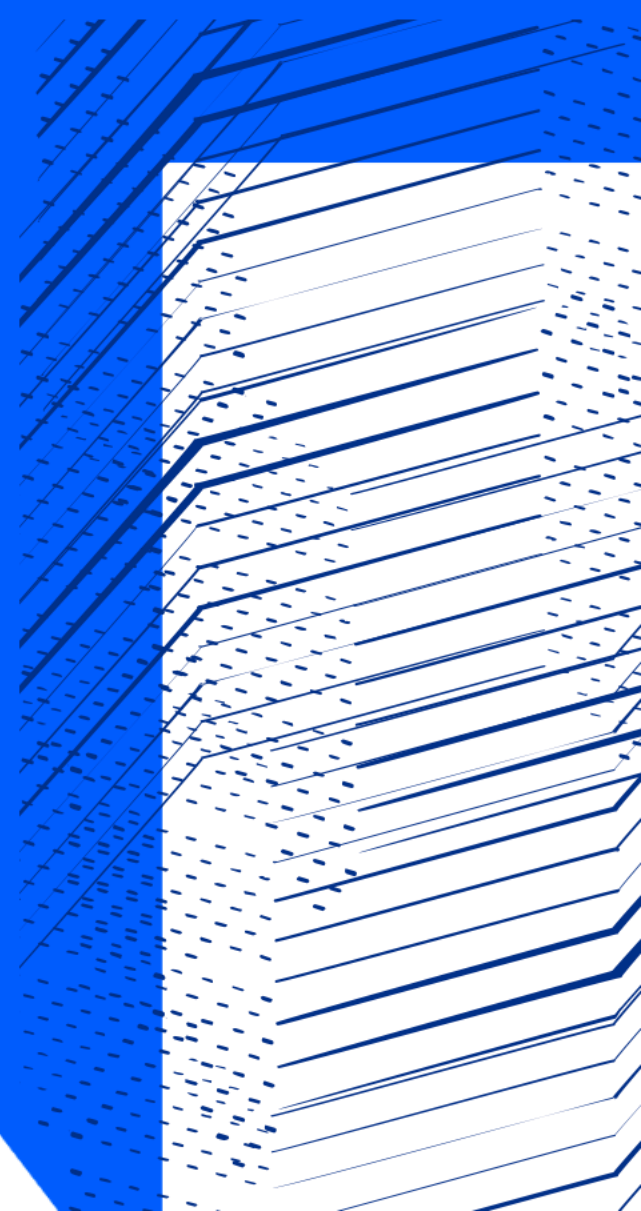




Science and
Technology
Facilities Council

Datalakes: Tier-1

Alastair Dewhurst



Introduction

- Thank you to Dave for introducing the Datalake Vision.
 - My talk follows on and will describe what the Tier-1 plans are.
 - We should consider LHC and non-LHC VOs separately.
- For the LHC VOs:
 - We will just be a part of their Datalakes, providing different QoS storage.
 - Maximizing what we provide for a fixed cost.
- For the non-LHC communities:
 - GridPP will provide the Datalake.
 - Built on Rucio (at RAL) and DIRAC (at Imperial)
 - Improving the service we deliver to them.



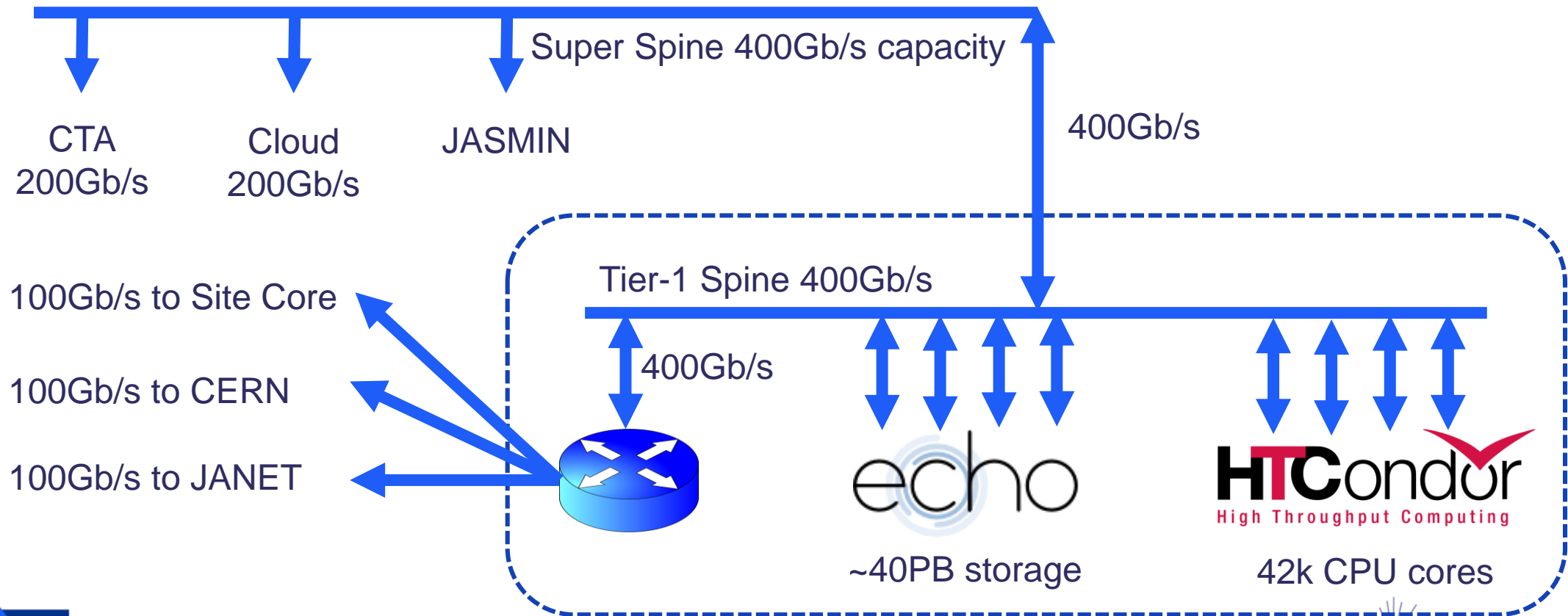
Science and
Technology
Facilities Council

Tier-1 as part of an LHC Datalake



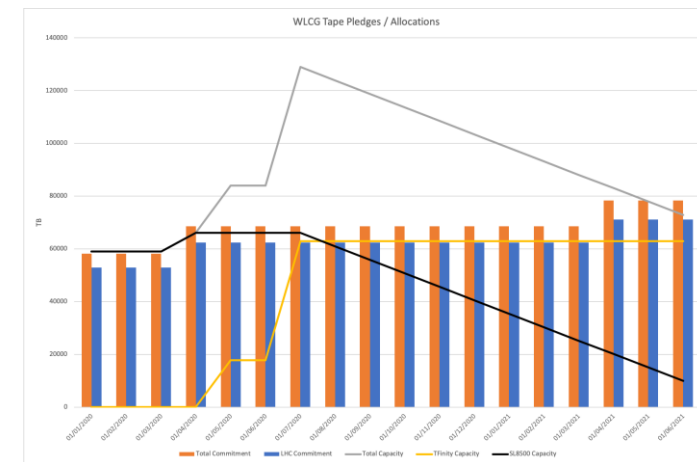
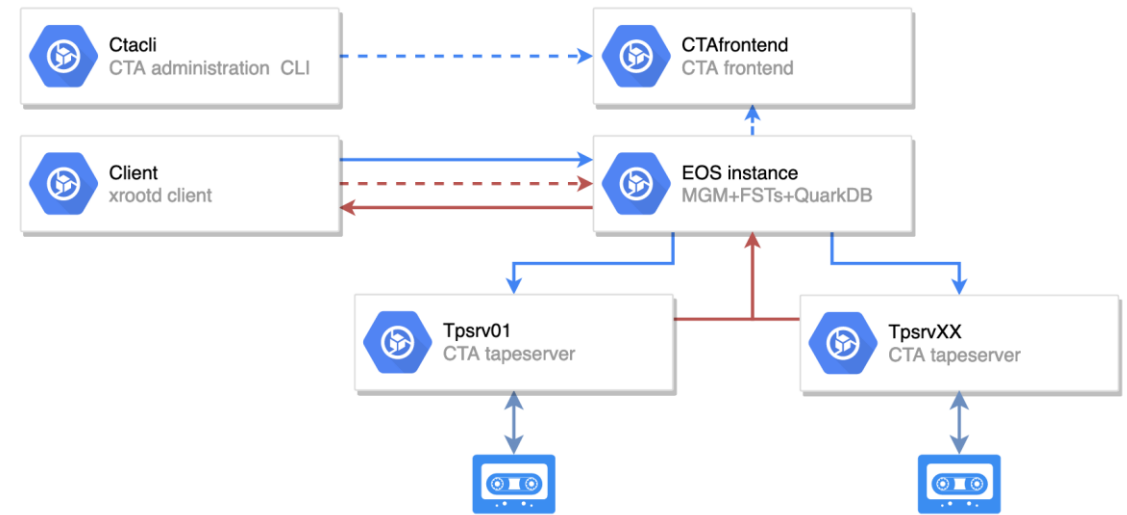
Core network infrastructure

- The Tier-1 is upgrading its network to ensure excellent connectivity to CERN, JANET and other STFC Services



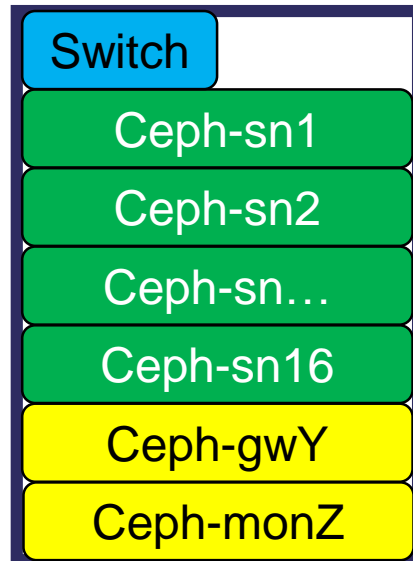
Tape - CTA migration

- RAL decided to migrate to CTA in October 2019.
 - Some delays due to COVID-19, but still hope it will be ready by start of Run 3.
- CTA is a pure tape system
 - “Traditional” Disk Cache will be in Echo.
 - Operating Tape drives at full speed efficiently requires SSD Cache.
 - CERN CTA performs very well in ATLAS Data Carousel.
- Migration to TFinity Tape Robot is going very well.
 - ~1.3PB migrated each week.
 - Aim to not renew Oracle maintenance costs in April 2021.



Disk - Echo evolution

- Ceph can easily support different QoS.
 - EC or Replication
 - HDD or SSD/NVMe
- I have got some SCD funding for a Ceph testbench and ¼ FTE for R&D.



- Trying to move Echo to a more converged infrastructure.
 - Easier to scale up.
- In future a mix of HDD and SSD on each storage node?
 - Non-EC SSD pool for data intensive workflows with random access.

XRootD and S3/Swift

- Echo can be accessed via XRootD or S3/Swift API.
- For at least the duration of Run 3 the LHC experiments will use XRootD to access their data for jobs.
 - Expect an increase in direct I/O and remote reading.
- CTA is built on XRootD.
- XRootD has been the Achilles heel of Echo so far.
 - Ian Johnson has been leading the group making improvements for Echo (more in Sam's talk).
- Significant interest in S3 API from IRIS communities (more in Ian's talk).

Compute - Utilizing SCD Cloud

- The SCD OpenStack cloud has grown massively in the last two years.
 - Now have 60k CPU cores, 300+ GPUs.
 - Significant effort that collaborates strongly with Tier-1.
- Tier-1 batch can spin up VMs on Cloud.
 - Ideal way of testing different types of compute.
- I secured SCD funding for 6 FPGAs which will be added to the Cloud.
 - Will be utilized by Switft-HEP project.



Science and
Technology
Facilities Council

UK Datalake for Non-LHC Communities



Multi-VO Rucio

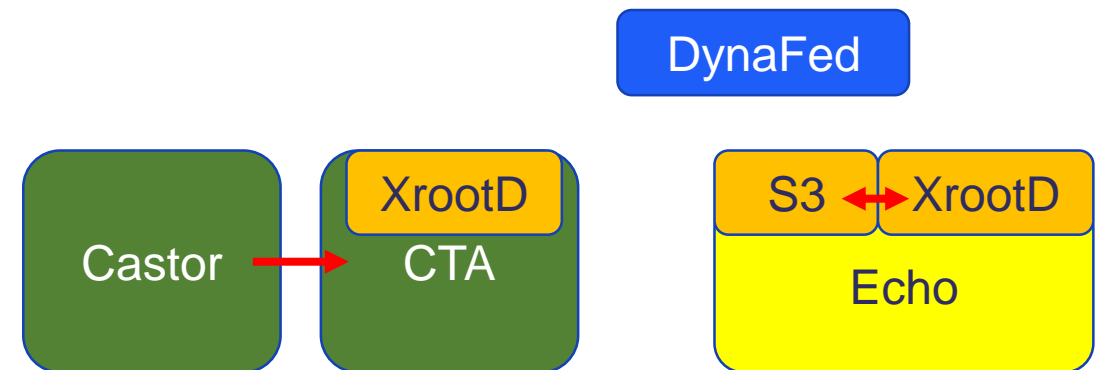
- Rucio is the “brains” of any Datalake.
- Ian Johnson has been leading Multi-VO Rucio project.
 - More details in his talk.
- Multi-VO Rucio aims to allow us to:
 - Support many VOs without requiring vast amounts of effort.
 - Rapidly add new VOs.
 - Share site configuration details.
- Aim to move Rucio to a Tier-1 production service 1st half of 2021.
- With EGI and Swift-HEP money it will be possible to fund an entire additional person working on Rucio.
 - Job advert is being worked on currently

Swift-HEP

- Swift-HEP will provide some effort to actively contribute to the DOMA Datalake work.
- Deliverables:
 1. Setup a UK data-lake prototype with [at least] four sites and compare metrics with the current system (Month 1-12)
 2. Implement additional Quality of Service information in the data management (Month 13 - 18)
 3. Produce recommendations to sites on how to optimise data access and deploy stateless storage (Months 19-24)
 4. Creation of a [Virtual] Analysis Facility (Months 19-30)

Tier-1 Data Services

- Tier-1 will provide Rucio, FTS and DynaFed services.
- Rucio will integrate with DIRAC at Imperial.
 - More info in Daniela's talk.
- Underlying storage can be hidden from users.
 - Reduce impact of Castor to CTA migration.
- DynaFed will simplify access to S3 endpoints.



What does this mean?

- If VOs want Compute and Storage, GridPP will recommend they use the DIRAC service.
 - Very little should change.
- If VOs want just Storage, GridPP will recommend Rucio.
 - They will need to learn Rucio commands but easier than SRM/XRootD.
 - Easier for Tier-1 Liaisons to provide wider support.
- From a site point of view.
 - Rucio has much more flexibility than DIRAC regarding storage endpoint configuration.
 - Site config changes can be rolled out to all VOs easily.
 - Easier to respond to problems:
 - e.g. for LSST data challenge we could replicated data to all sites.
 - Reduce impact of data loss/availability

Summary

- The Tier-1 has plans to evolve, Network, Tape, Disk and Compute services to meet the challenges of Run 3.
- The Tier-1 is well placed to test the different QoS that might be required in a data lake.
- The Tier-1 has finished development of a multi-VO Rucio instance.
 - Funding has been secured for future developments to contribute to the DataLake work.



Science and
Technology
Facilities Council

Questions?