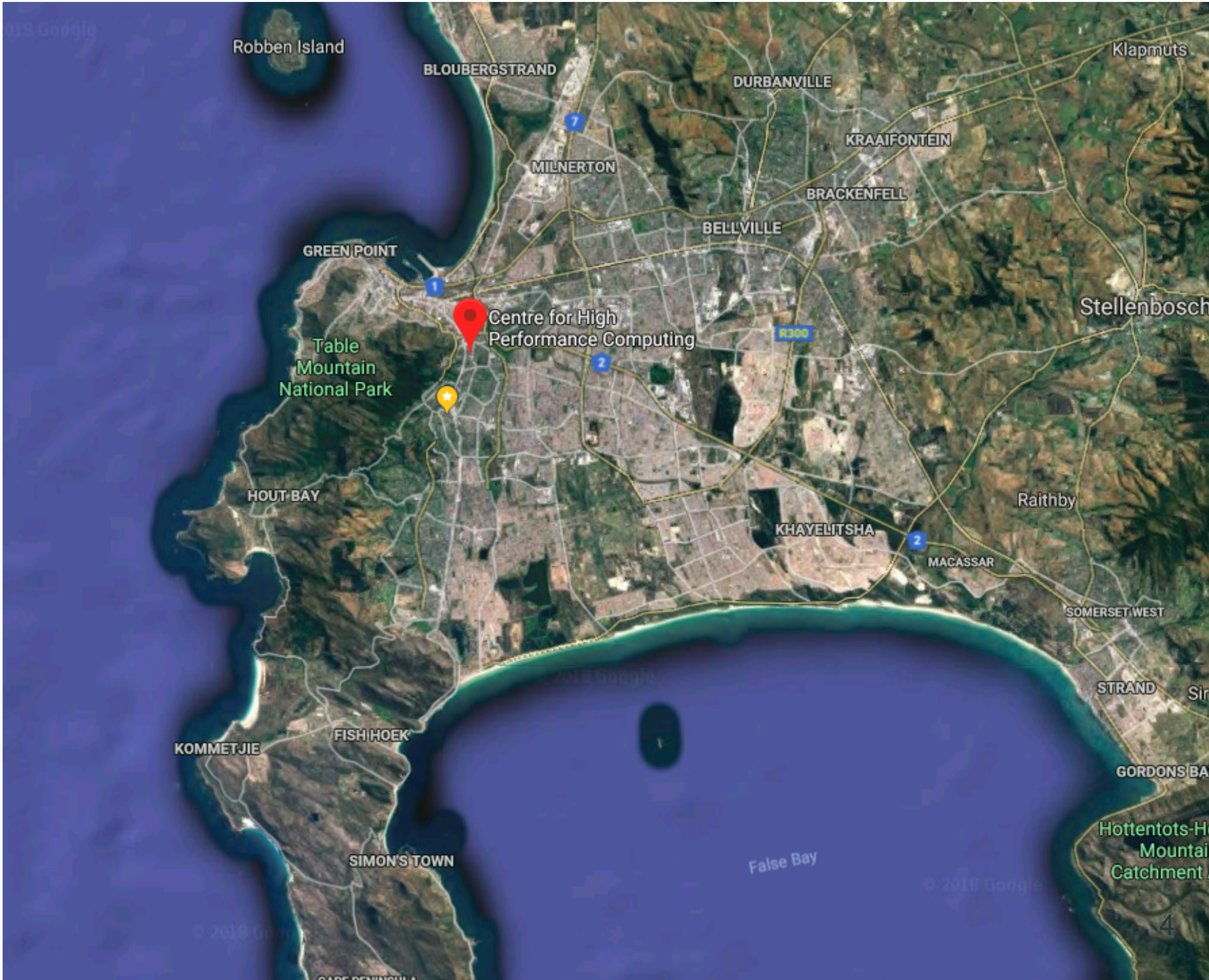
Sean Murray & William Phukungoane
CHPC,CSIR
University of Cape Town

April 17 2017

Outline

Location



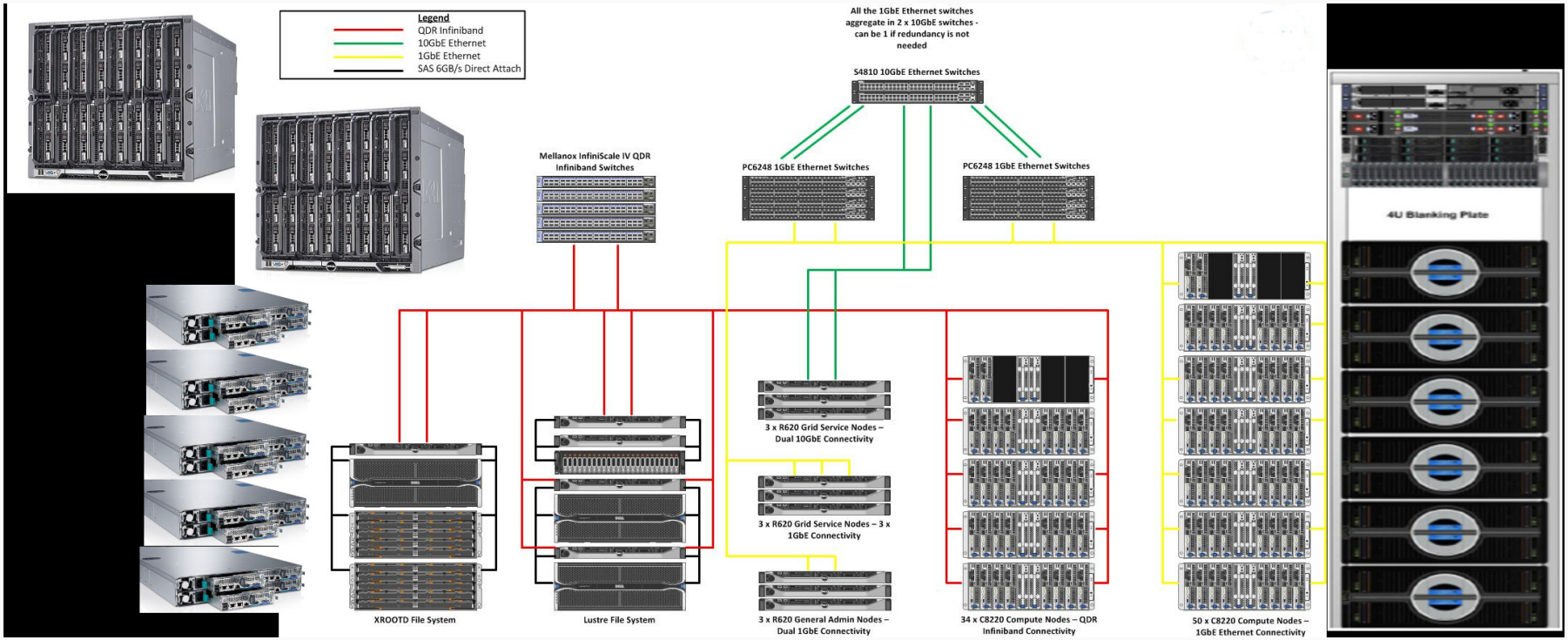

Robben Island
BLOUBERGSTRAND
DURBANVILLE
Klapmuts
KRAAIFONTEIN
MILNERTON
BRACKENFELL
BELLVILLE
GREEN POINT
Centre for High
Performance Computing
R300
Stellenbosch
Table
Mountain
National Park
HOUT BAY
Raithby
KHAYELITSHA
MACASSAR
SOMERSET WEST
STRAND
FISH HOEK
KOMMETJIE
GORDONS BA
SIMON'S TOWN
False Bay
Hottentots-H
Mountai
Catchment

Commitments

Resources Delivered and pledged.

expriment	index	2017	2017	2018	2018
ALICE	cpu	12000	6000	10000	10000
	disk	348TB	100TB	1500TB	1500TB
ATLAS	cpu	12000	0	10000	10000
	disk	262T0B	0TB	700TB	700TB

Computing Infrastructure



Current hardware

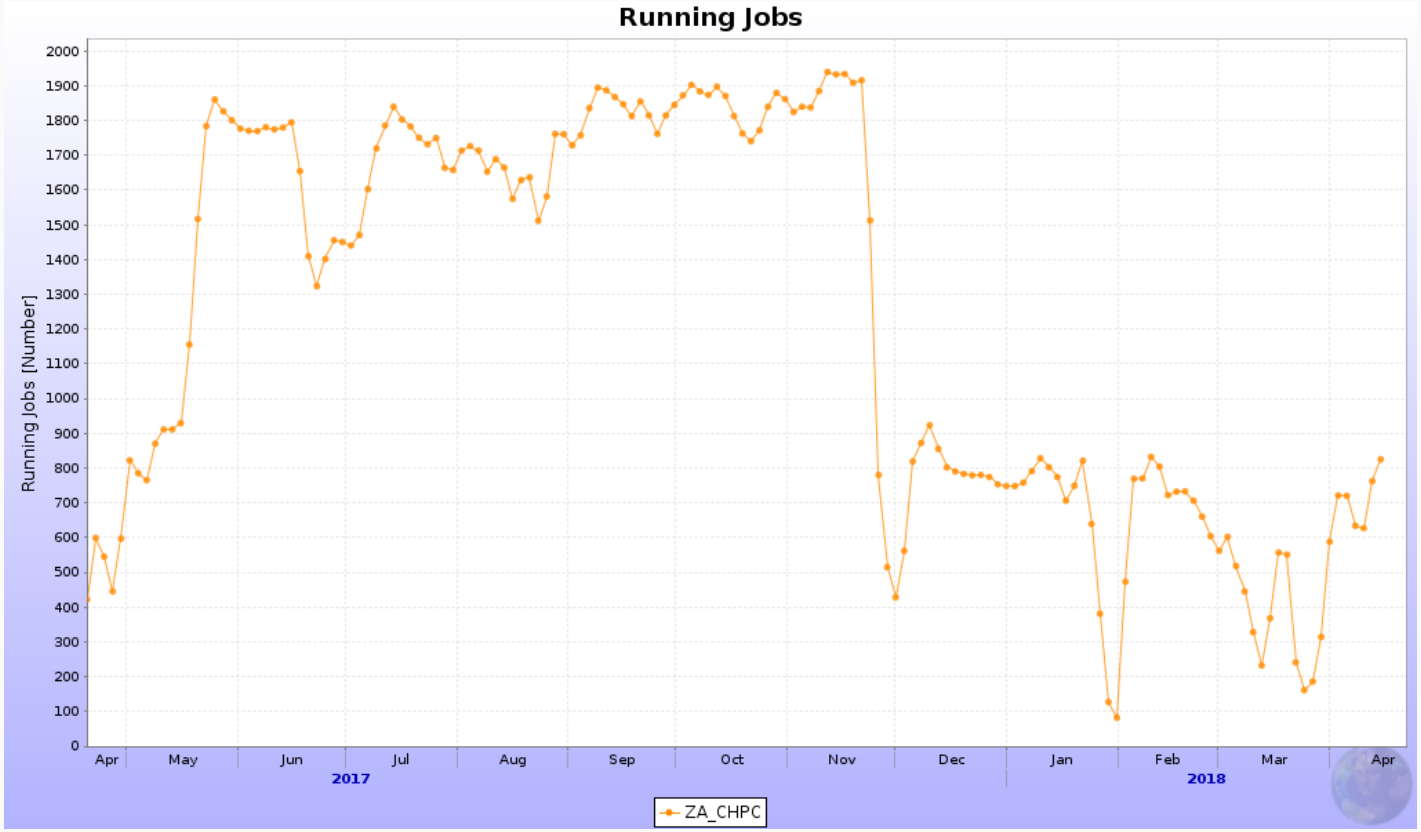
- 50 nodes of 48 cores 192GB RAM and 1.6TB of SSD, 2x bonded 1G ethernet
- 29 nodes of 48 cores 96GB RAM and 1TB, QDR infiniband.
- 2x M1000E (16 blades) ”new”
- 5xC6100 8x 6core xeon each. ”new”
- 9 management servers, lower spec
  - compute element (head node,ce),
  - storage element 2 redirectors, 2 storage nodes with direct attached multipath storage
  - authentication, monitoring, provisioning.

Current Storage

- 383TB EOS for ALICE
- 252TB EOS for ATLAS
- 1PB Lustre via QDR infiniband, no ups and Xyratex proprietary hardware.

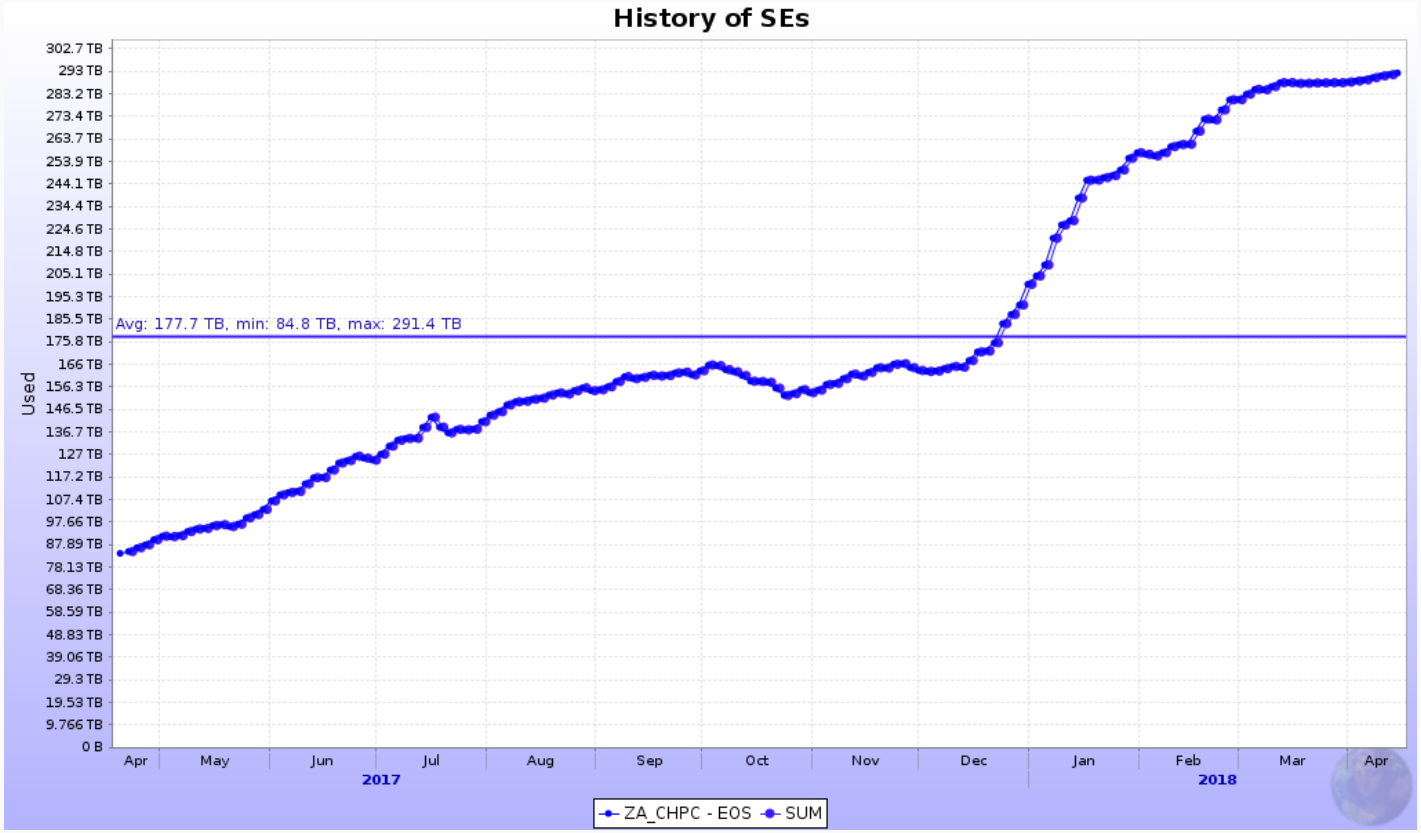
Storage is limited to c. 4Gbps which limits processing hence not all cores are allocated.

Current Performance last year

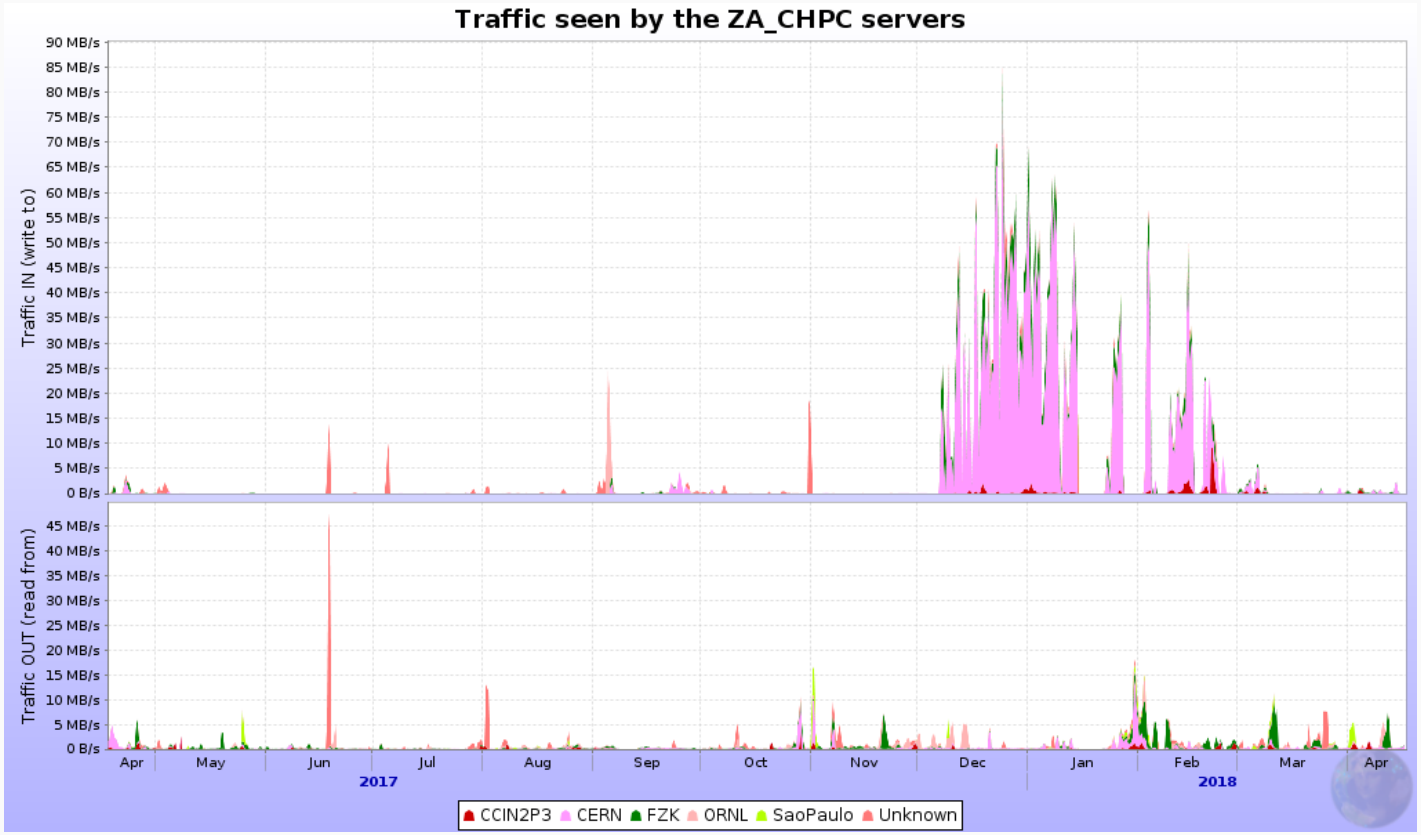


- 1.7M ALICE jobs in last year, 877k 2016, 455k 2015.
- 1222 Avg 2017 980 Avg 2016 704 Avg 2015.
- 293TB of 383TB, 91TB last year. Dec 15-Jan 29 we took in 100TB, 30% of our storage, mostly CERN and FZK

Graph of Storage Use



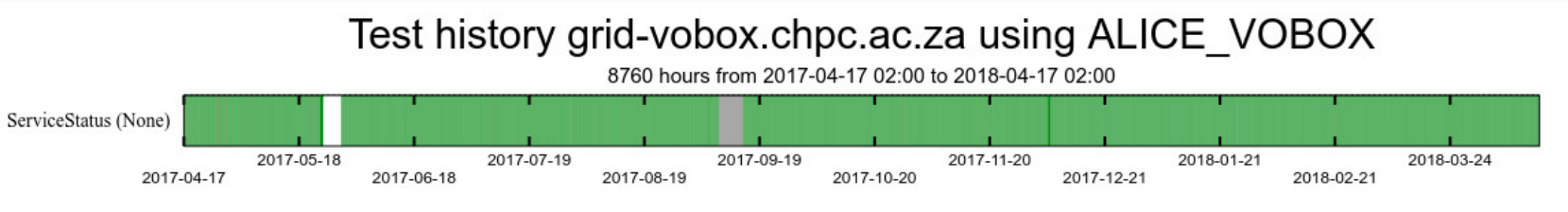
Graph of Storage Transfers

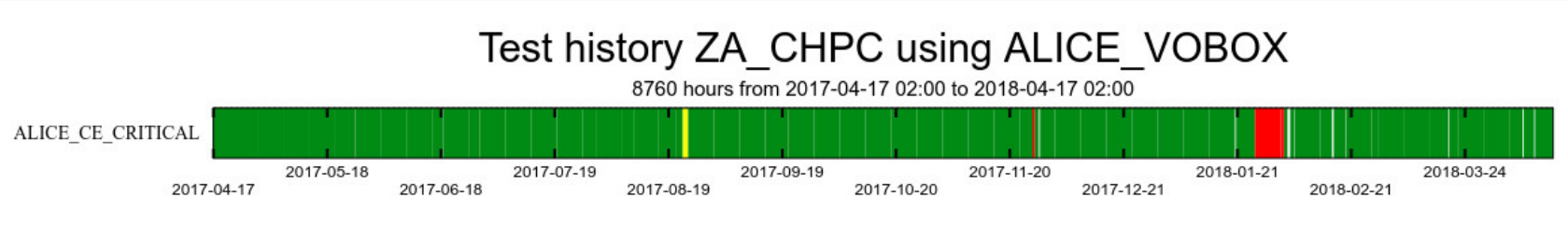


100TB

from 15 December to 29 January.

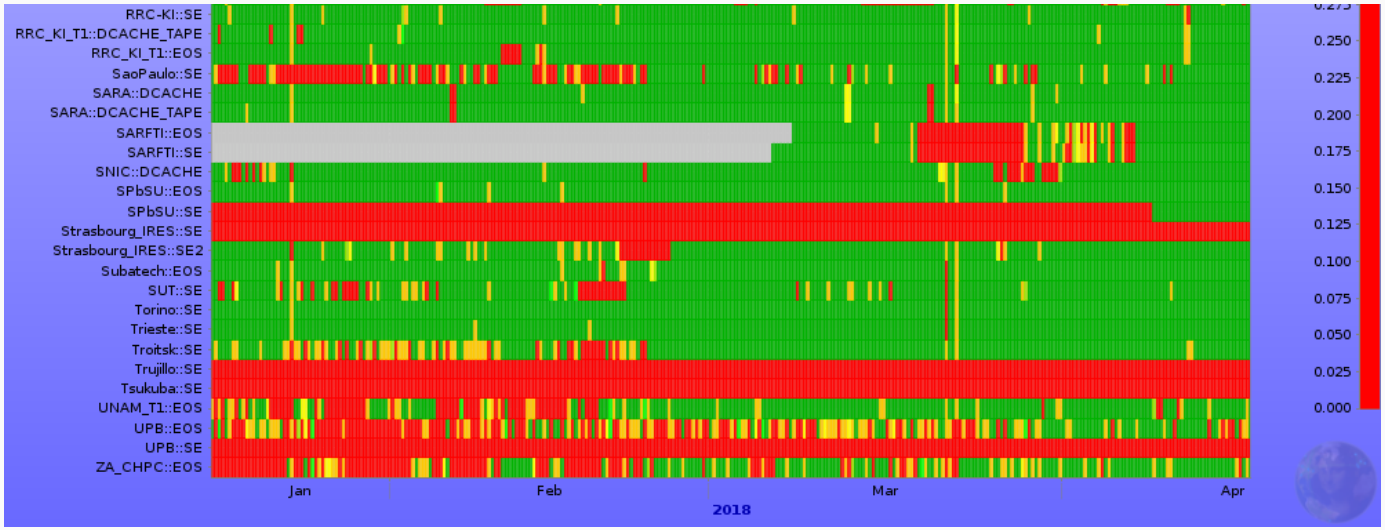
Availability / Reliability





Function	Last 365 days
Availability	96%
Reliability	95%
Storage	72%

EOS ”downtime”



Storage down time is primarily due to storage test failures due to network starvation. Its now gotten so bad it happens 7 or 8 times a week.

Downtimes

The big ones are, ignoring network starvation :

- 25 January Failure 1.5 week, to get new certificates.
- 27 Nov Melted busbar, we managed to patch power across and keep system up, while power was switched off and redone.

Network maximisation

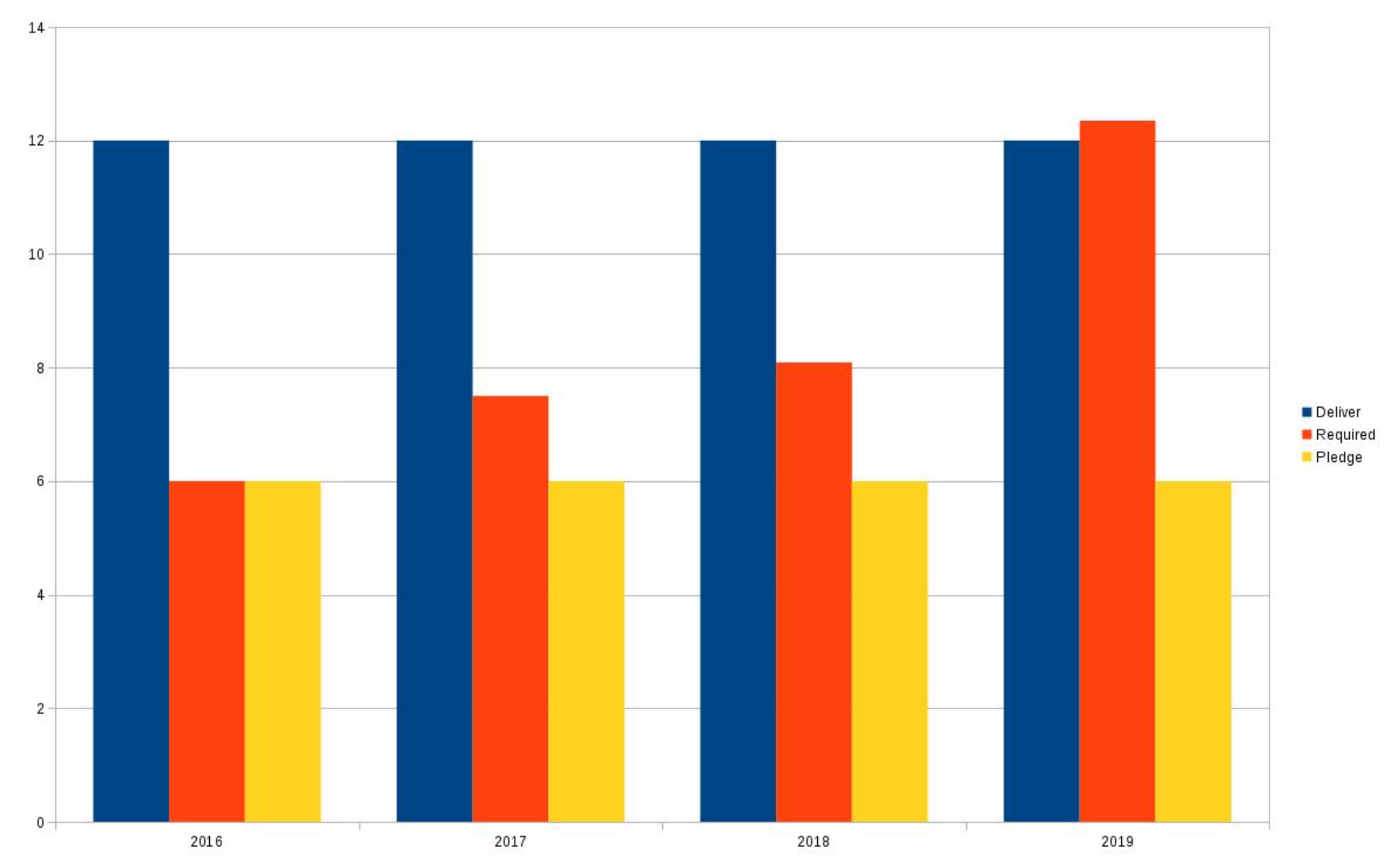
We have maxed out our network at 2.1Gbps regularly.

- CE at 1Gbps
- EOS at 1Gbps
- Vobox doing a 100Mbps.

provided and (required)

Resource	2016	2017	2018	2019
CPU kHS06	7.5 (6)	7.5 (7.5)	10(8.09)	15(12.35)
Storage TB	384 (550)	384 (682)	15000(1100)	??

Cpu graphically hepspec06-10



Outline

cpu growth

- No real plans here(money), migrate computing into the lower spec 29 nodes
- Project for openstack cloud onsite.
- We picked up 2 M1000e blade systems and 5 dell c6100.
- openshift on our M1000e and C6100 cpu additions. Build farm, testing, and batch when idle.
- We could allocate 3400 cores, hepscp06=12.

Storage

- After much frustation with the 100TB of Lustre it will become central storage for openshift

Storage

- After much frustation with the 100TB of Lustre it will become central storage for openshift
- We have 1PB of lustre, connected to 29 grid nodes and 1 M1000e via QDR.

Storage

- After much frustation with the 100TB of Lustre it will become central storage for openshift
- We have 1PB of lustre, connected to 29 grid nodes and 1 M1000e via QDR.
- we were allocated 2M ZAR, to meet pledges of storage, 1.5PB ALICE.

Storage

- After much frustation with the 100TB of Lustre it will become central storage for openshift
- We have 1PB of lustre, connected to 29 grid nodes and 1 M1000e via QDR.
- we were allocated 2M ZAR, to meet pledges of storage, 1.5PB ALICE.
- Addtional 2.4M ZAR found from excess budgets from SA-CERN collaboration excess funding.

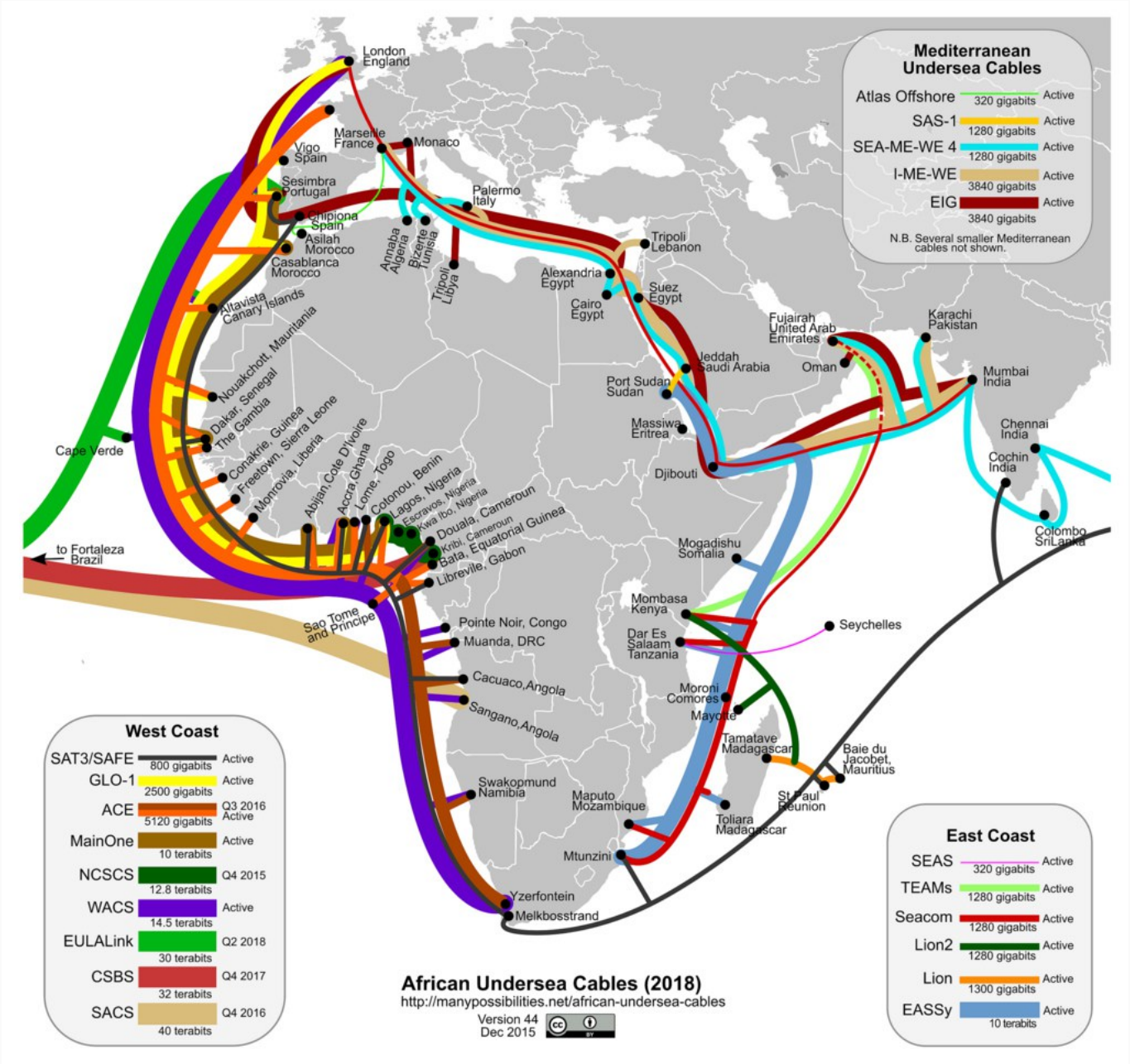
Storage

- After much frustation with the 100TB of Lustre it will become central storage for openshift
- We have 1PB of lustre, connected to 29 grid nodes and 1 M1000e via QDR.
- we were allocated 2M ZAR, to meet pledges of storage, 1.5PB ALICE.
- Addtional 2.4M ZAR found from excess budgets from SA-CERN collaboration excess funding.
- In Total 4.4M ZAR (280k EUR) for new EOS Storage, strings attached.

Storage

- After much frustation with the 100TB of Lustre it will become central storage for openshift
- We have 1PB of lustre, connected to 29 grid nodes and 1 M1000e via QDR.
- we were allocated 2M ZAR, to meet pledges of storage, 1.5PB ALICE.
- Addtional 2.4M ZAR found from excess budgets from SA-CERN collaboration excess funding.
- In Total 4.4M ZAR (280k EUR) for new EOS Storage, strings attached.
- Attempts to make EOS generic off lustre storage for site.

Network Connection



- 10G network installed
- new ipv4 and ipv6 allocated.
- Tested on 1 node on ipv4 and ipv6.
- still waiting for sfp purchase, currently using a donation.
- We are on the backbone, naked.
- We are still only paying for 212Mbps.

Outline

29 nodes

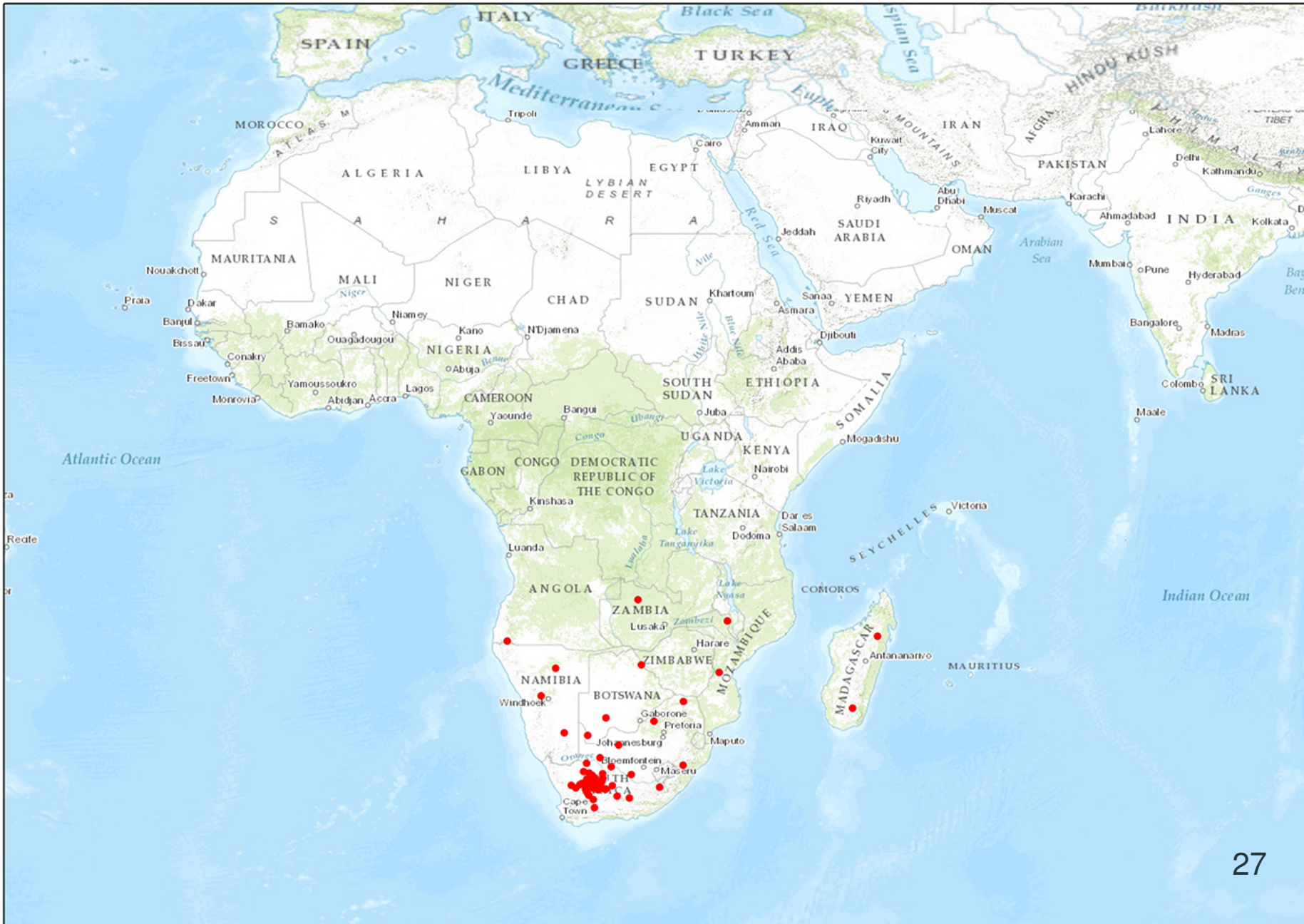
- 29 nodes local HEP, and grid, and generic usage.
- M1000e and C6100 for slush fund addition. (donations)
- code based on CODE-RADE, or LHC experiments from CVMFS.
- Local Storage for users, eos and beegfs.

The idea was to run the 29 nodes, 1392 cores with 1PB beegfs as a proof of concept of a analysis facility rather than a Tier1.

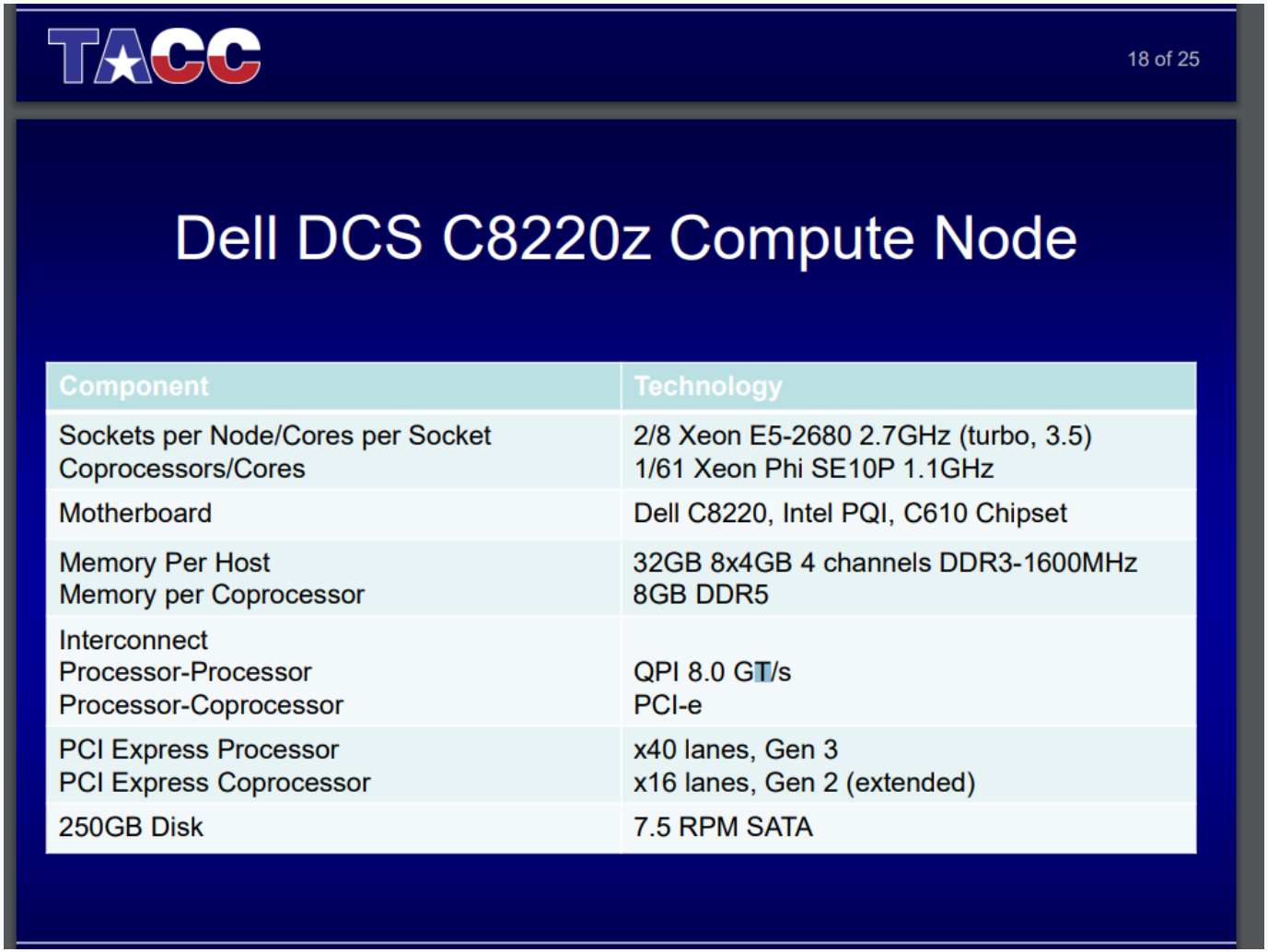
Outline

Square Kilometer Array

Square Kilometer Array, partners



Stampede



20 racks

currently mid-Atlantic.

Mozambique President

Mozambique

Mauritus

Mauritius

Mauritus

Outline

Summary

Things are improving.

- The big impediment of network is now a purely technical config issue, and money, we are using a donated network at the moment.
- Storage is shortly to exceed the required amount for the first time.
- Ample cpu.