CHPC WLCG Tier2 Facility and SA Grid Operations

Sean Murray
ALICE
CHPC
CSIR

October 24 2017

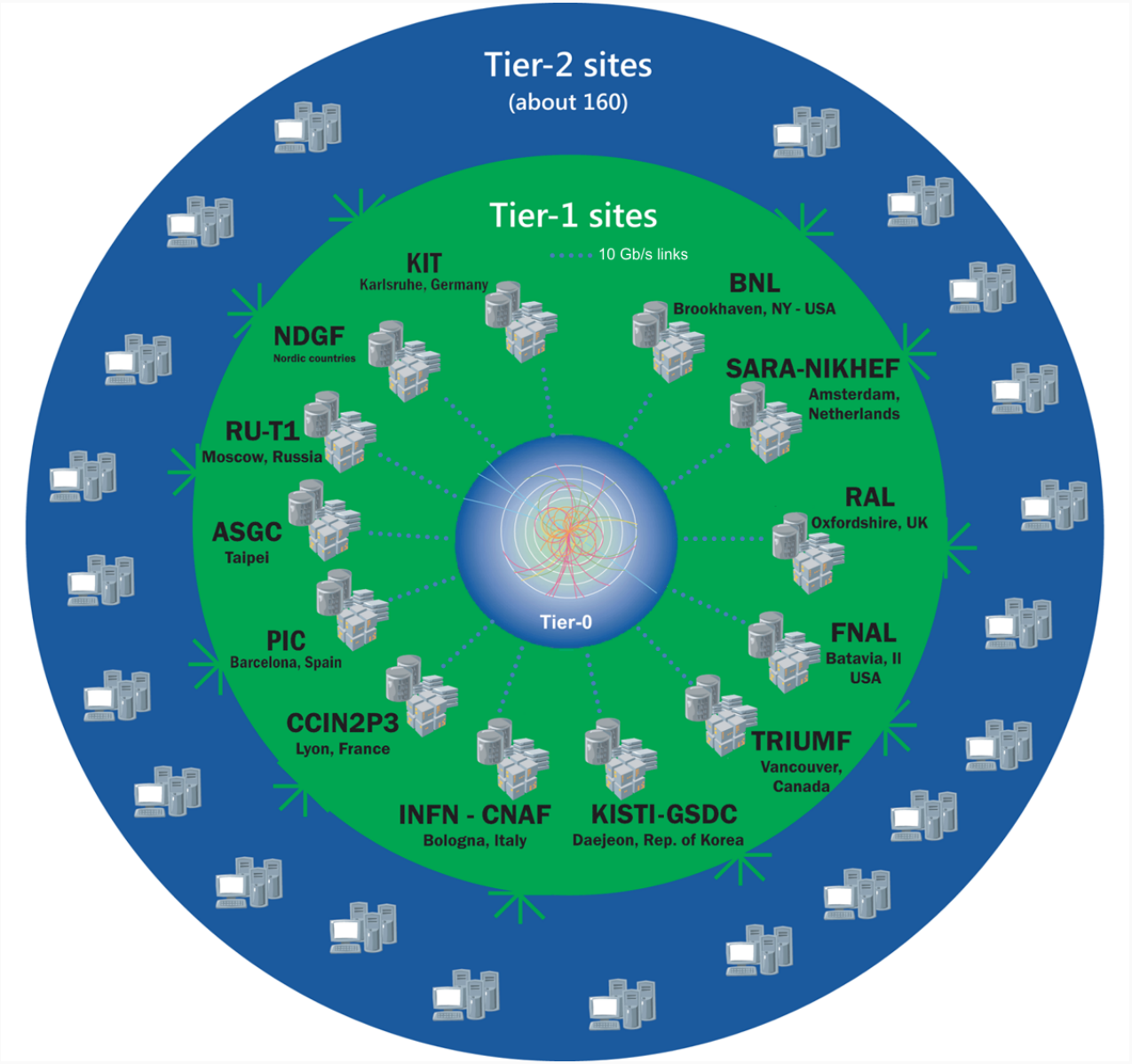
Outline

WLCG

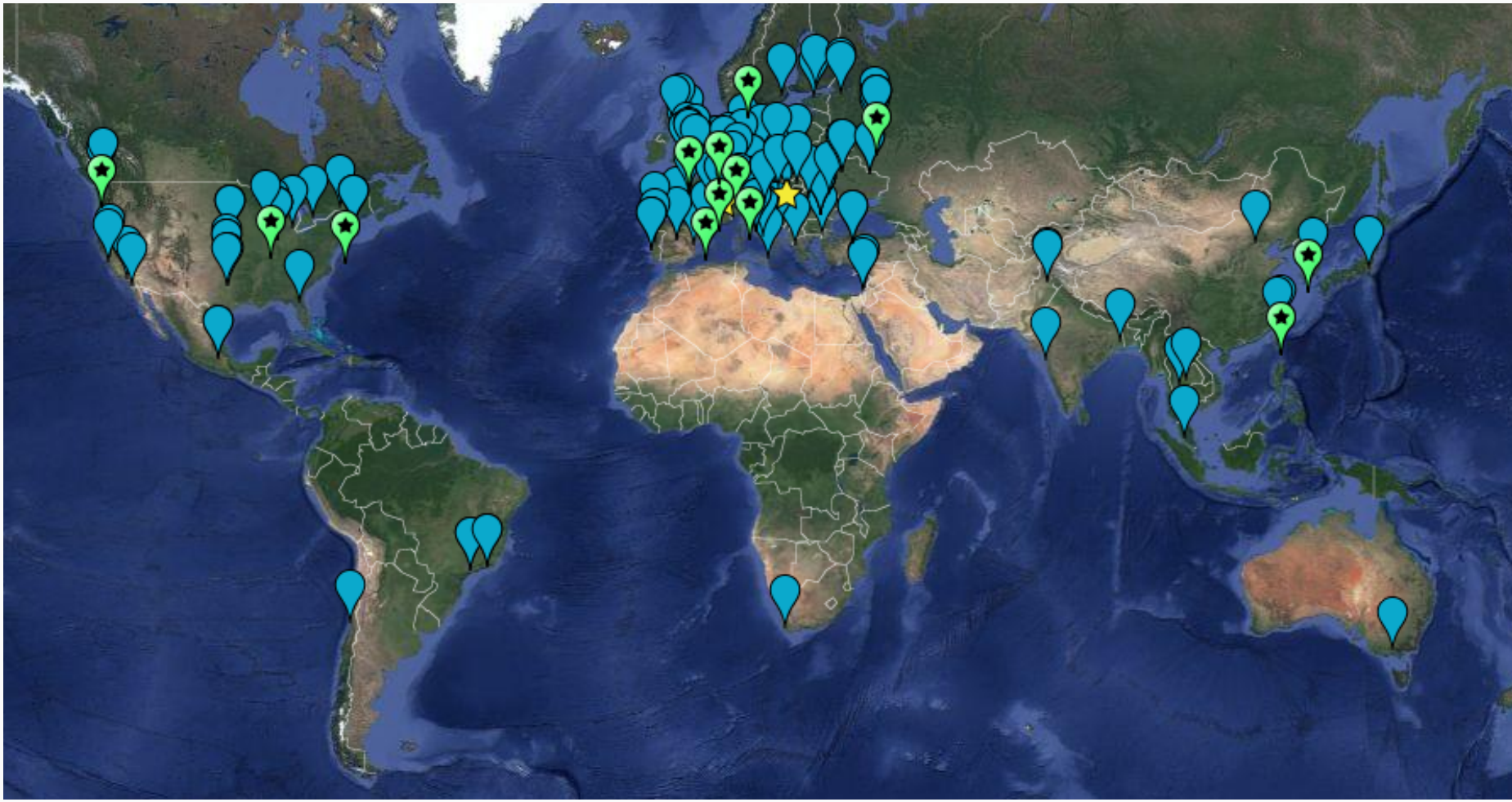
SAGrid, user analysis

Backup Slides

What is the WLCG



WLCG Map of Sites



Commitments

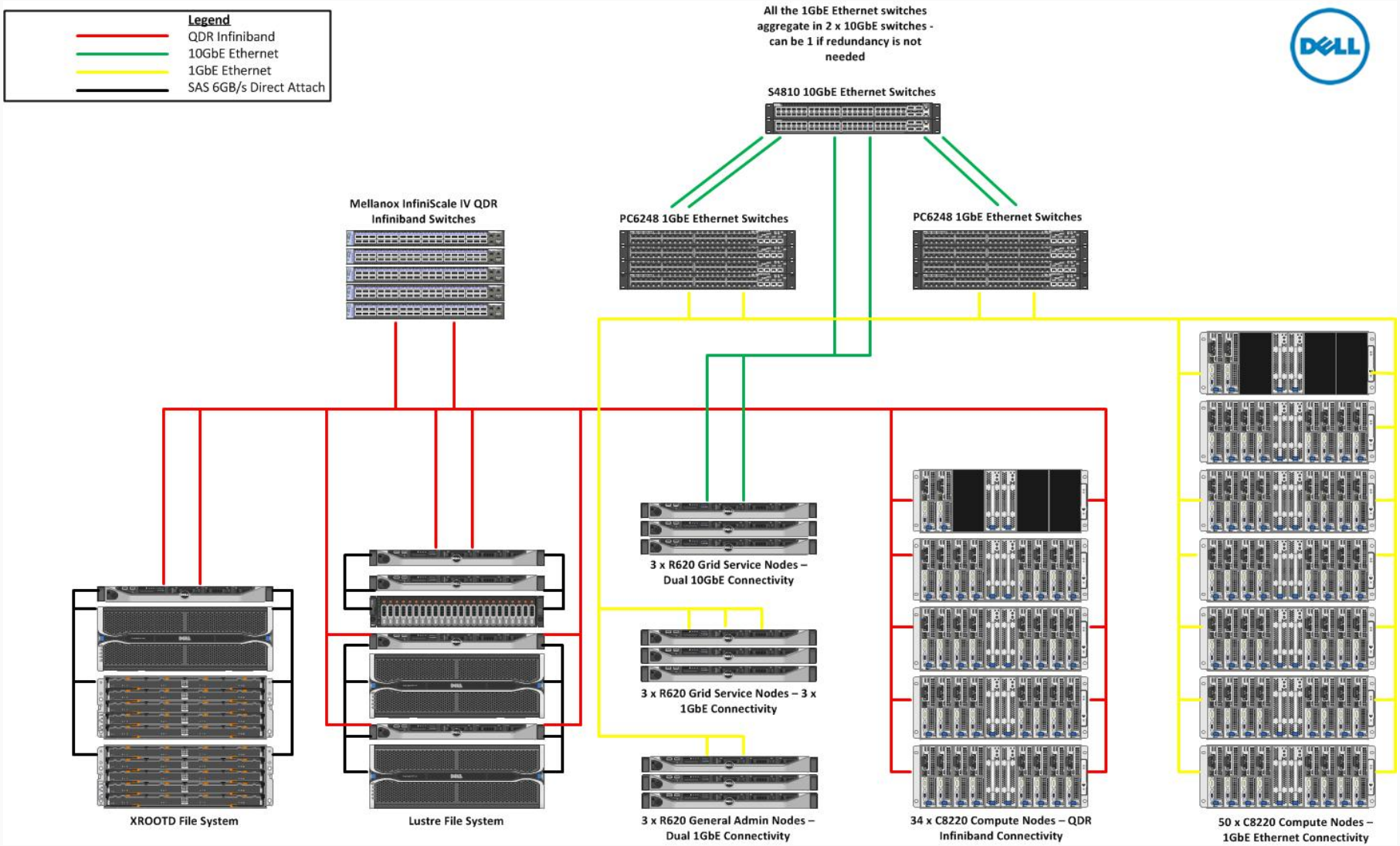
According to Tender :

- 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) [2018=10k]
- 100TB storage [2018=1.5PB and 800TB]
- All ALICE.
- now equal ALICE and ATLAS (?)
- 212Mbps International Traffic.

Computing Infrastructure



Current hardware

- 50 nodes of 48 cores 192GB RAM and 2x800TB of SSD, 2 bonded 1G.
- 28 nodes of 48 cores 96GB RAM and 1TB SATA, QDR infiniband
- c.100TB of Lustre on the 28 nodes with QDR infiniband. –dead
- c. 1PB luster 4 IB FDR IB, – dead migrating to beegfs/eos
- 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, site bdii.
- 16 blades of MD1000e, 384 cores. **IB connection**

Current Storage

- 383TB EOS for ALICE, down from 440TB
- 252 TB EOS for ATLAS, down from 400TB
- 107 TB lustre for 34 nodes.
- 1001 TB lustre 4 management nodes, 6 OSTs
- Quotes coming for 2PB EOS.
- 400TB ALICE
- 400TB ATLAS

Reduction in data sizes is due to reorganisation for reliability.

Storage questions

- c. 1PB hardware lock in ?
- c. 1PB eos/beegfs/lustre reinstall
- 2PB EOS configured how?

Availability / Reliability

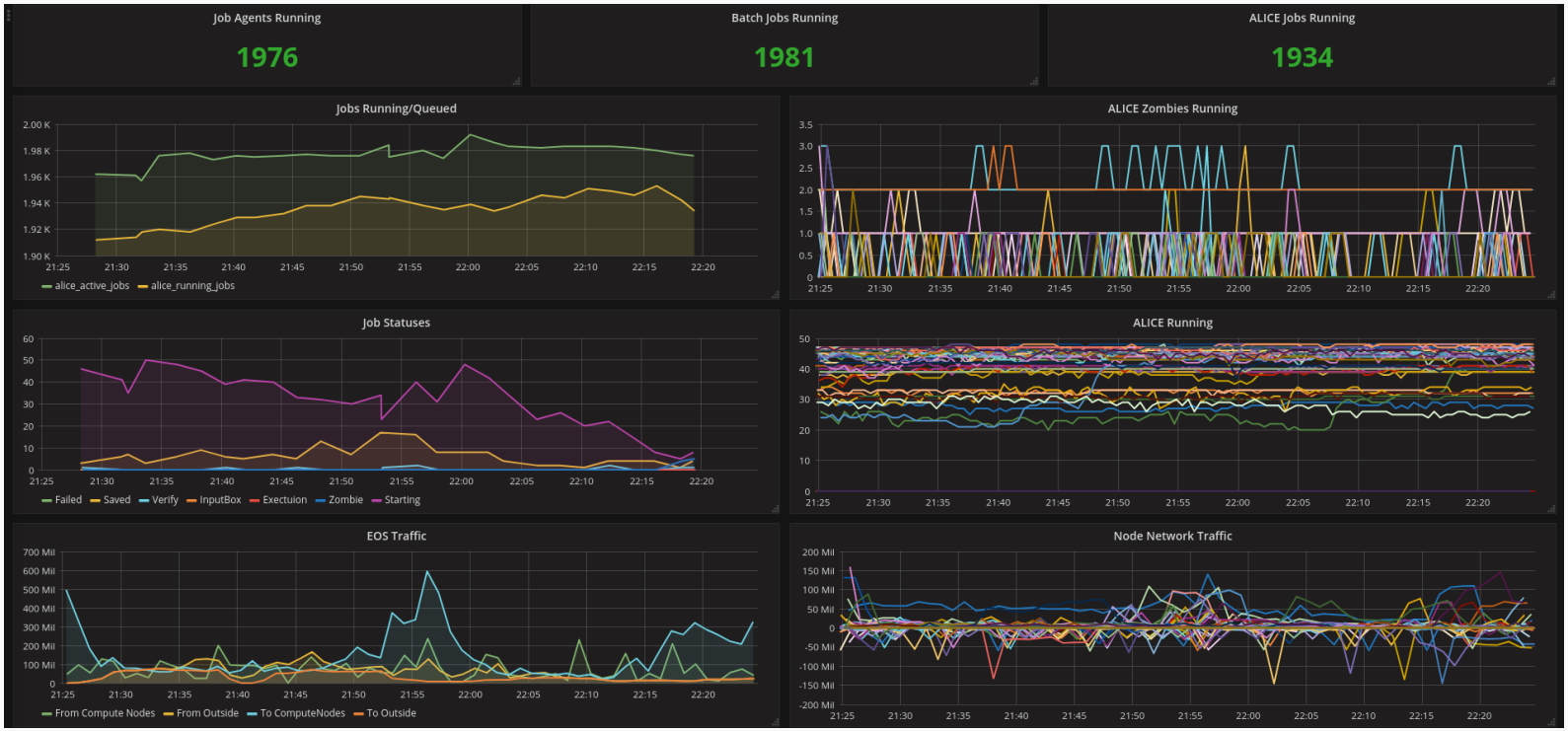
So who monitors us :

- WLCG
- EGI
- ALICE
- ATLAS
- Local via zabbix/grafana and racktables(replacing with foreman plugin)

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

Function	July	August	September	Year
Availability	100	100	100	98.6
Reliability	100	100	100	99.5

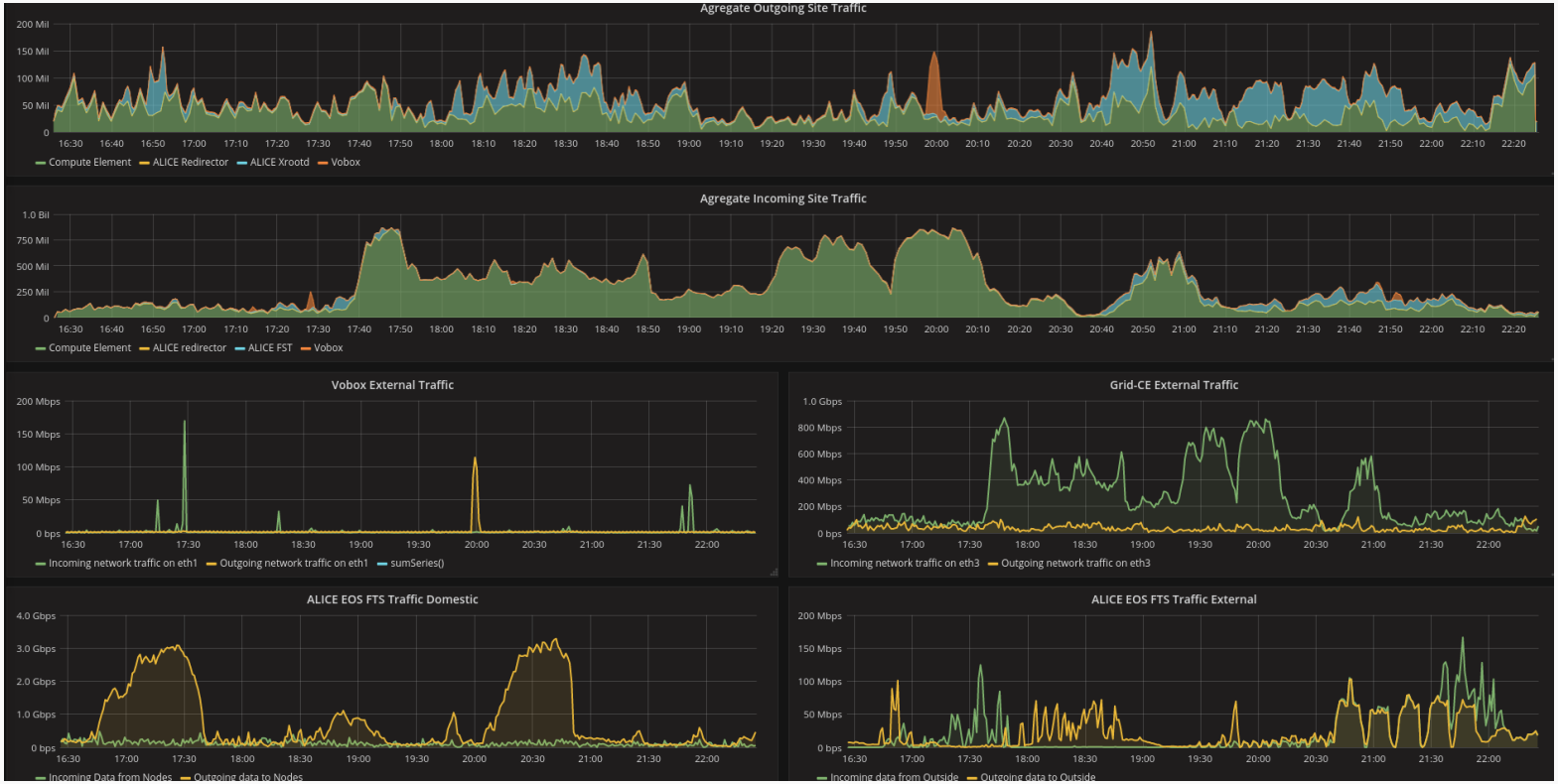
Grafana ALICE



This is currently being expanded to pull in more from MonaLisa, and more experiment and storage specific metrics to help to trivialy diagnose and forewarn problems. We want to know before everyone else.

How else are we to hide our errors ....

Grafana ALICE



Tier2 Processing

- ATLAS processing, pilot jobs start and then get work packages, data from Italy(T1).
- ALICE processing, job agents and leave it to central services, vobox.
- ops jobs fail due to starvation from alice. currently a dedicated machine, batch upgrade to fix.
- all software is distributed via cvmfs, both experiments.
- Both experiments have something similar to code-rade, an entire ecosystem of modules, versions, builds, etc.
- still no official validation on CC7.

Tier2 Storage operations

Thankfully both experiments use xrootd. ATLAS will soon deprecate SRM.

- 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, upgrade in eos required for ipv6(xrootd4).

Devops

- I am a software engineer and writing software is what I enjoy, preferably computational physics and hardware stuff.
- So in the well trodden path of a theoretical physicist.
- Think of a cow as a sphere .....
- Lets approximate sys admin as a software problem.
- I want everything in git and tested, and those tests automated.

Philosophy

- Keep It Simple Simpleton.
- Only use nonproprietary software.
- Stand on whom evers shoulders/heads/toes/knees I can. (Leverage the tools available)
- Abstract sys admin to a Software Engineering problem.
- Let me get back to the things that I enjoy ALICE O2, computational physics, machine learning etc.

Simple Solution

Switch everything off for 4 or 5 months while we redo **everything, BUT:**

Sadly we have an MOU to fullfill.

the plan

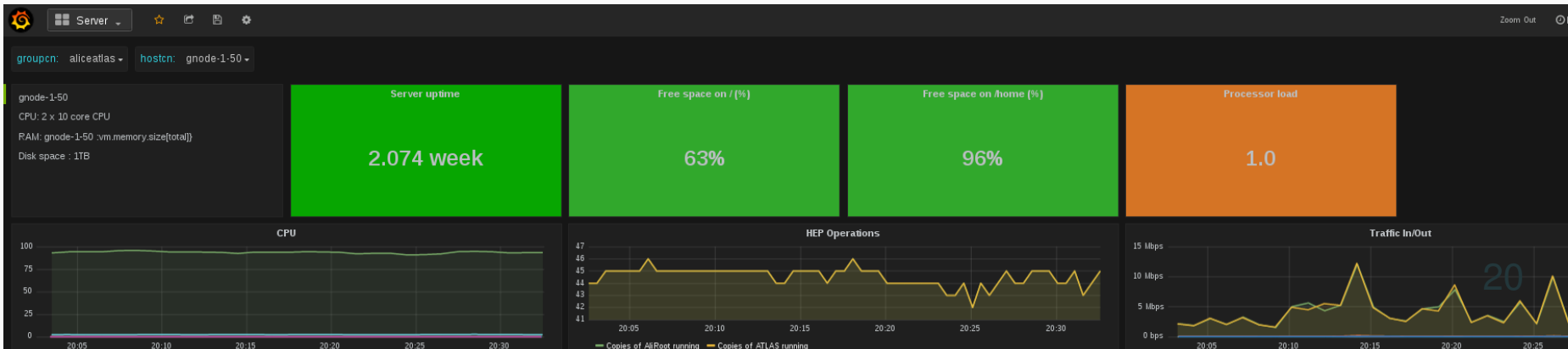
- Keep site up and processing maintaing our Tier 2 status.
- incrementally replace systems.
- We now have a MD1000e, on infiniband, a place to test configurations and deligate to batch we not playing.

upgrades

- attempts to delay till CC7 validation have failed.
- Puppify, to auto site deployment.
- Transition to foreman from xcat.
- 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 vobox and ce.
- Storage, we now have approval for getting quotes, to fullfil our pledges for next year.


ALICE
Zoom Out
Job Agents Running
1800
Batch Jobs Running
1800
ALICE Jobs Running
1797
Jobs Running/Queued
Batch Jobs Running
Batch Jobs Queued
Running Jobs
Jobs Agents
Job Statuses
alice_failed_jobs
alice_saved_jobs
alice_error_v_jobs
alice_error_jb_jobs
alice_error_e_jobs
alice_zombie_jobs
ALICE EOS
No datapoints
ALICE Zombies Running
ALICE Running


ATLAS
Zoom Out
Total ATLAS Jobs Running
3
Jobs Running/Queued
Total ATLAS Jobs Running
Total ATLAS Jobs Queued


Server
Zoom Out
groupcn: aliceatlas
hostcn: gnode-1-50
gnode-1-50
CPU: 2 x 10 core CPU
RAM: gnode-1-50 :vm.memory.size[total]}
Disk space : 1TB
Server uptime
2.074 week
Free space on / (%)
63%
Free space on /home (%)
96%
Processor load
1.0
CPU
HEP Operations
Copies of AliRoot running
Copies of ATLAS running
Traffic In/Out

Outline

28 nodes

28 of the 34 nodes came back for SAGrid and HEP when idle.

- 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.
- O2 dev for O2 development with dds/mesos/pbs. (no gpu/fpga)
- Local Storage for users, eos and lustre.

deployment/provisioning

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. This might change.

puppet

A method/system to maintain a system in a known state.
hiera, a key value store used by puppet.
We store all site specific information in hiera, tightly integrated into puppet..

AAROC Devops and r10k

- r10k branches are mapped to environments (dev, testing, production). Option to replace with foreman
- 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.
- all puppet except for eos is standard from other places, yaim is a pain.
- is mostly the R10k file and hiera for site specific information.

Testing

Devops requires some form of automated testing.

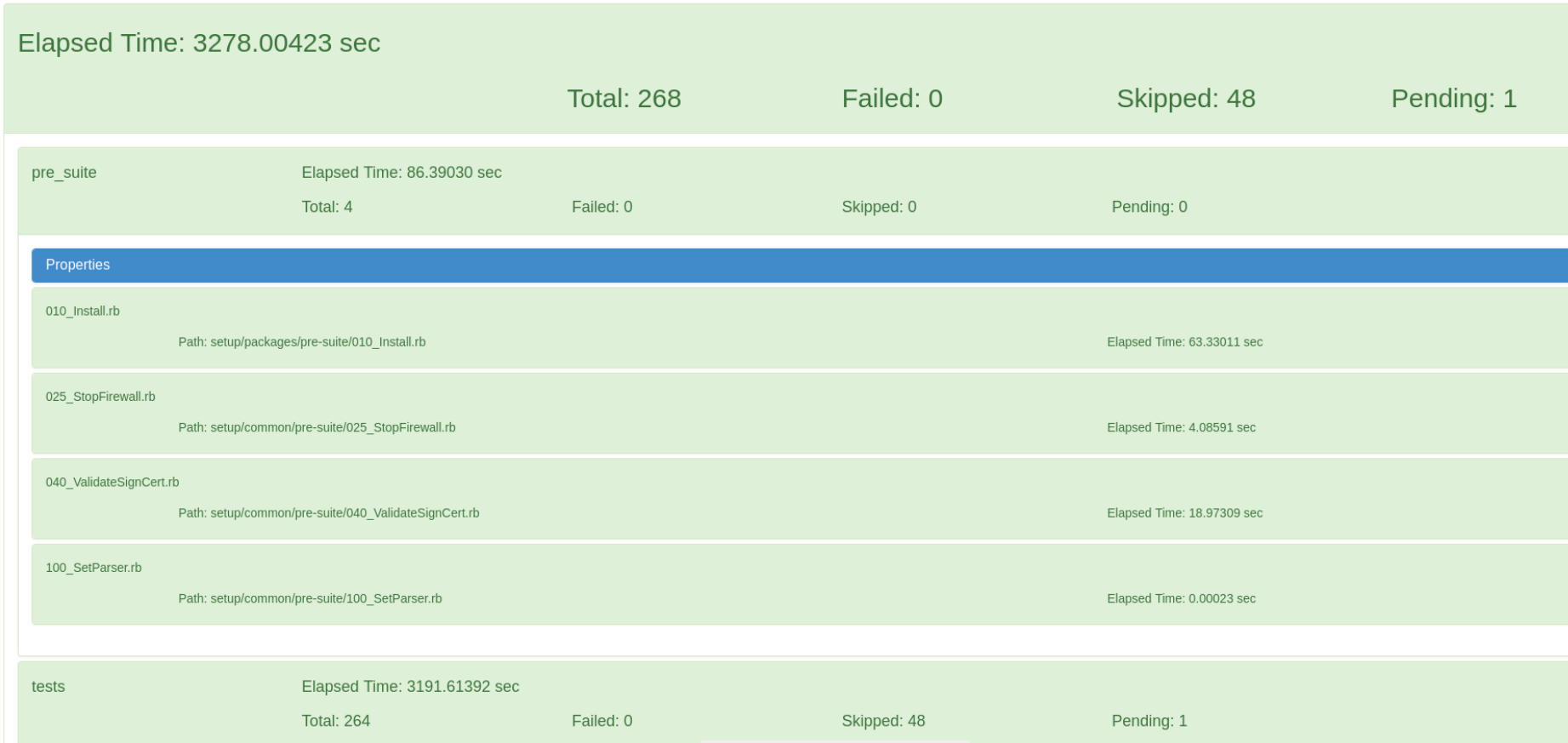
- Unit testing via rspec.
- Function testing.
- Fact testing.
- System testing.

Beaker

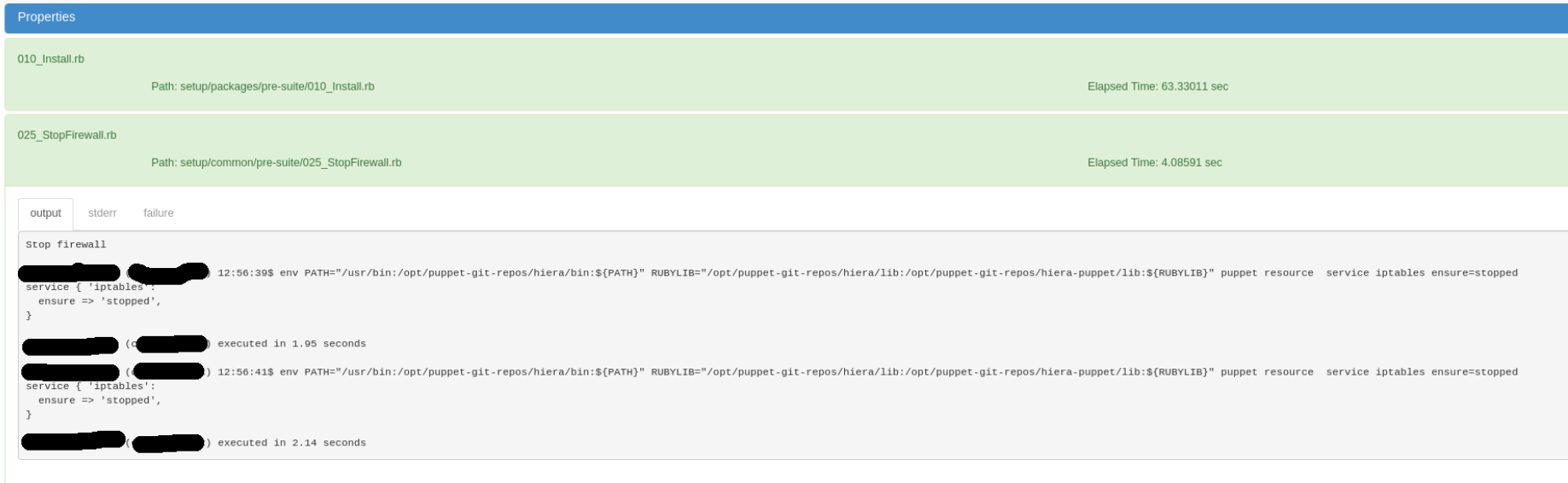
- Getting there ....
- Foreman templates written for some of beaker.
- Trying to integrate beaker tests and zabbix tests so tests are the same live and simulated. (pet project)

Beaker Pretty interface

Not surprisingly we get to a software engineering test case website ...



Beaker Pretty interface

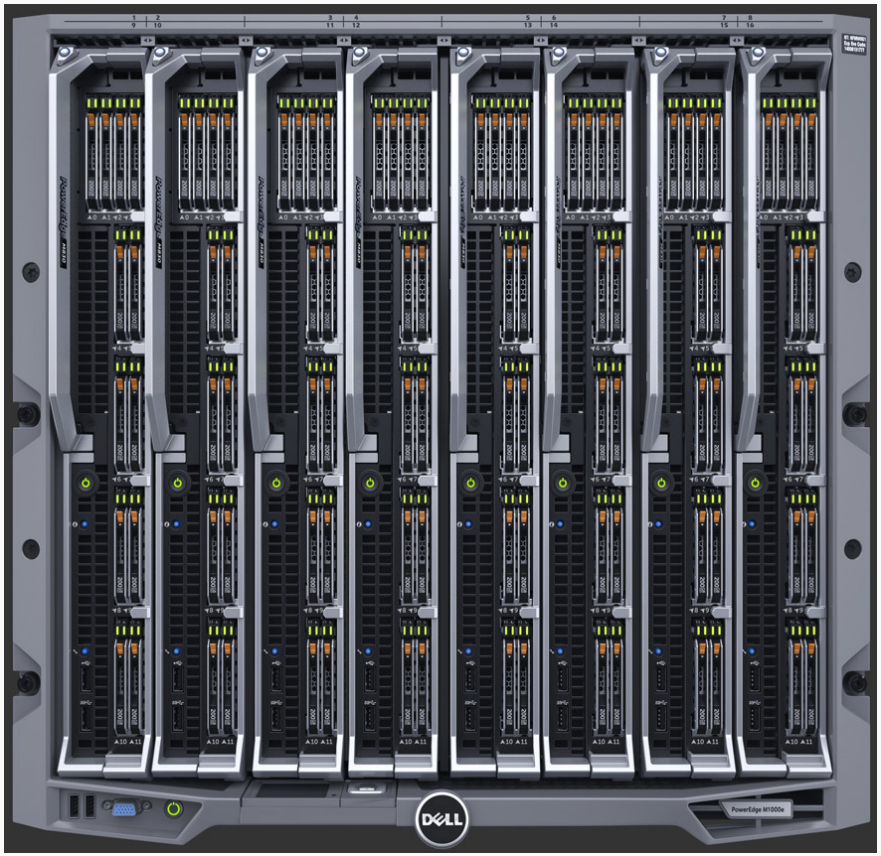


Where to test

This is currently done on my desktop .....

But not a standard dekstop

BUT



Monitoring

- zabbix all data ends up here.
- grafana on zabbix.
- racktables with zabbix link, foreman plugin seems to do an easier job for our purposes.
- foreman
- monalisa
- have singular sources of authority. idrac mostly.
- all zabbix templates, items etc. on github, not really devops, and deployed by node type.

TheForeman

- Deploy and maintain the entire lifecycle of a machine.
- Deploy to cloud (numerous),
- deploy containers.
- monitor health e.g. puppet status.
- Combined with Katello and pulp, maintain package configurations (ignored)

We currently use it as almost a cobbler replacement, this might change.

TheForeman

- Deploy and maintain the entire lifecycle of a machine.
- Deploy to cloud (numerous),
- deploy containers.
- monitor health e.g. puppet status.
- Combined with Katello and pulp, maintain package configurations (ignored)
- Why ditch so much functionality?

We currently use it as almost a cobbler replacement, this might change.

TheForeman

- Deploy and maintain the entire lifecycle of a machine.
- Deploy to cloud (numerous),
- deploy containers.
- monitor health e.g. puppet status.
- Combined with Katello and pulp, maintain package configurations (ignored)
- Why ditch so much functionality?
- puppet-yum gets us back to **everything** in a series of git repos. Remember where I am coming from.

We currently use it as almost a cobbler replacement, this might change.

TheForeman

- Deploy and maintain the entire lifecycle of a machine.
- Deploy to cloud (numerous),
- deploy containers.
- monitor health e.g. puppet status.
- Combined with Katello and pulp, maintain package configurations (ignored)
- Why ditch so much functionality?
- puppet-yum gets us back to **everything** in a series of git repos. Remember where I am coming from.
- Replace r10k (at least how we use it) and keep the different branches of puppet for dev/test/prod.

We currently use it as almost a cobbler replacement, this might change.

network

This is critically important for us.

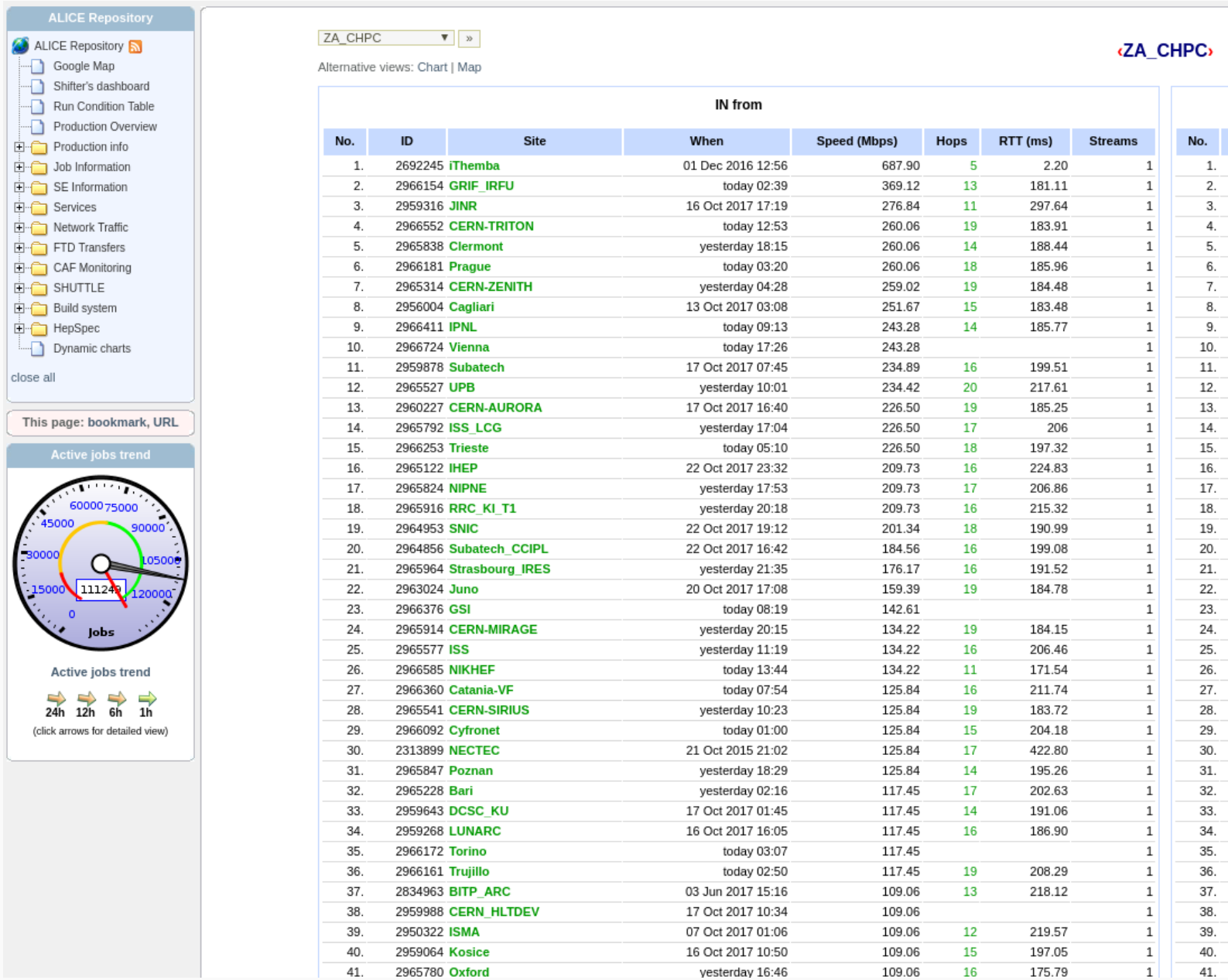
- All public interfaces are 1G cables across server room.
- change to dual 10G connections to switch.
- 10G fibre to SANREN.
- Shallow buffered swith, but in HA config.

Testing

We have been through a bout of testing, now moving to own network so can test further.

- Move outside CHPC network.
- Dedicated 10G to SANREN.
- PerfSonar installed.
- Test cleanly on our minimally buffered 10G Force10 Switch. Anybody want to loan us a 10G deep buffered switch to test.

Testing



No.	ID	Site	When	Speed (Mbps)	Hops	RTT (ms)	Streams	No.
1.	2692245	iThemba	01 Dec 2016 12:56	687.90	5	2.20	1	1.
2.	2966154	GRIF_IRFU	today 02:39	369.12	13	181.11	1	2.
3.	2959316	JINR	16 Oct 2017 17:19	276.84	11	297.64	1	3.
4.	2966552	CERN-TRITON	today 12:53	260.06	19	183.91	1	4.
5.	2965838	Clermont	yesterday 18:15	260.06	14	188.44	1	5.
6.	2966181	Prague	today 03:20	260.06	18	185.96	1	6.
7.	2965314	CERN-ZENITH	yesterday 04:28	259.02	19	184.48	1	7.
8.	2956004	Cagliari	13 Oct 2017 03:08	251.67	15	183.48	1	8.
9.	2966411	IPNL	today 09:13	243.28	14	185.77	1	9.
10.	2966724	Vienna	today 17:26	243.28			1	10.
11.	2959878	Subatech	17 Oct 2017 07:45	234.89	16	199.51	1	11.
12.	2965527	UPB	yesterday 10:01	234.42	20	217.61	1	12.
13.	2960227	CERN-AURORA	17 Oct 2017 16:40	226.50	19	185.25	1	13.
14.	2965792	ISS_LCG	yesterday 17:04	226.50	17	206	1	14.
15.	2966253	Trieste	today 05:10	226.50	18	197.32	1	15.
16.	2965122	IHEP	22 Oct 2017 23:32	209.73	16	224.83	1	16.
17.	2965824	NIPNE	yesterday 17:53	209.73	17	206.86	1	17.
18.	2965916	RRC_KI_T1	yesterday 20:18	209.73	16	215.32	1	18.
19.	2964953	SNIC	22 Oct 2017 19:12	201.34	18	190.99	1	19.
20.	2964856	Subatech_CCIPL	22 Oct 2017 16:42	184.56	16	199.08	1	20.
21.	2965964	Strasbourg_IRES	yesterday 21:35	176.17	16	191.52	1	21.
22.	2963024	Juno	20 Oct 2017 17:08	159.39	19	184.78	1	22.
23.	2966376	GSI	today 08:19	142.61			1	23.
24.	2965914	CERN-MIRAGE	yesterday 20:15	134.22	19	184.15	1	24.
25.	2965577	ISS	yesterday 11:19	134.22	16	206.46	1	25.
26.	2966585	NIKHEF	today 13:44	134.22	11	171.54	1	26.
27.	2966360	Catania-VF	today 07:54	125.84	16	211.74	1	27.
28.	2965541	CERN-SIRIUS	yesterday 10:23	125.84	19	183.72	1	28.
29.	2966092	Cyfronet	today 01:00	125.84	15	204.18	1	29.
30.	2313899	NECTEC	21 Oct 2015 21:02	125.84	17	422.80	1	30.
31.	2965847	Poznan	yesterday 18:29	125.84	14	195.26	1	31.
32.	2965228	Bari	yesterday 02:16	117.45	17	202.63	1	32.
33.	2959643	DCSC_KU	17 Oct 2017 01:45	117.45	14	191.06	1	33.
34.	2959268	LUNARC	16 Oct 2017 16:05	117.45	16	186.90	1	34.
35.	2966172	Torino	today 03:07	117.45			1	35.
36.	2966161	Trujillo	today 02:50	117.45	19	208.29	1	36.
37.	2834963	BITP_ARC	03 Jun 2017 15:16	109.06	13	218.12	1	37.
38.	2959988	CERN_HLTDEV	17 Oct 2017 10:34	109.06			1	38.
39.	2950322	ISMA	07 Oct 2017 01:06	109.06	12	219.57	1	39.
40.	2959064	Kosice	16 Oct 2017 10:50	109.06	15	197.05	1	40.
41.	2965780	Oxford	yesterday 16:46	109.06	16	175.79	1	41.

Flies in the ointment

- Just when you think you on top of it .....

Flies in the ointment

- Just when you think you on top of it .....
- Can not run tests on git update, waaaay to slow, but that is the least of the problems.

Flies in the ointment

- Just when you think you on top of it .....
- Can not run tests on git update, waaaay to slow, but that is the least of the problems.
- ALICE Job agents run for 28 hours

Flies in the ointment

- Just when you think you on top of it .....
- Can not run tests on git update, waaaay to slow, but that is the least of the problems.
- ALICE Job agents run for 28 hours
- ATLAS pilot jobs run for c. 48 hours.

Flies in the ointment

- Just when you think you on top of it .....
- Can not run tests on git update, waaaay to slow, but that is the least of the problems.
- ALICE Job agents run for 28 hours
- ATLAS pilot jobs run for c. 48 hours.
- Any real test would take c. 72 hours to test properly with out annoying people.

Flies in the ointment

- Just when you think you on top of it .....
- Can not run tests on git update, waaaay to slow, but that is the least of the problems.
- ALICE Job agents run for 28 hours
- ATLAS pilot jobs run for c. 48 hours.
- Any real test would take c. 72 hours to test properly with out annoying people.
- We could just not care ....

Flies in the ointment

- Just when you think you on top of it .....
- Can not run tests on git update, waaaay to slow, but that is the least of the problems.
- ALICE Job agents run for 28 hours
- ATLAS pilot jobs run for c. 48 hours.
- Any real test would take c. 72 hours to test properly with out annoying people.
- We could just not care ....
- But we want real tests, it helps you sleep at night.

bandwidth - Does ANYBODY ever have enough.

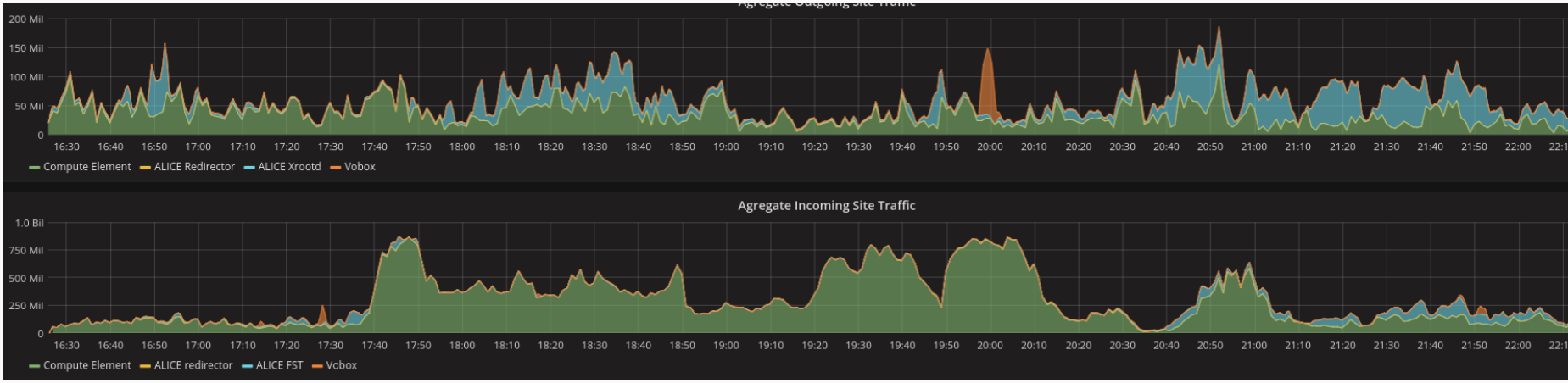
We were graciously donated a 1Gbps test.

We used it all.

We need more. . . . . .

We have cables on order and then will upgrade to 10Gbps for testing purposes.

bandwidth yesterday



Note the blue part, as our storage gets used the blue, international storage transfers becomes more significant than the rest. This will become even worse once we get our 2PB installed (and used).

Summary

Summary

- remove as much human contact as possible.

Summary

- remove as much human contact as possible.
- abstract as much away to bot and psuedo bots

Summary

- remove as much human contact as possible.
- abstract as much away to bot and psuedo bots
- let the science take priority.

Summary

- remove as much human contact as possible.
- abstract as much away to bot and psuedo bots
- let the science take priority.
- William and I can sit back, drink espresso and watch some graphs ....

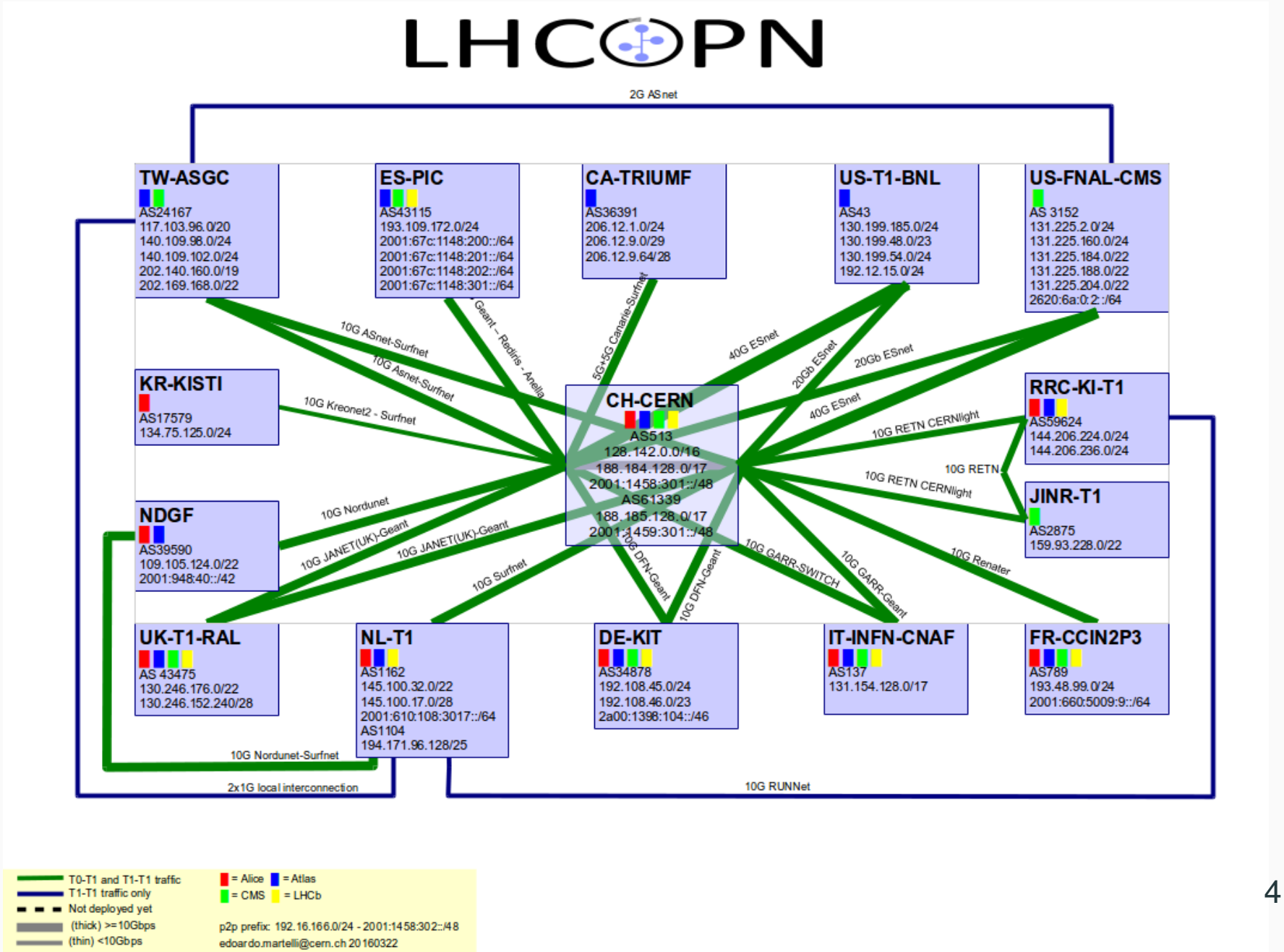
Summary

- remove as much human contact as possible.
- abstract as much away to bot and psuedo bots
- let the science take priority.

- FROM THE BEACH !

Outline

LHCOPN


LHCOPN
2G ASnet
TW-ASGC
AS24167
117.103.96.0/20
140.109.98.0/24
140.109.102.0/24
202.140.160.0/19
202.169.168.0/22
ES-PIC
AS43115
193.109.172.0/24
2001:67c:1148:200::/64
2001:67c:1148:201::/64
2001:67c:1148:202::/64
2001:67c:1148:301::/64
CA-TRIUMF
AS36391
206.12.1.0/24
206.12.9.0/29
206.12.9.64/28
US-T1-BNL
AS43
130.199.185.0/24
130.199.48.0/23
130.199.54.0/24
192.12.15.0/24
US-FNAL-CMS
AS 3152
131.225.2.0/24
131.225.160.0/24
131.225.184.0/22
131.225.188.0/22
131.225.204.0/22
2620:6a:0:2::/64
KR-KISTI
AS17579
134.75.125.0/24
CH-CERN
AS513
128.142.0.0/16
188.184.128.0/17
2001:1458:301::/48
AS61339
188.185.128.0/17
2001:1459:301::/48
RRC-KI-T1
AS59624
144.206.224.0/24
144.206.236.0/24
JINR-T1
AS2875
159.93.228.0/22
NDGF
AS39590
109.105.124.0/22
2001:948:40::/42
UK-T1-RAL
AS 43475
130.246.176.0/22
130.246.152.240/28
NL-T1
AS1162
145.100.32.0/22
145.100.17.0/28
2001:610:108:3017::/64
AS1104
194.171.96.128/25
DE-KIT
AS34878
192.108.45.0/24
192.108.46.0/23
2a00:1398:104::/46
IT-INFN-CNAF
AS137
131.154.128.0/17
FR-CCIN2P3
AS789
193.48.99.0/24
2001:660:5009:9::/64
10G ASnet-Surfnet
10G Asnet-Surfnet
Geant - Restris - Anella
5G+5G Canarie-Surfnet
40G ESnet
20Gb ESnet
20Gb ESnet
40G ESnet
10G Kreonet2 - Surfnet
10G RETN CERNlight
10G RETN
10G RETN CERNlight
10G Nordunet
10G JANET(UK)-Geant
10G JANET(UK)-Geant
10G Surfnet
10G DFN-Geant
10G DFN-Geant
10G GARR-SWITCH
10G GARR-Geant
10G Renater
10G Nordunet-Surfnet
2x1G local interconnection
10G RUNNet
T0-T1 and T1-T1 traffic
T1-T1 traffic only
Not deployed yet
(thick) >=10Gbps
(thin) <10Gbps
= Alice
= Atlas
= CMS
= LHCb
p2p prefix: 192.16.166.0/24 - 2001:1458:302::/48
edoardo.martelli@cern.ch 20160322

IOU table

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 accelerator operation	6 hours	6 hours	12 hours	99%	n/a
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 Centre[9] 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

mou table

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