



RAL Tier-1 Status and Plans

- also Oxford T2 -

Catalin Condurache

Rutherford Appleton Laboratory

STFC UK Research and Innovation

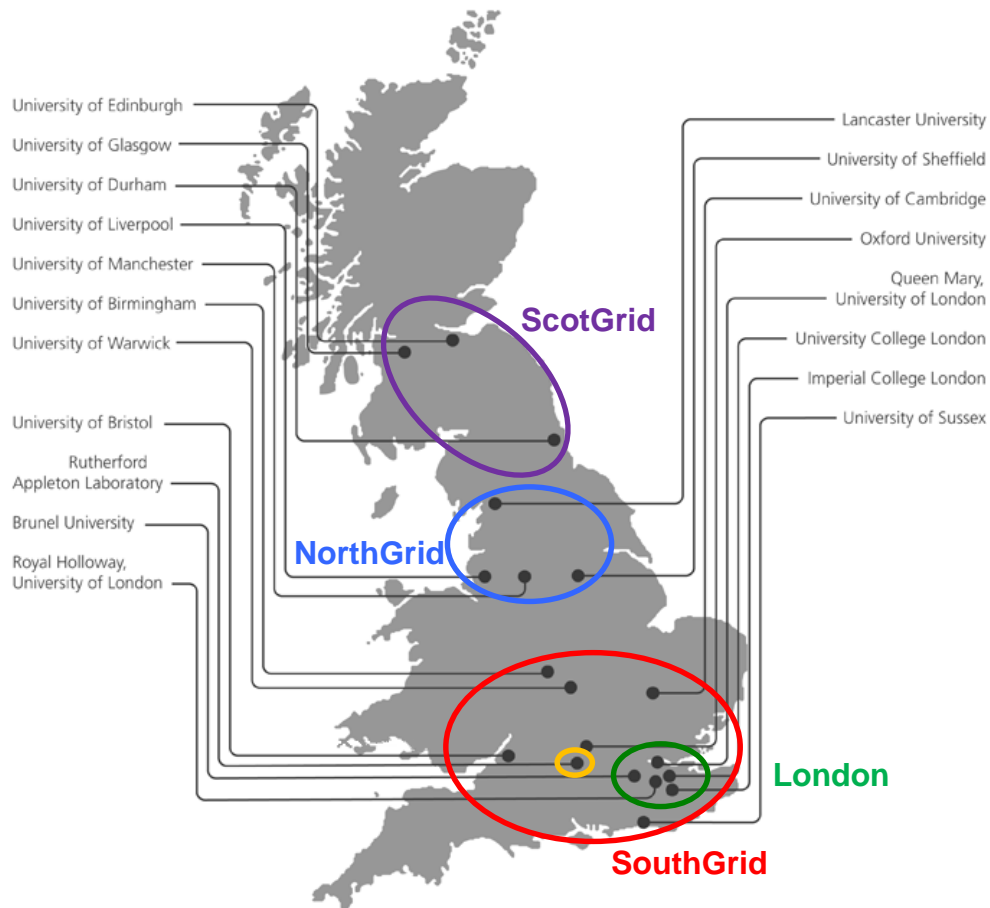
Outline

- UK GridPP Collaboration
- Since last report
- Oxford T2 status and plans
- RAL Tier-1 Centre
 - Components status and plans
 - ALICE highlights
- Storage for ALICE @ RAL
- UK funding status

Outline

- UK GridPP Collaboration
- Since last report
- Oxford T2 status and plans
- RAL Tier-1 Centre
 - Components status and plans
 - ALICE highlights
- Storage for ALICE @ RAL
- UK funding status

UK GridPP Collaboration



- Operating since 2002
- Makes case funding cases every ~4 years
- Now in cycle GridPP5
- ~10% of WLCG
- 19 institutes
- Tier-1 @RAL – supports all 4 LHC exps
- 4 distributed Tier-2 centres
- Policy: 10% of CPU and 5% of storage available to “non-LHC VOs”

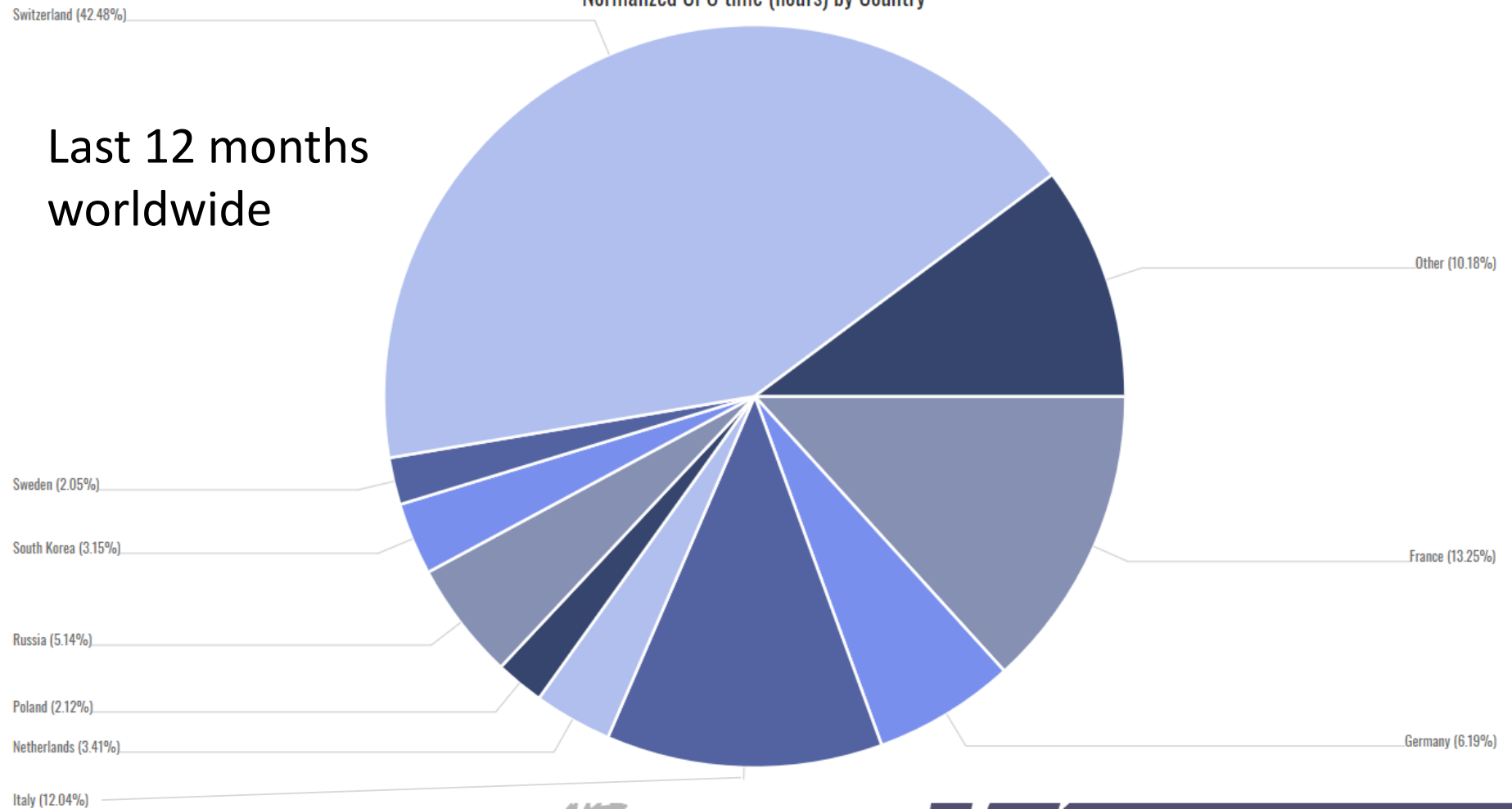
Outline

- UK GridPP Collaboration
- **Since last report**
- Oxford T2 status and plans
- RAL Tier-1 Centre
 - Components status and plans
 - ALICE highlights
- Storage for ALICE @ RAL
- UK funding status

ALICE CPU Accounting

Last 12 months
worldwide

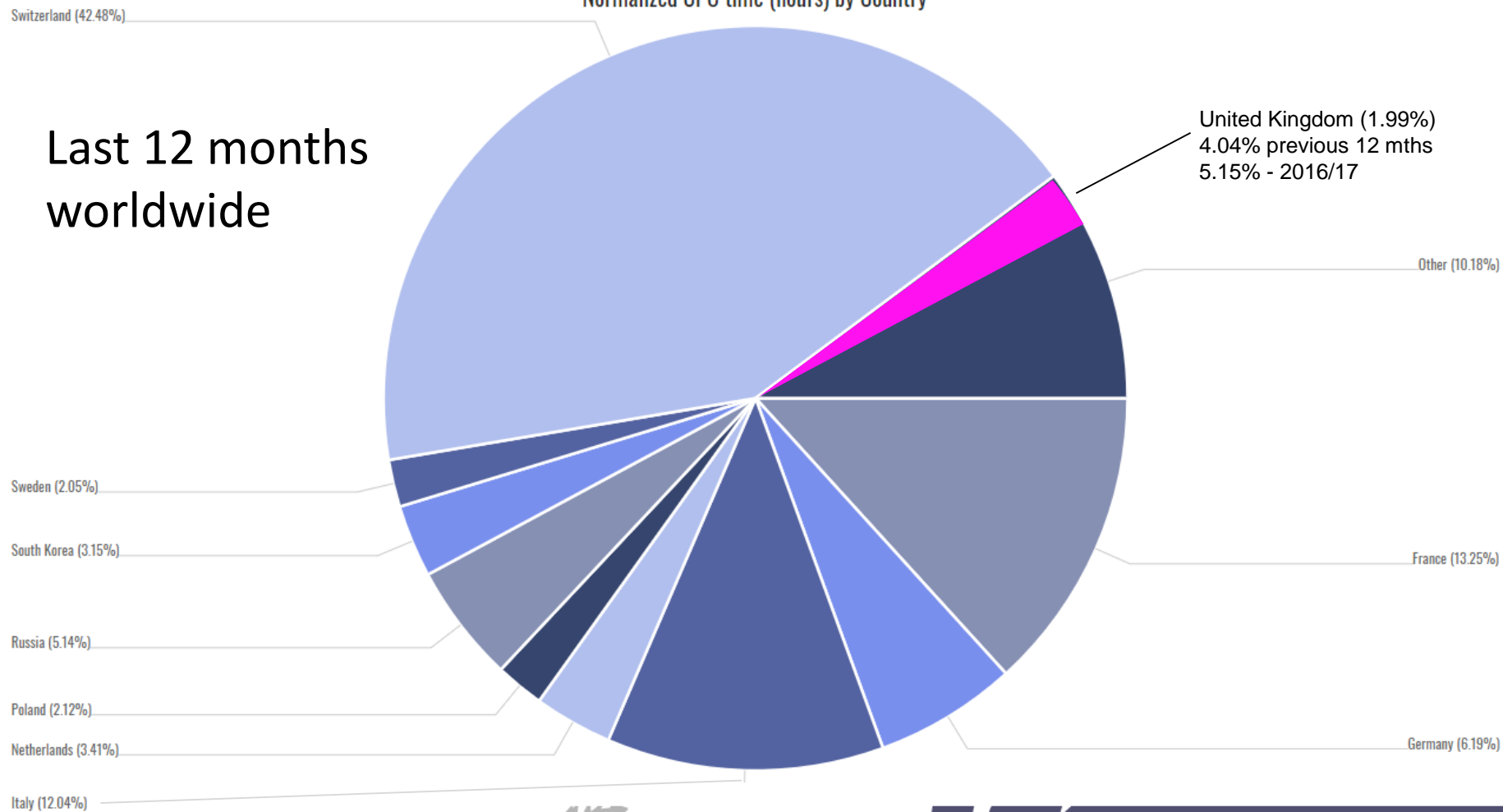
Normalized CPU time (hours) by Country



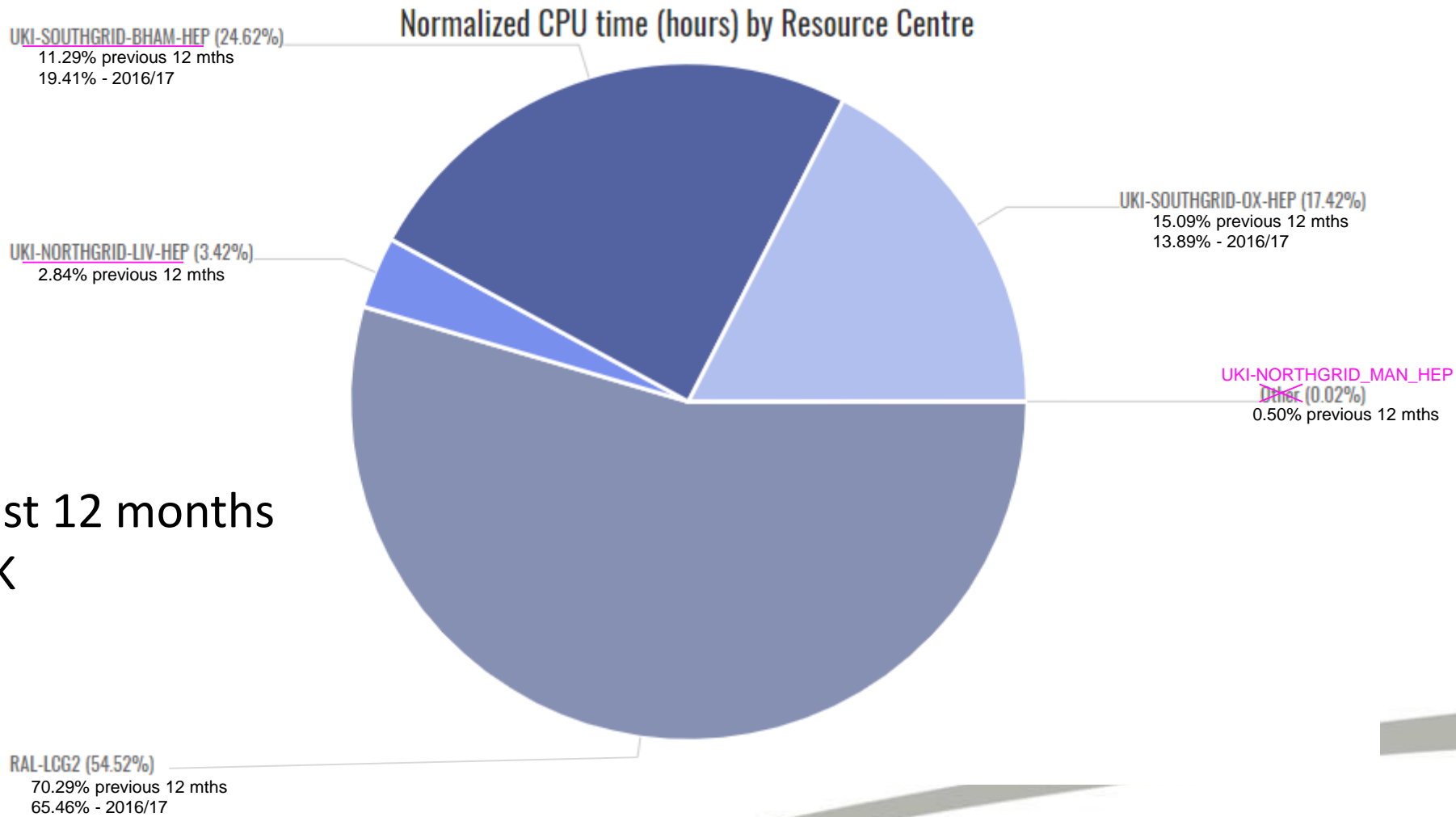
ALICE CPU Accounting

Last 12 months
worldwide

Normalized CPU time (hours) by Country



ALICE CPU Accounting

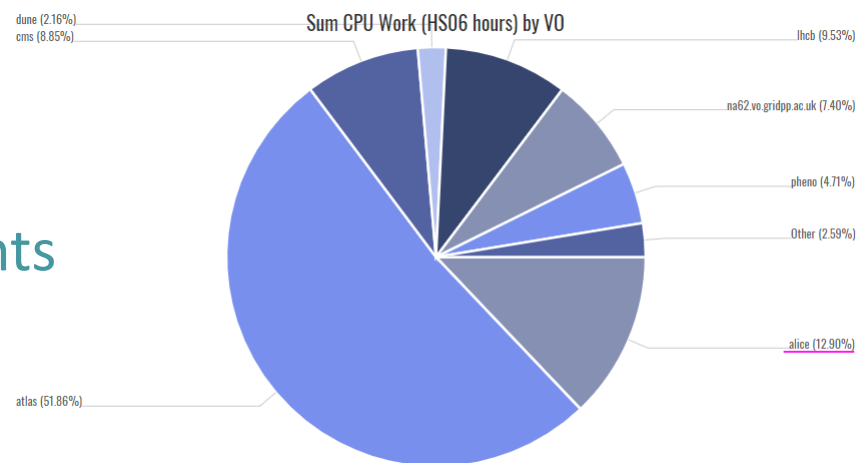


Outline

- UK GridPP Collaboration
- Since last report
- **Oxford T2 status and plans**
- RAL Tier-1 Centre
 - Components status and plans
 - ALICE highlights
- Storage for ALICE @ RAL
- UK funding status

Oxford Tier-2 Status and Plans

- *UKI-SOUTHGRID-BHAM-HEP*
- NO disk storage!
- Pledged to deliver 40% CPU of UK Tier-2 ALICE requirements



Oxford Tier-2 Status and Plans

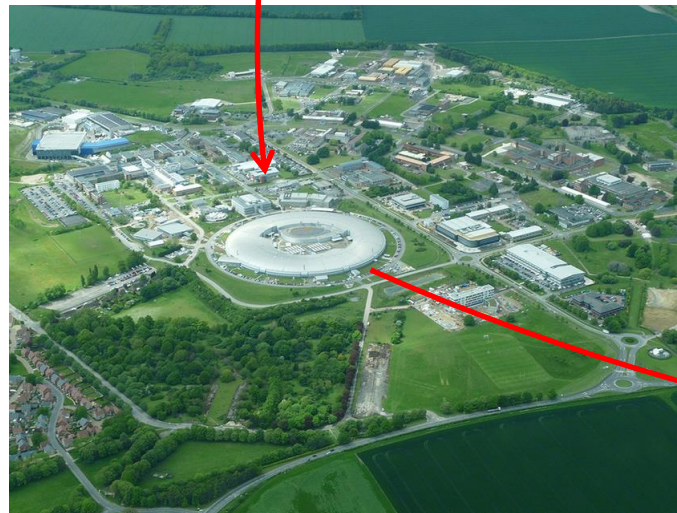
- Migrated to Centos7
- ALICE successfully running on new Centos7 cluster
- The Grid Cluster runs HT Condor behind ARC-CE
- 500 slots reserved for ALICE
 - when there is no other work => no of slots increased
- SL6 CE decommissioned
- VOBOX still SL6 – should upgrade to Centos7?
- “ALICE support is fantastic!” - Kashif Mohammad

Outline

- UK GridPP Collaboration
- Since last report
- Oxford T2 status and plans
- **RAL Tier-1 Centre**
 - **Components status and plans**
 - **ALICE highlights**
- Storage for ALICE @ RAL
- UK funding status

RAL Tier-1 Centre

- *RAL-LCG2*
- Hosted and run by STFC Rutherford Appleton Laboratory
- 15 miles south of Oxford on Harwell Campus



RAL Tier-1 Centre - Hardware

- CPU: ~350k HS06 (~30k cores)
 - FY18/19: procurement 2664 cores, ~34k HS06
- Storage
 - Castor: ~5PB usable – retirement in progress
 - Ceph: ~36PM raw / ~28PB usable
 - FY18/19: additional 11.3PB raw / 8.2PB usable
 - Acceptance testing
 - Tape: 10k slot SL8500
 - 80PB capacity (T10KD)
 - ~46PB physics data

RAL Tier-1 Centre – Batch Farm

- Currently ~30,000 slots over ~760 nodes providing ~350,000 HS06 (WLCG pledge is 272,725 HS06)
- Running HTCondor + ARC-CE
 - Starts jobs inside Docker container images
 - <https://hub.docker.com/r/stfc/grid-workernode-c6>
 - <https://hub.docker.com/r/stfc/grid-workernode-c7>
 - decouples the execution environment from the base OS
- Cloud bursting
 - Small amount of WNs in SCD OpenStack cloud when resources are available
- Currently evaluation of ARC6 and HTCondor-CE

RAL Tier-1 Centre – Batch Farm

- Since Aug'18 we have been running latest mainline kernel builds from ELRepo
 - Stable overlays implementation
 - Re-implemented memory management sub-system
 - Both required for a reliable container infrastructure
- ~11k cores partitioned off for only multi-core jobs
 - Better scheduling as draining of single-core slots happens far less aggressively
 - Partitioned may be increased in the future as the job mix changes

RAL Tier-1 Centre – Storage - Castor

- Past 12 months
 - Work to decommission ‘disk-only’ CASTOR pools in favour of Ceph-based storage and move tape access to shared CASTOR cache pool
 - ATLAS - 31 Oct 2018
 - Non-LHC users – 15 Nov 2018
 - CMS – 5 Dec 2018
 - LHCb – expected 2019Q2
 - ALICE – 2019Q{3,4} ?

RAL Tier-1 Centre – Storage - Castor

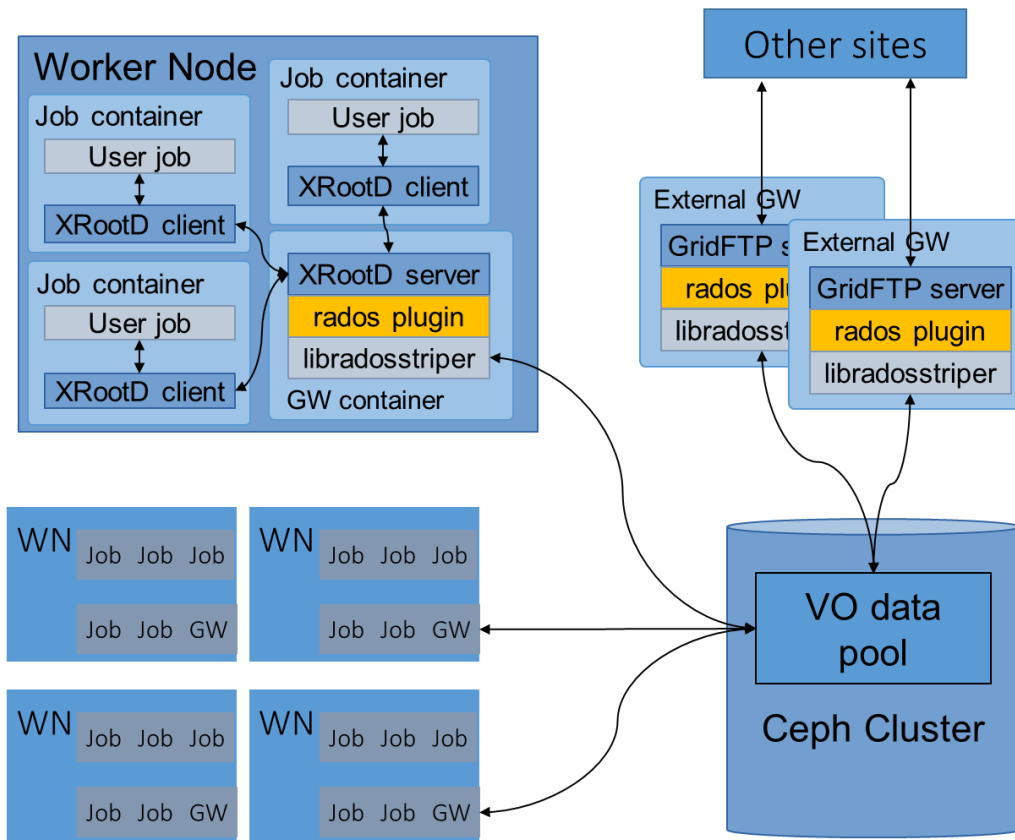
- New tape cache pool
 - CASTOR v2.1.17-35
 - New Oracle RAC, re-implemented configuration
- Once all users migrated, will enable 'Recommended Access Order'
 - Optimized order in which to read tape

RAL Tier-1 Centre – Storage - Echo

- Erasure-Coded High-throughput Object store
- Big Ceph cluster for Tier-1 object storage
 - 181 storage nodes
 - 36/28PB raw/usable – 16PB data stored
 - Erasure Coded (8+3) data pools
 - 64MB rados objects
- Three major Ceph versions so far
 - Currently Luminous

RAL Tier-1 Centre – Storage - Echo

- Tier-1 batch farm and external access to Echo

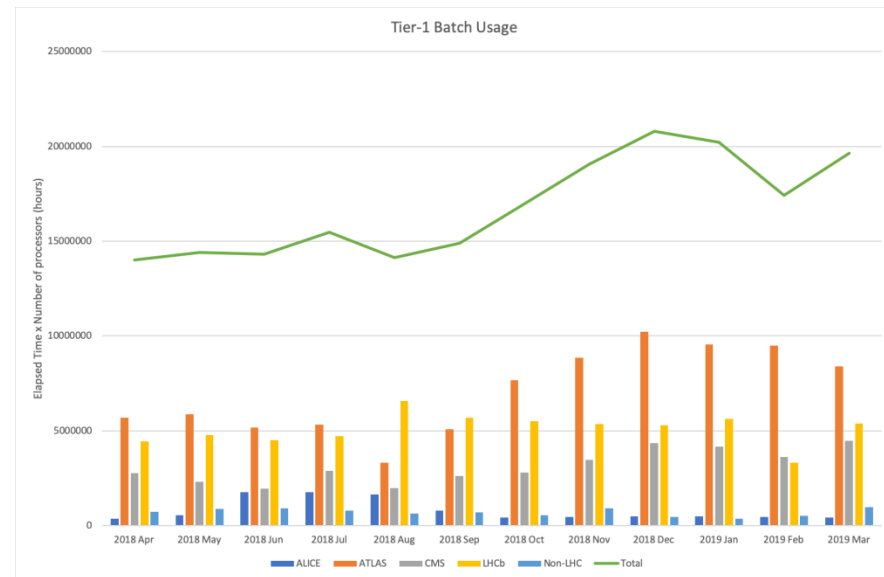


RAL Tier-1 Centre – Misc.

- Dual stack IPv4/6 available on Tier-1 network
 - Current issue – IPv6 packets loss
- 100 Gb/s connectivity to RAL site active since Summer 2018
- Some/slow progress on LHCONE connectivity

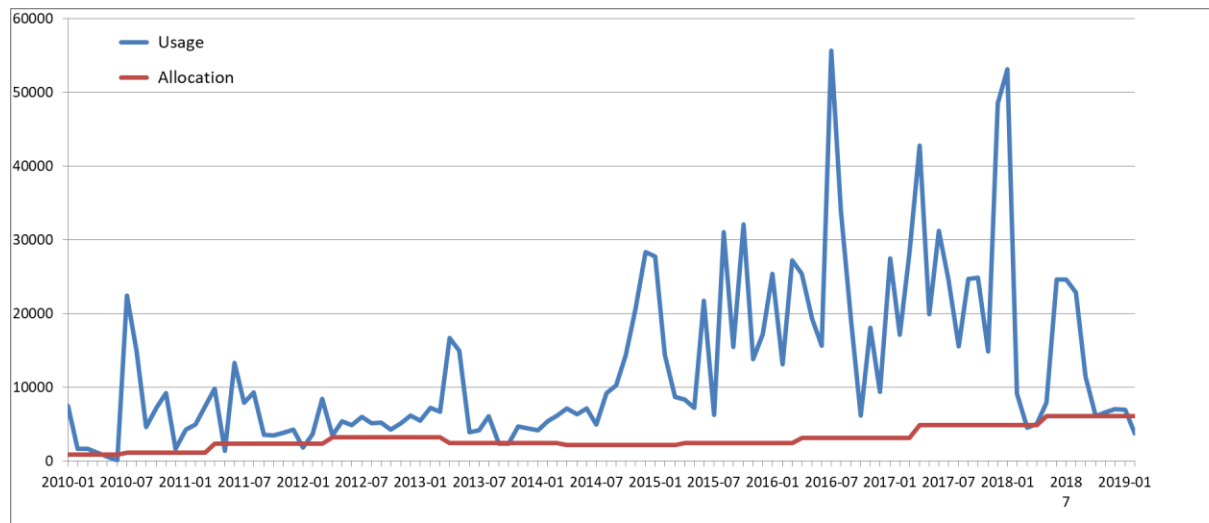
RAL Tier-1 Centre – More on ALICE

- Decreasing UK contribution to ALICE CPU time
 - 1.99% - last 12 months
 - 4.04% - previous 12 mths
 - 5.15% - 2016/17
- ALICE has always got its fair share
- RAL insufficient pledge for 6 months in 2018
- Since Oct'18, larger focus on multi-core jobs for ATLAS, CMS to provide their allocation when averaged over the year

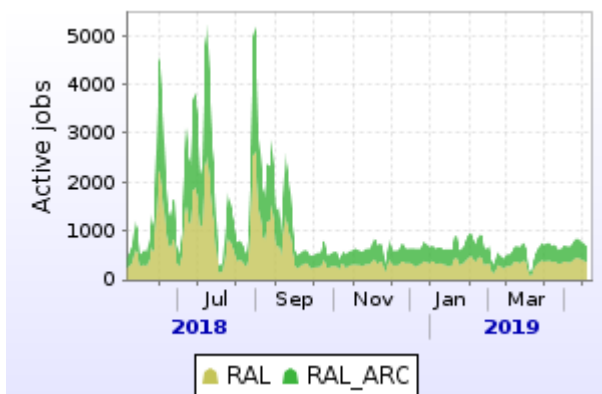


RAL Tier-1 Centre – More on ALICE

- CPU fairshare – 2.6% (7300 HS06) in 2019
- CPU efficiencies – 85-95% for ALICE
- No limit on opportunistic use of spare cycles jobs
- ~24.6K HS06 CPU used in June and July 2018



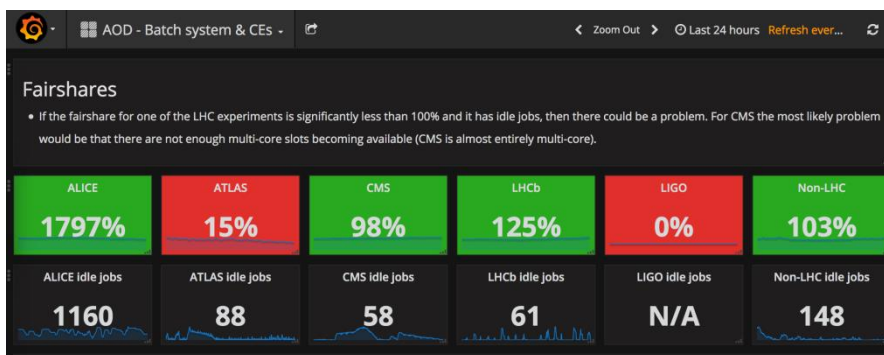
RAL Tier-1 Centre – More on ALICE



Two VOBOXes



Tier-1 Dashboard – ALICE – last 30 days



Fairshares 8 July 2018

Outline

- UK GridPP Collaboration
- Since last report
- Oxford T2 status and plans
- RAL Tier-1 Centre
 - Components status and plans
 - ALICE highlights
- **Storage for ALICE @ RAL**
- UK funding status

ALICE Storage @RAL – CASTOR

- ALICE is the only remaining user of disk only endpoints on CASTOR (550 TB)
 - Hardware over 4 years old
- CASTOR Gen (AliEn SE ALICE::RAL::SE) manages disk-only data for ALICE (428 TB data, 122 TB free) accessed by external clients and Tier-1 batch farm jobs
- Feb'19 – ALICE migrated to new CASTOR WLCG Tape instance (AliEn SE ALICE::RAL::TAPE)
 - Shared buffer (~2 PB) and shared drives (20) offer significantly better throughput

ALICE Storage @RAL – CASTOR

- ALICE data currently on tape – 727 TB
- Migration to CASTOR WLCG Tape required the setup of a separate CASTOR headnode (XRootD redirector) sitting behind the DNS alias *xrootd-alice-tape.gridpp.rl.ac.uk*
- Necessary because the ALICE XRootD token authz cannot coexist with the GSI authn used by other VOs
- No problems/issues so far

ALICE Storage @RAL – ECHO

- Ongoing work to enable ALICE to use the Tier-1 Ceph storage cluster (Echo)
- Full 2019 ALICE pledge (720 TB) available in Echo
 - Not yet usable because the XRootD config on the Ceph gateways needs to be modified (for ALICE token authz)
- RAL test endpoint (AliEn SE ALICE::RAL::CEPH) passes all ALICE external tests except the 3rd party copy (aka TPC)
 - Once solved, also need to ensure ALICE batch jobs run efficiently

What is th

Disk storage elements

AliEn SE		Catalogue statistics					Storage-provided information					Functional tests				Last day add tests		Demotion			
SE Name	AliEn name	Size	Used	Free	Usage	No. of files	Type	Size	Used	Free	Usage	Version	EOS Version	add	get	rm	3rd	Last OK add	Successful	Failed	factor
1. RAL - CEPH	ALICE::RAL::CEPH	953.7 GB	0	953.7 GB	-	0	CEPH	-	-	-	-	Xrootd v4.8.5					Test...	08.05.2019 17:01	24	0	0
Total		953.7 GB	0	953.7 GB		0		0	0	0				1	1	1	0				

ALICE Storage @RAL – ECHO

- We have been in long correspondence with XRootD developers to resolve the TPC issue
 - It seems that the authz token strings are stripped from the CGIs when the '--tpc' flag in xrdcp is switched on
- It has been registered as an issue and hopefully there will be a fix in XRootD 4.10 release

Outline

- UK GridPP Collaboration
- Since last report
- Oxford T2 status and plans
- RAL Tier-1 Centre
 - Components status and plans
 - ALICE highlights
- Storage for ALICE @ RAL
- UK funding status

UK Funding Status

- GridPP6 bid has been submitted
- It will cover April 2020 to March 2024
- **Assuming the bid is fully funded** the Tier-1 will provide 3% of ALICE requirements (up from 2% now)
- UK Tier-2s will provide 2.1% of ALICE requirements
- See REBUS

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



Thank you!
Any questions?