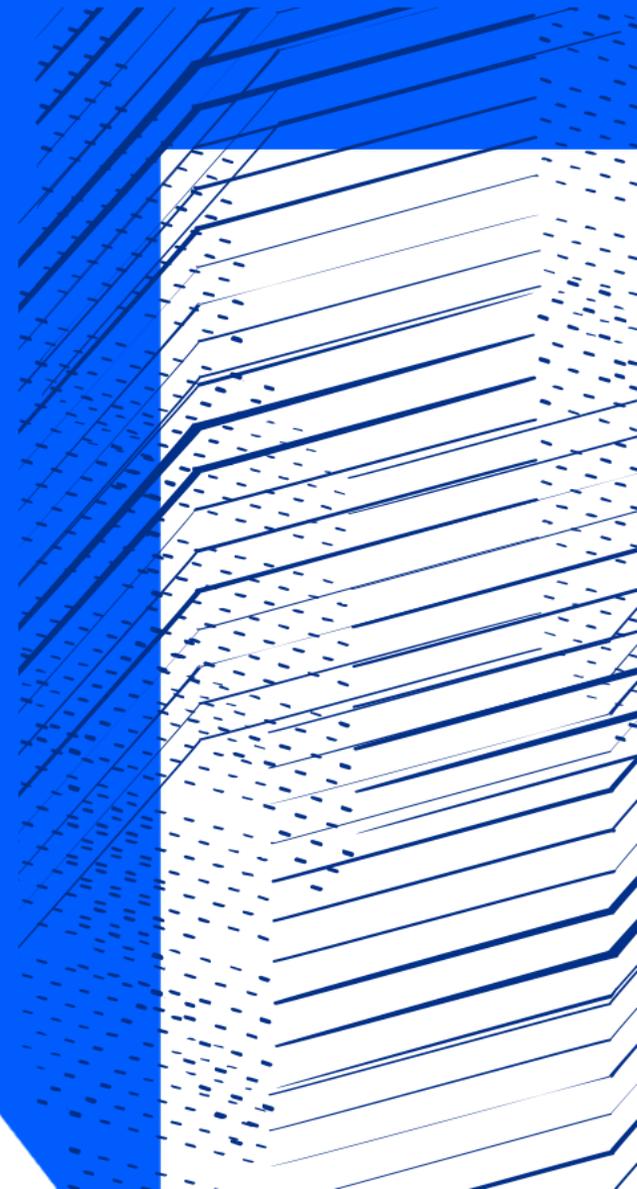




Science and  
Technology  
Facilities Council

# Data Management

Alastair Dewhurst, on behalf of many.



# Introduction

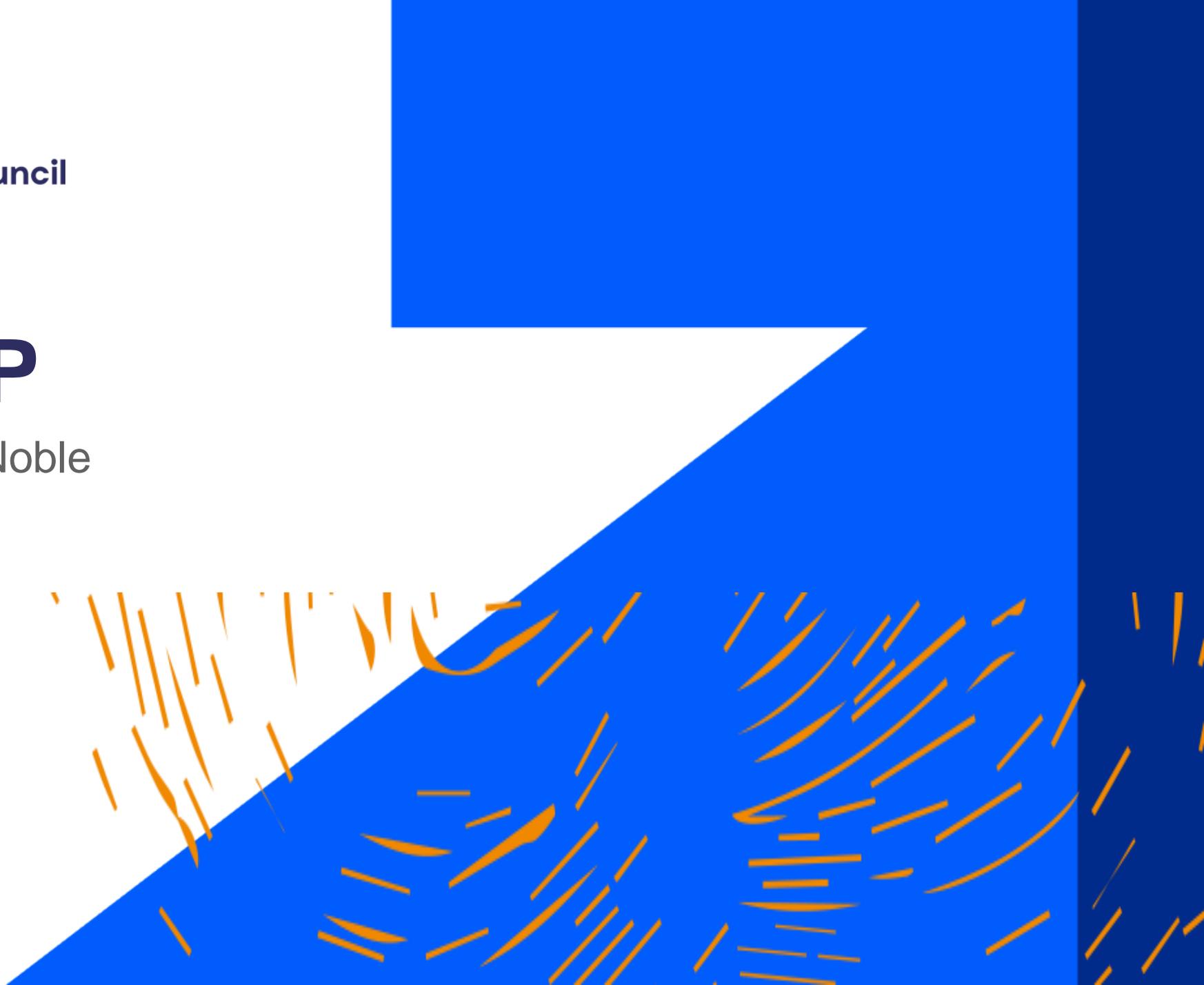
- Going to cover 3 main topics:
- Swift-HEP
  - Work has been done by Tim Noble (who is not available today).
  - Covers the work ongoing to meet objectives.
- ExCALIBUR 1a and data centre evolution
  - What we learnt from ExCALIBUR 1a and how this is feeding into our data centre evolution.
- Quick summary of data management work going on in the rest of the WLCG and beyond.



Science and  
Technology  
Facilities Council

# Swift-HEP

On behalf of Tim Noble



# SWIFT-HEP Objectives

- Build a UK data-lake using Rucio to move data management to an infrastructure level and reduce overall effort required by experiments and infrastructure.
  - Configure at least 4 core sites with permanent storage
  - Configure additional sites including cloud, cache storage, and no local storage)
  - Create metrics to compare current usage to data-lake
- Implement additional QoS information in data management (13-18 months)
  - Rucio development for supporting data intensive workflows
  - Prioritising data to relevant storage reliability based on age

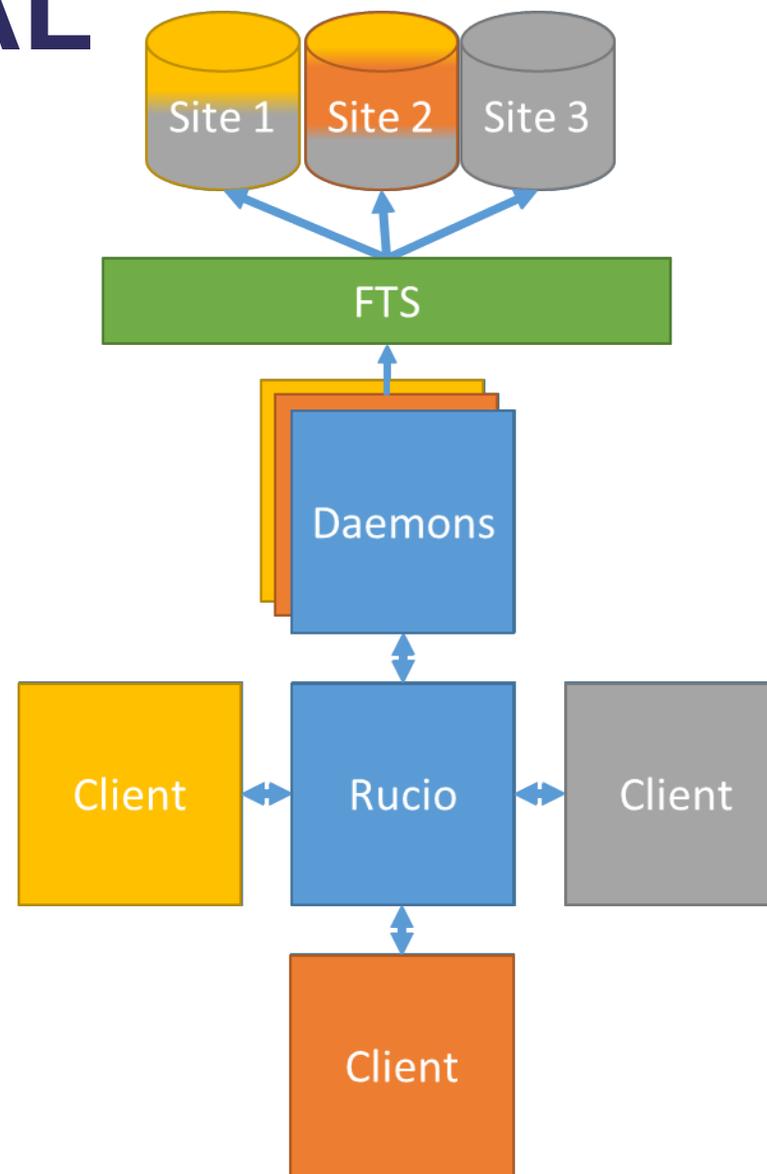
# Rucio

- Intelligent data movement
- High Capacity
- Heterogeneous network and storage infrastructure
- Adaptive replication
- Data recovery
- Seamless FTS integration
- Funding from SWIFT-HEP and EGI-ACE



# Multi-VO Rucio at RAL

- VMs on SCD OpenStack cloud
  - VM for access (Bastion)
  - Server
  - ELK
  - 3 daemon container sets – one set per VO
- 2 machines running the databases

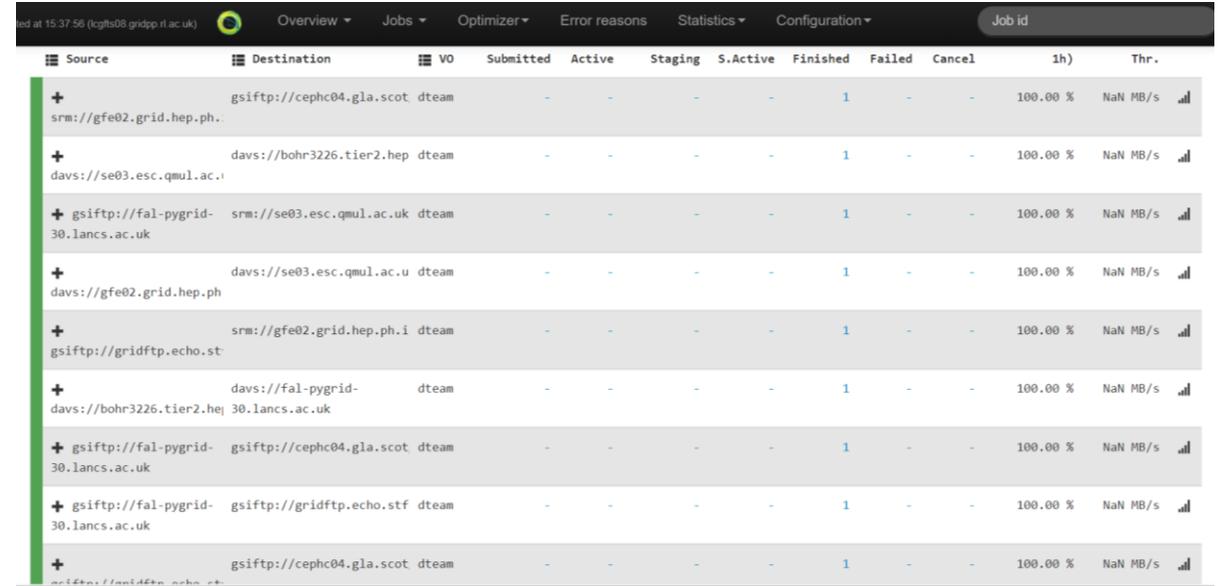


# Advantages of Multi-VO Rucio

- Running a Rucio instance that supports multiple VOs:
  - Maintained by RAL not by individual experiments
  - One instance to support and maintain
  - Lower levels of load
  - Shared RSE configuration
  - New VOs are quick to add – work with VO admin to setup their environment
- More contact with Developers and larger communities using Rucio to know how best to utilise

# Progress - UK data-lake prototype

- 6 permanent storage sites configured
  - Hourly upload to each
  - Then transfer those files to all other sites
- Monitoring currently using Rucio and FTS to monitor transfer status

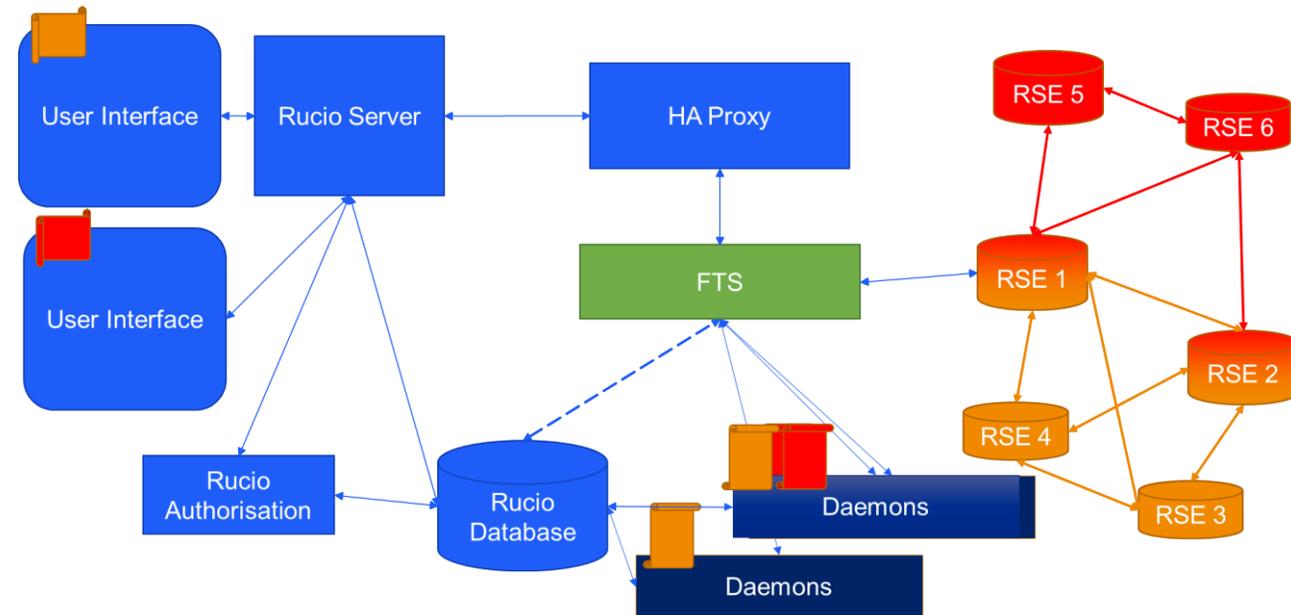


The screenshot shows a monitoring dashboard with a table of transfer jobs. The table has columns for Source, Destination, VO, Submitted, Active, Staging, S.Active, Finished, Failed, Cancel, 1h, and Thr. The jobs listed are:

Source	Destination	VO	Submitted	Active	Staging	S.Active	Finished	Failed	Cancel	1h	Thr.
+	gsiftp://cephc04.gla.scot	dteam	-	-	-	-	1	-	-	100.00 %	NaN MB/s
+	davs://bohr3226.tier2.hep	dteam	-	-	-	-	1	-	-	100.00 %	NaN MB/s
+	gsiftp://fal-pygrid-30.lancs.ac.uk	srms://se03.esc.qmul.ac.uk	dteam	-	-	-	1	-	-	100.00 %	NaN MB/s
+	davs://se03.esc.qmul.ac.u	dteam	-	-	-	-	1	-	-	100.00 %	NaN MB/s
+	srms://gfe02.grid.hep.ph.i	dteam	-	-	-	-	1	-	-	100.00 %	NaN MB/s
+	davs://fal-pygrid-30.lancs.ac.uk	dteam	-	-	-	-	1	-	-	100.00 %	NaN MB/s
+	gsiftp://fal-pygrid-30.lancs.ac.uk	gsiftp://cephc04.gla.scot	dteam	-	-	-	1	-	-	100.00 %	NaN MB/s
+	gsiftp://fal-pygrid-30.lancs.ac.uk	gsiftp://gridftp.echo.stf	dteam	-	-	-	1	-	-	100.00 %	NaN MB/s
+	gsiftp://cephc04.gla.scot	dteam	-	-	-	-	1	-	-	100.00 %	NaN MB/s

# Progress - Development for daemons

- Building on top of previous Multi-VO developments
- Developing a selection of certificate based on the VO submitting the job, but also allow for bulk submissions for scale up of the Multi-VO use.
- Continue to work with Rucio developers to improve Rucio.



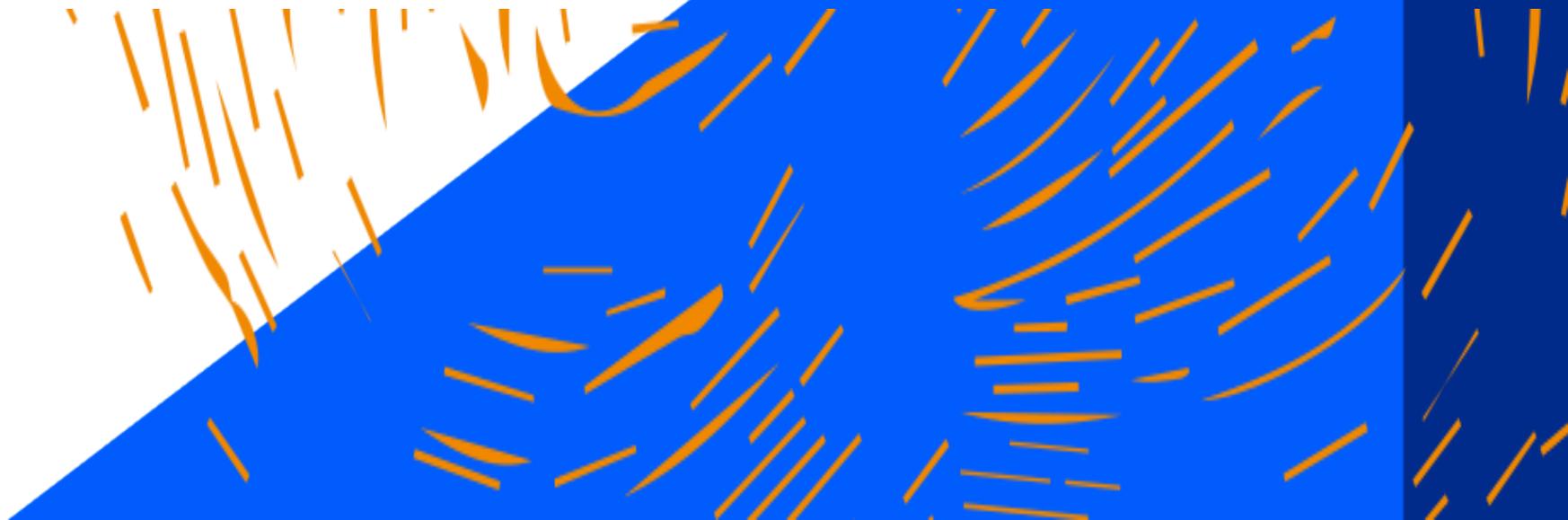
# Future plans

- SWIFT-HEP Developments
  - Increase number of sites and storage technologies
  - Improve monitoring of Rucio and resources used for analysis
  - QoS Developments for Rucio
- EGI-ACE Developments
  - Improvements to usability and accessibility for new users
- Internal Developments
  - General improvements to service
    - Containerisation
    - Gain experience with on boarding a variety of communities and requirements



Science and  
Technology  
Facilities Council

# ExCALIBUR 1a and Data Centre Evolution

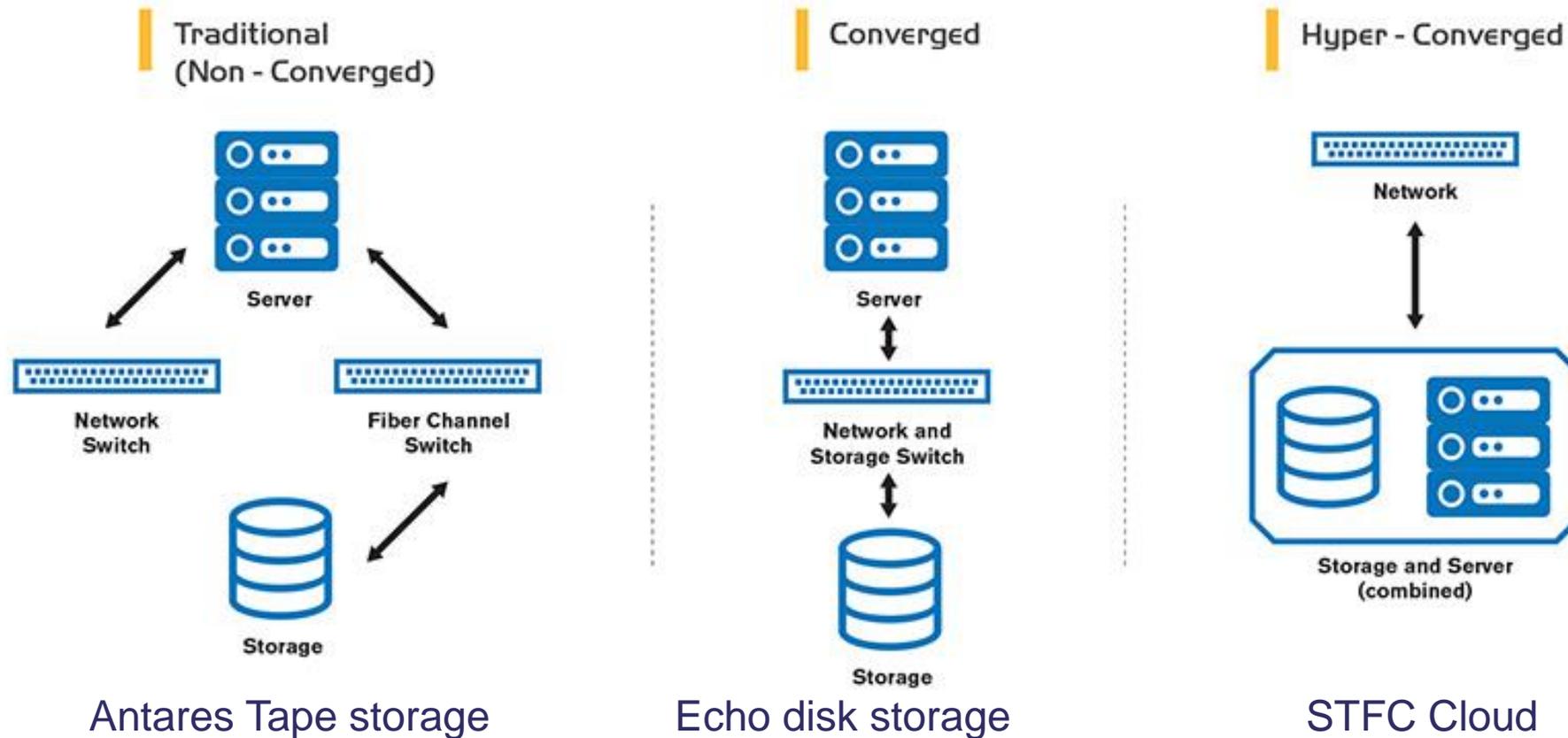


# ExCALIBUR 1a

- 6 months of effort to investigate using Kubernetes to manage data transfer services.
  - Work carried out by Tom Birkett.
- Kubernetes is a container orchestration system for automating computer application deployment, scaling and management.
- It is traditionally much harder to quickly scale data services compared to those that are CPU bound as you need to ensure that the storage and network also scale.

# Data Centre Architecture

- Power of Kubernetes increases as

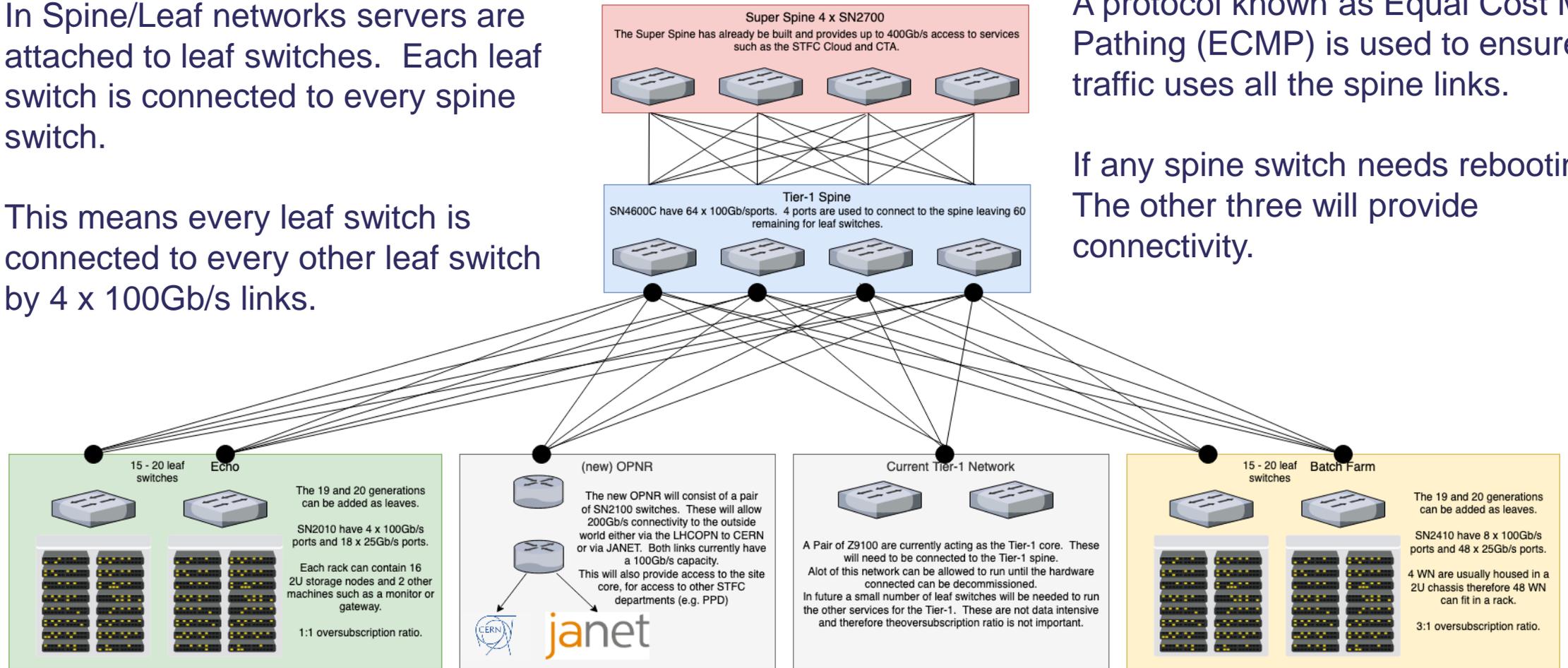


# Network Evolution

In Spine/Leaf networks servers are attached to leaf switches. Each leaf switch is connected to every spine switch.

This means every leaf switch is connected to every other leaf switch by 4 x 100Gb/s links.

## Tier-1 Network Architecture



A protocol known as Equal Cost Multi Pathing (ECMP) is used to ensure traffic uses all the spine links.

If any spine switch needs rebooting, The other three will provide connectivity.

# Storage Evolution

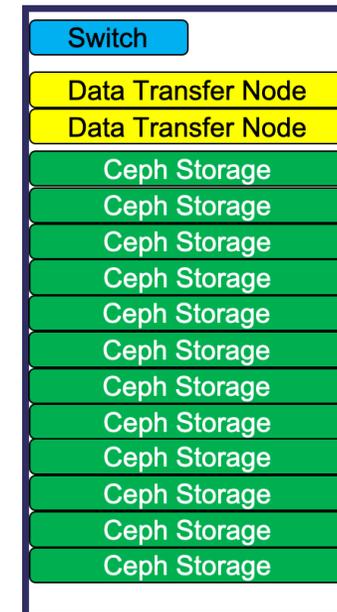
## ARIDED

- Funded by IRIS
- Designed to be a high performance file system mounted across STFC Cloud
- CephFS cluster with all SSD hardware - 1.5PB raw capacity.
- Hardware ordered in August.
- Delivery expected in December.

## Echo

- Existing Object storage for LHC experiments.
- Increasing to 96PB raw capacity.

New procurements have a more converged architecture.



# Plans

- Work is ongoing to containerize our applications:
  - FTS, Rucio, XCaches etc
- Data Centre is continuing to evolve towards a more converge infrastructure.
  - This allows us to continue to scale out available resources.
  - Easier to provide access to a range of resources (e.g. FPGAs)
  - This will allow us to better take advantage of software defined solutions (e.g. Kubernetes) for managing services.
- We are part of an ExCALIBUR 1b bid (ExaTEPP) with lattice QCD group.
  - Rucio developments relevant to Lattice QCD community.
  - Work on Parallelizing data transfers to improve throughput.



Science and  
Technology  
Facilities Council

# Data Management in GridPP and beyond.



# WLCG DOMA

- WLCG DOMA group is coordinating
  - [Document lists current ongoing development.](#)
- UK is involved in several such as:
  - Opportunistic Storage demonstrator
  - XCache
  - Protocol Migration
  - Data Challenges
  - Network Packet marking

# Opportunistic Storage at ECDF



Have “Unstable” storage



Registered it as “Volatile” storage with ATLAS



Use-case: When/if a file replica at Edinburgh becomes damaged ATLAS will failover to find a good replica at another site



Put it into production with no redundancy, expected everything to break



2-3 months later... Everything is still working as per day 1.

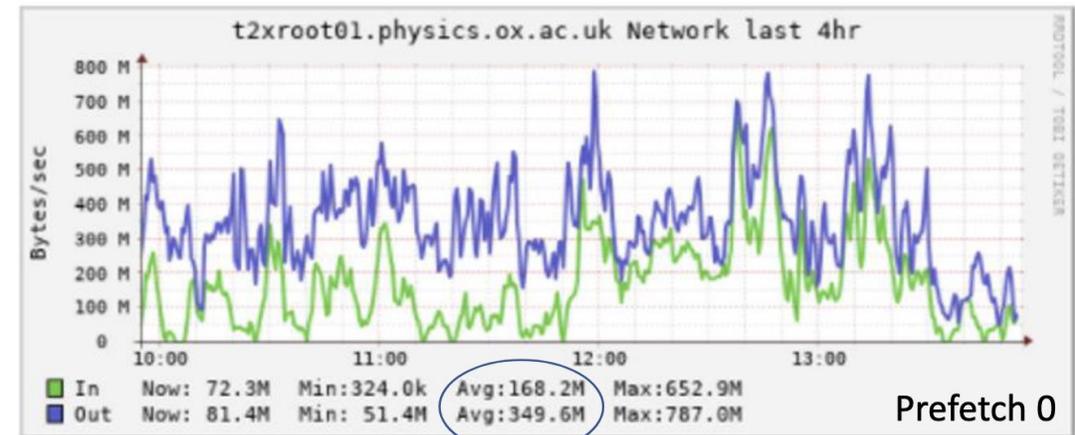


Slightly unexpected, and surprising. Gives us an opportunity to be intentionally a bad actor.

<https://indico.cern.ch/event/1054156/timetable/#20210903.detailed>

# XCache storage at Oxford

- Oxford is decommissioning its disk storage.
- Testing data access from RAL both directly and using an XCache.
- Xcache hopes to:
  - Improve CPU efficiency
  - Reduce failures
  - Reduce the amount of external bandwidth being used.

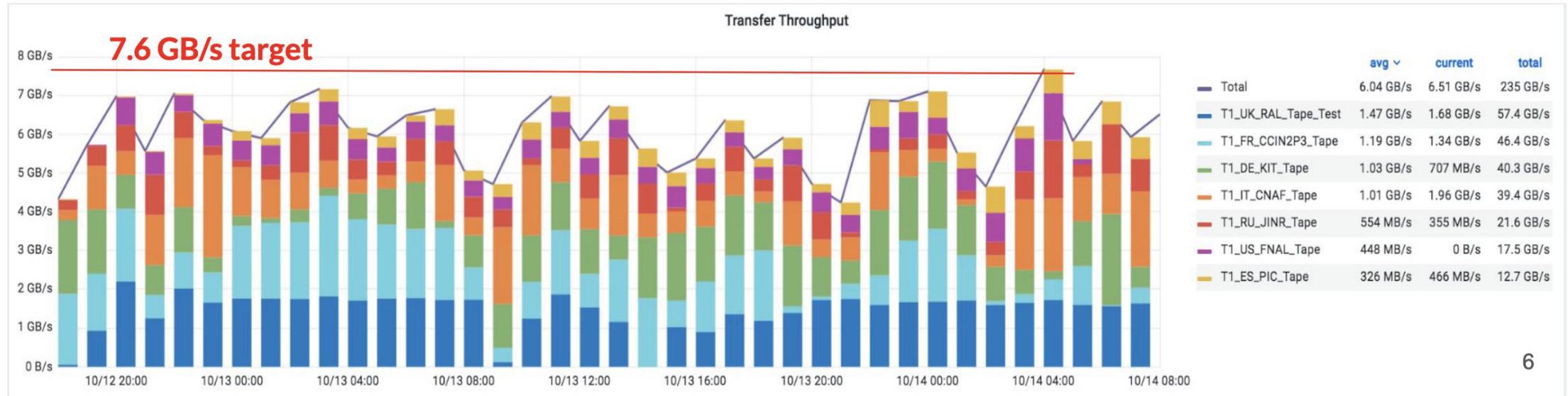


~50% reduction in traffic (whole file caching)

# Tape challenge

- At the start of October the WLCG ran a tape challenge to ensure sufficient performance for start of Run 3.

VO	Throughput (GB/s)
ALICE	0.08
ATLAS	1.6
CMS	0.9
LHCb	1.46
Total	4.04



<https://indico.cern.ch/event/1089983/>

# New Communities

- Rucio is also funded by EGI-ACE to on-boarding of new communities
  - Two new communities using Rucio
  - Tutorials, webinars, and documentation for new users
- LSST has also started testing Rucio.
  - 9PB of storage planned at RAL.

# Conclusion

- ExCALIBUR 1a allowed us to look at potential ways to manage our infrastructure more effectively.
  - This has contributed to the way we evolve our data centre.
- Swift-HEP objectives for the first year have already been (mostly) met.
- UK is effectively contributing to the R&D work being coordinated by the WLCG DOMA group.
- We are continuing to generate interest from new communities.



Science and  
Technology  
Facilities Council

# Questions?