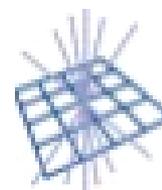




Science & Technology Facilities Council  
Rutherford Appleton Laboratory



**GridPP**  
UK Computing for Particle Physics

# ATLAS Experiment Support

Tim Adye

Rutherford Appleton Laboratory

GridPP 41 Collaboration Meeting  
Ambleside, University of Cumbria

31<sup>st</sup> August 2018



ATLAS

# Outline

1. Introducing myself
2. Outline of the ATLAS Liaison role
3. Some issues / projects I am working on
4. Plans and ongoing issues / projects for ATLAS and the UK Cloud

I have included various links in these slides, as reference in case they are of help – or even just vague interest.

I'm sorry that some links are to internal ATLAS pages, so only of use to ATLAS members.

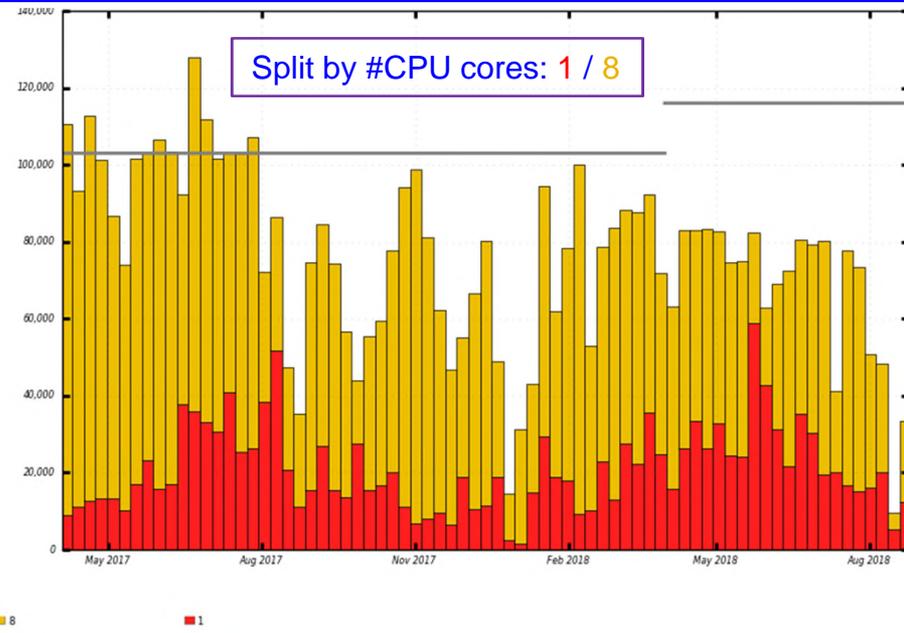
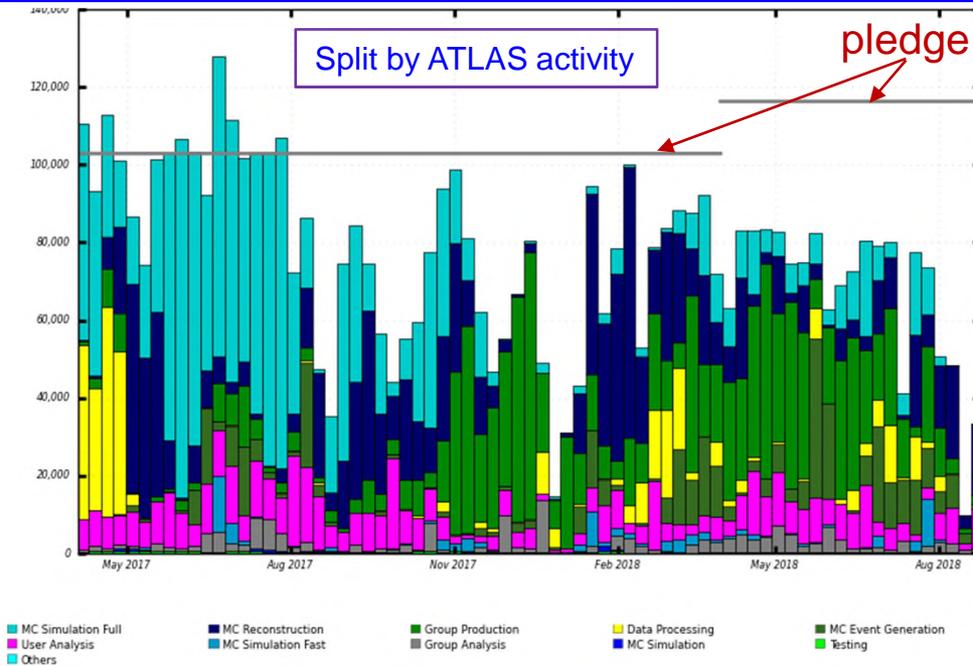
# Introducing myself

- In May, I took over as ATLAS experiment liaison from **Alastair Dewhurst**
  - Alastair had become RAL Tier 1 Manager
- Some previous responsibilities:
  - Managed the BaBar Tier A Centre at RAL (2002–7)
    - similar sort of role to ATLAS liaison, but with earlier technologies (mostly pre-Grid)
      - shared BaBar–RAL liaison work with **Emmanuel Olaiya** (now of CMS)
  - Developed BaBar data management and distribution system
    - in collaboration with **Alessandra Forti** and others
  - I joined ATLAS in 2007, mostly working on physics analysis and performance
    1. Higgs search and measurements (statistical combination)
      - I ran one of the first large-scale ATLAS *analysis* uses of the Grid
        - Determined “ $5\sigma$ ” discovery significance using unprecedentedly-large toy MC simulations
    2. Track reconstruction performance and software validation
    3. ATLAS Phase II Upgrade: performance of the planned new tracking detector, “ITk”
      - Developed ART (ATLAS Release Tester) tests for ITk → explained on later slide
    4. Shift duty supporting ATLAS analysis users on the Grid

# ATLAS experiment liaison

- Ensure that ATLAS gets its work done on RAL Tier-1
  - **Oversee ATLAS jobs and data transfer status and performance**
    - Use ATLAS monitoring to compare with Tier-1 monitoring
  - **Run ATLAS Frontier service at RAL Tier-1**
    - ATLAS Conditions database (Oracle) central servers at CERN, ccIN2P3, TRIUMF, RAL
    - Not used by other VOs (CMS runs Frontier just at CERN)
  - **Liaise with RAL Tier-1**
    - Work in Tier-1 Centre, frequent discussions with staff.
  - **Feed back from/to ATLAS**
    - Collaboration meetings, ADC+other meetings, ATLAS eGroups, weekly+quarterly reports to Roger Jones
- Help “ATLAS UK Cloud Squad” keep the Tier-2s working for ATLAS
  - **Alessandra Forti, Peter Love, Elena Korolkova**
    - aided by GridPP storage group and site admins from ATLAS UK Tier-2s
  - **Chair UK Cloud support weekly meeting**
    - See Indico pages for minutes
- Project work to develop ATLAS distributed computing to better use GridPP services

# ATLAS CPU use of Tier-1

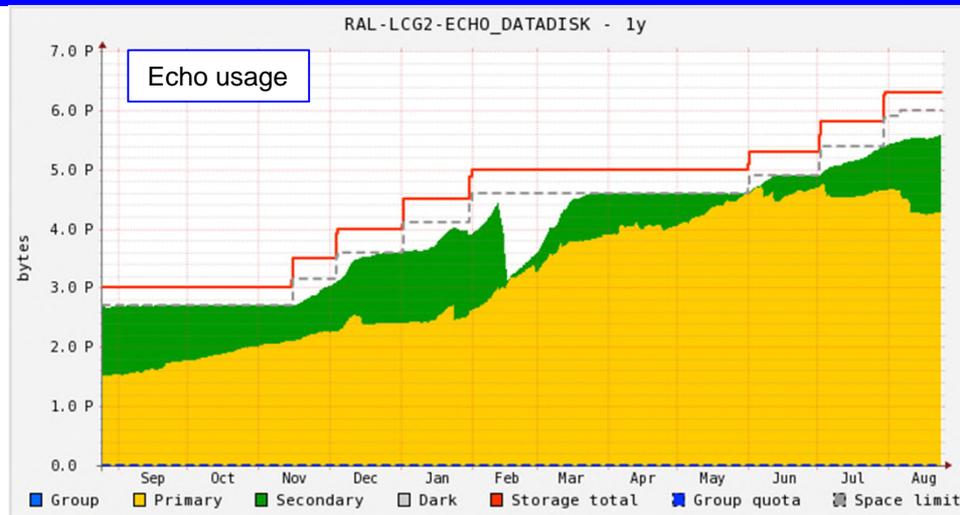
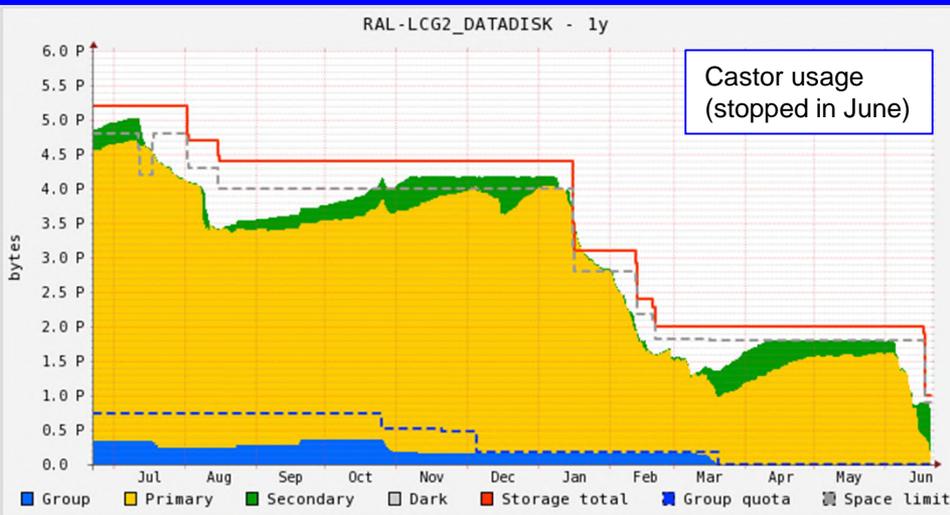


- Since August 2017, ATLAS has used **60–80 kHS06**
  - RAL Tier-1 ATLAS pledge: **103 kHS06** (till Apr18), then **116 kHS06**
- Consider specific reasons for drops in usage:
  - Oct–Dec: possibly caused by incorrect ATLAS PanDA configuration
    - 1 vs 8-core jobs sent to wrong queue
  - CERN Christmas closure: drop in ATLAS work
  - 10–17 Aug 2018: one week Echo downtime
- New capacity not fully deployed until August 2018 → **116% pledge**
  - Previously deployed capacity: 97% (Jun-Oct), 89% (Nov-Jun) of FY17/18 pledge
- These issues still do not fully explain why ATLAS use remained mostly below 80 kHS06

# Monitoring Tier-1 CPU usage

- From mid-August, ATLAS should be able to exceed pledged use of the farm
  - Farm capacity is now above pledged amount
  - Configuration fixes applied → explained yesterday by James Adams
- Monitor closely in next weeks to ensure we achieve this
  - See if need further improvements on ATLAS (PanDA) and/or Tier-1 (Condor) sides
- Tier-1 Capacity Planning Meeting agreed there should be monthly (and quarterly) reviews of CPU usage of each VO
  - Monthly meeting with Tier-1 batch farm experts and experiment liaisons
  - If the usage is below pledge we need to understand why
  - Produce a report for the PMB VO contacts to officially sign off

# Disk data migration from Castor → Echo



- ATLAS has been migrating data off RAL Castor disk over the last year
- ATLAS has been increasingly using Echo since Feb 2016
- 4 July: switched off last PanDA queue accessing Castor disk
- Jul-Aug: transfer remaining primary data off Castor disk
  - Decided to drop ~15k of old logfiles rather than wait for them to be copied
- Next steps:
  1. Should now be ready to confirm that ATLAS doesn't need any remaining files (just dark data)
    - Just waiting for last word from ATLAS DDM group
  2. Remove DDM endpoints from AGIS and Rucio
  3. Remove old PanDA queues. Remainder only access Echo disk and Castor tape.

# ATLAS Tier-1 tape tests for “tape carousel”

- ATLAS investigating the feasibility of a “Tape Carousel” workflow
  - Copy data from tape to disk and process in a controlled way
    - Will allow keeping AOD data on tape
    - Process AOD to create derived datasets using carousel
- First tests just of tape reading performance at various Tier 1 sites
- Tier-1 is replacing Castor tape stager (see Rob Appleyard’s talk yesterday)
  - Castor disk service is being decommissioned
- Plan to combine ATLAS performance test for RAL with Tier-1 test of new hardware
  - Transfer 204 TB of AOD data from tape → Echo, controlled by Rucio
  - New Castor tape stager is up and running, but not yet in production
  - Has same tape backend, so need to set up a dummy Rucio SE and register data as if it were newly written to tape
    - I’m writing a Rucio tool to do the registration without initiating file transfer
  - Can then initiate Rucio replication request from test SE to Echo
  - Plan to start test next week

# Oracle support at the RAL Tier-1

- Oracle currently used at the Tier-1 for
  1. Castor disk service – being decommissioned now
  2. Castor tape service – to be replaced in next 2 years
  3. ATLAS Frontier conditions database mirror at RAL
- Oracle is expensive to run
  - May no longer be covered by CERN license
  - Not been able to hire an Oracle database administrator, and use of contractors is expensive
- RAL Tier-1 has decided to phase out Frontier support on Oracle
  - Plan for sometime after July 2019, to allow the completion of Run2 reprocessing during LS2
  - The Tier-1 will continue to provide high performance/resilient database services
    - Looking at hiring a new non-Oracle database administrator shortly
- Implications for ATLAS conditions database service
  - Frontier requires Oracle
    - Some MySQL support, but cannot mirror from Oracle database at CERN
    - If continue to use Frontier, then will have to drop RAL mirror
      - Rely on 3 other sites, or even reduce to CERN-only (as for CMS)
  - ATLAS investigating alternative conditions databases for Run3

# ATLAS plans

- ATLAS UK Cloud Squad is closely involved in wider ATLAS developments
  - eg. Alessandra Forti is ATLAS Infrastructure Coordinator
  - See following slides (many kindly provided by Alessandra)
  
- IPv6
  - Alastair Dewhurst will continue to follow IPv6 developments in ATLAS and WLCG groups

# CentOS7 and Python upgrades

- ATLAS will start to actively encourage sites to migrate SL6 → CentOS7 in the autumn
  - Not be mandatory until 2019, but SL6 is starting to cause problems for some new developments due to the default version of Python
- SL6 releases can run on CentOS7 but native CentOS7 releases cannot run on SL6
  - This is affecting only nightly test releases at the moment but it will soon cause problems for production
- More mandatory deadlines will be set at the beginning of 2019
  - 8 UK sites already have CentOS7 nodes in production
- Information on how to upgrade is on [ATLAS CentOS7Readiness TWiki](#)
  
- Python 2.6 → 2.7 → 3
  - OS defaults: SL6=Python 2.6, CentOS7=Python 2.7
  - ATLAS production/analysis software uses its own version of Python in CVMFS
    - mostly decoupled from OS
  - except: Pilot supports Python 2.6 (for now) and 2.7
    - Python 3 version in preparation.

# Containers

- Development is on-going, tied to Pilot 2 development and deployment
- So far most of the effort is concentrated on production queues
  - Manchester, Glasgow, Lancaster, Imperial, RAL all have a test queue
- Also starting work on integrating user containers and containers for software preservation
  - Potential for standalone containers to solve many problems for analysis users and groups
  - Requires more development to be able to run them on the Grid – [draft plan](#)
- More info for sites on [ADCContainersDeployment TWiki](#)

# Non-GridFTP third party copy

- At CHEP a new WLCG WG was created to setup the infrastructure and test WebDAV (https) and XRootD as replacements for GridFTP
  - Effort led by ATLAS and CMS
  - Plan is to deploy at least one alternative at all sites by the end of 2019
  - UK heavily involved for both experiments
    - RAL, IC, Glasgow, and QMUL already listed as test sites. Manchester will follow soon.
  - More information on [LCG ThirdPartyCopy Twiki](#)
  
- We are interested in WebDAV at RAL also because
  1. We would like to integrate S3 endpoints seamlessly into Rucio
  2. RAL provides S3 storage and runs a test Rucio instance for SKA
    - and in the near future for DUNE as well

# WebDAV testing

- I made a quick test to see what WebDAV services were known to ATLAS
  - Sites and storage endpoints are known to have varying levels of WebDAV support
    - Passive only, third-party push or pull, or full support
- Test Setup: pick all sites listed in AGIS (ATLAS Grid Information System) with:
  - Production “davs” SCRATCHDISK endpoint and HammerCloud datasets availability
  - `fts-transfer-submit -s https://fts3-test.gridpp.rl.ac.uk:8446 -t SCRATCHDISK -f xfer40.txt`
- Initial Results
  - 40 sites match criteria
  - For 23 sites, FTS transfers worked first time!
    - BUT – most via streaming, not third-party copy
    - Very low statistics (and only small files tested) but transfer rates comparable to GridFTP
      - See backup slide for numbers
  - AGLT2 gets a **gold☀star** as most transfers via it were third-party
    - Only site tested that is currently able to mediate third-party transfer via push and pull

# Third party WebDav transfers

- Summarise results according to storage system:

		Source			
		dCache	DPM	StoRM	S3
Destination	dCache	Green	Green	Green	Green
	DPM	Green	Green	Orange/Green	Green
	StoRM	Green	Orange/Green	Orange/White	Orange/White
	S3	Green	Green	Orange/White	Orange/White

- Have not tested EOS
- Green** means a successful third party transfer, mediated by FTS, has been performed between the specified storage
- Orange/Green** means a successful streaming transfer has been performed, however with the correct configuration third party transfers should work
- Orange/White** – untested but we believe only streaming transfers are currently possible

# New(ish) developments in ATLAS workflows

- Event Service: processes a few events at a time
  - Pre-emptable, minimizing loss from failures.
  - Has been used by all UK sites. Particularly helpful to keep running just before downtime.
- Harvester: tighter integration of PanDA into batch system jobs
  - Dynamic allocation of work to jobs, depending on available local resources
  - Plan to roll it out for ATLAS UK Cloud week of 10-14 September
    - First step is to merge single and multi-core PanDA queues (“UCORE”)

# ADC project work

- I recently developed ART (ATLAS Release Tester) tests for the ITk (tracking detector upgrade)
  - ART is a new system developed by **Tulay Cuhadar Donszelmann** (Sheffield) for automated physics performance and software validation on the Grid
    - Replaces previous system (RTT), which ran on CERN batch farm
    - Runs test jobs on each nightly release
    - Returns a status indicating that the job ran OK, and produced expected results
    - Status flags for each night, release branch, test package, and test sample all presented within the PanDA monitor display
  - For the ITk (and other subsystems) we also required histogram comparisons (DCube) available for checking by expert shifters
    - Tulay kindly implemented the system I proposed for downloading the results to EOS disk for web display
- In the longer term, I am keen to renew my interest in Distributed Data Management
  - Perhaps get involved in Rucio development

# Summary

- ATLAS experiment liaison:
  - Ensure that ATLAS gets its work done on RAL Tier-1
  - Help “ATLAS UK Cloud Squad” keep the Tier-2s working for ATLAS
  - Project work to develop ATLAS distributed computing
- First look at
  - ATLAS CPU use of Tier-1
  - Disk migration from Castor → Echo
  - ATLAS Tier-1 tape tests for “tape carousel”
- Plans
  - Phase out of Oracle support at RAL Tier-1
  - SL6 → CentOS7 and Python 2.6 → 2.7 → 3
  - Containers
  - WebDAV and/or XRootD as replacements for GridFTP
  - Changing ATLAS workflows
  - Projects: ART and Rucio development

**Backup**

# WebDAV tests – some more detail

- Very basic test of WebDAV access
  - Just to see what's trivially accessible
    - Not yet tried to get other sites working
    - Missing tests probably my fault
- Pick all sites listed in AGIS with:
  - “davs” endpoint access to
    - HammerCloud dataset and
    - SCRATCHDISK
  - “production” status
    - 40 sites, 23 FTS transfers work first time
- fts-transfer-submit -s https://fts3-test.gridpp.rl.ac.uk:8446 -t SCRATCHDISK -f xfer40.txt
- Test: UKI-LT2-QMUL ↔ all others (both ways)
  - 3rd party transfer only used for AGLT2
    - Uses streaming via FTS controller for all others
- Test: AGLT2 ↔ all others (both ways)
  - Uses 3rd party transfer with most sites
    - marked black
  - Streaming still used for a few sites
    - marked red

Push MB/s	Pull MB/s	Store	Peer site
0.51	0.33		AGLT2
0.40		dpm	BEIJING-LCG2
0.36			CA-SFU-T2
0.03		dCache	CA-VICTORIA-WESTGRID-T2
0.49	0.39	dCache	CSCS-LCG2
0.26	0.03	dCache	DESY-ZN
0.11	0.28		EELA-UTFSM
0.75	0.38	dpm	FMPHI-UNIBA
0.43		dCache	GoeGrid
0.54	0.32	dCache	IEPSAS-Kosice
0.38		dpm	IN2P3-LAPP
0.55		dpm	INFN-COSENZA
0.36		dpm	INFN-NAPOLI-ATLAS
0.30		dpm	INFN-ROMA1
0.35	0.25	dpm	RO-02-NIPNE
0.58	0.03		RRC-KI-T1
	0.45	dpm	TOKYO-LCG2
	0.26	dpm	TR-10-ULAKBIM
0.50	0.33		UKI-LT2-QMUL
0.37		dpm	UKI-LT2-RHUL
0.35		dpm	UKI-NORTHGRID-LANCS-HEP
0.38		dpm	UKI-SOUTHGRID-CAM-HEP
0.45	0.27		WEIZMANN-LCG2
0.61	0.37	dCache	wuppertalprod