Science & Technology Facilities Council

UK Research and Innovation

Developments in Disk and Tape Storage at the RAL Tier 1

Speaker: Rob Appleyard

Storage at RAL: An Overview

- CASTOR for WLCG tape storage
  - Re-implemented service as a dedicated tape endpoint
- Ceph-based Echo storage system for disk
  - Migration from CASTOR nearly complete
- Ceph-backed OpenStack private cloud
- CASTOR for local facilities tape storage
- New tape robot

Echo

What is Echo?

- **E**rasure-**C**oded **H**igh-throughput **O**bject store
- Ceph backed object store
  - Not a file system
  - Access provided via GridFTP, Xrootd, S3, Dynafed
- Erasure coding, not replication
- Current: 29PB usable, 40PB raw, soon to be 39PB usable.
- In production since Feb 2017
- ATLAS and CMS fully migrated

Erasure Coding

- Echo uses an 8/3 EC scheme
  - Each object (actually - placement group):
    - 8 data shards, 3 parity shards, each shard on a different storage node (73% usable)
  - Largest known (to us) EC Ceph cluster
- Benefits from EC?
  - Price
  - Throughput
  - Fault tolerance
- Disadvantages?
  - Sparse reads of the data are inefficient
  - Complexity (sync between 11 machines)
  - Less community expertise

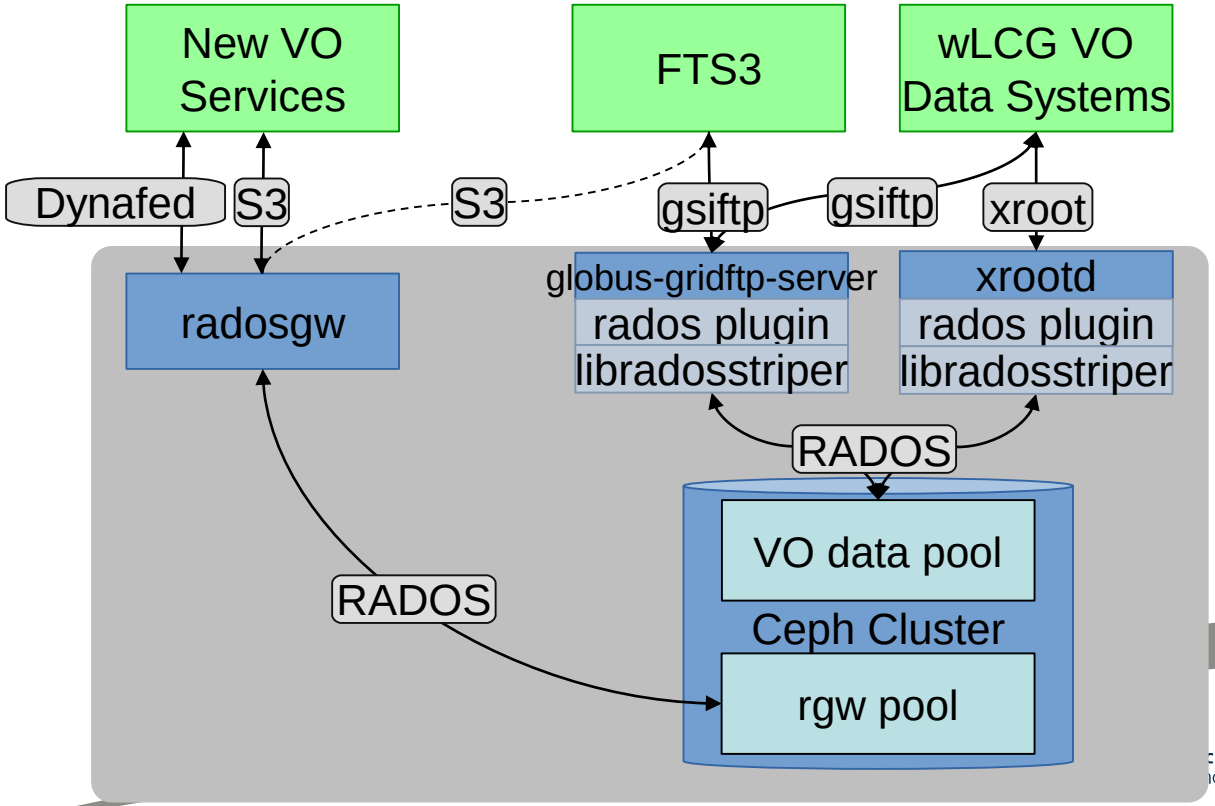
Typical Storage Node Spec

- 128GB RAM
- 24*12TB HDD
- 2*8 core (16HT) CPU (Intel E5-2620V4)
- 2*10Gb networking (1 for external access, 1 for internal cluster operations)

Access (1)

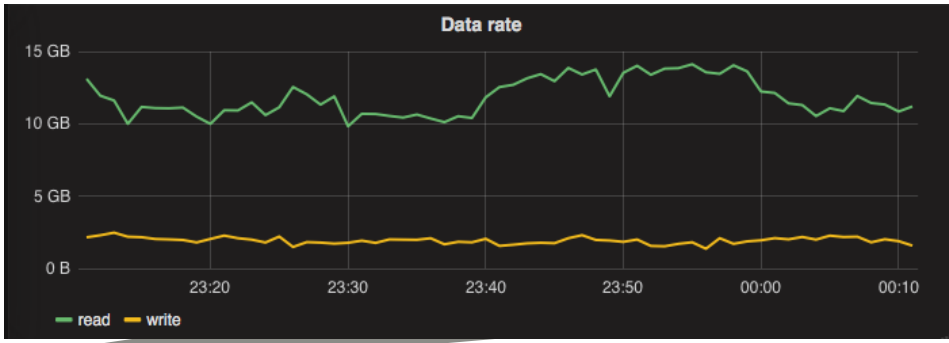
- Core design goal of Echo: Maximise throughput, avoid SPOFs
  - CRUSH means no MD server to slow us down– all gateways have a copy of the CRUSH map,
    - Monitors only maintain the master copy
  - Each batch node has a gateway, so data can pass direct to batch node from 8 storage nodes
  - 7 external gateways

Access (2)


New VO Services
FTS3
wLCG VO Data Systems
Dynafed
S3
S3
gsiftp
gsiftp
xroot
radosgw
globus-gridftp-server
rados plugin
libradosstriper
xrootd
rados plugin
libradosstriper
RADOS
RADOS
VO data pool
Ceph Cluster
rgw pool

Echo Status

- 20GB/s throughput on a routine basis
- No data loss (directly) attributable to hardware failures
- Day-to-day management overhead decreasing
- ATLAS & CMS fully migrated
- LHCb waiting on DIRAC fixes
  - All data mirrored
- ALICE Real Soon Now



Disk read errors

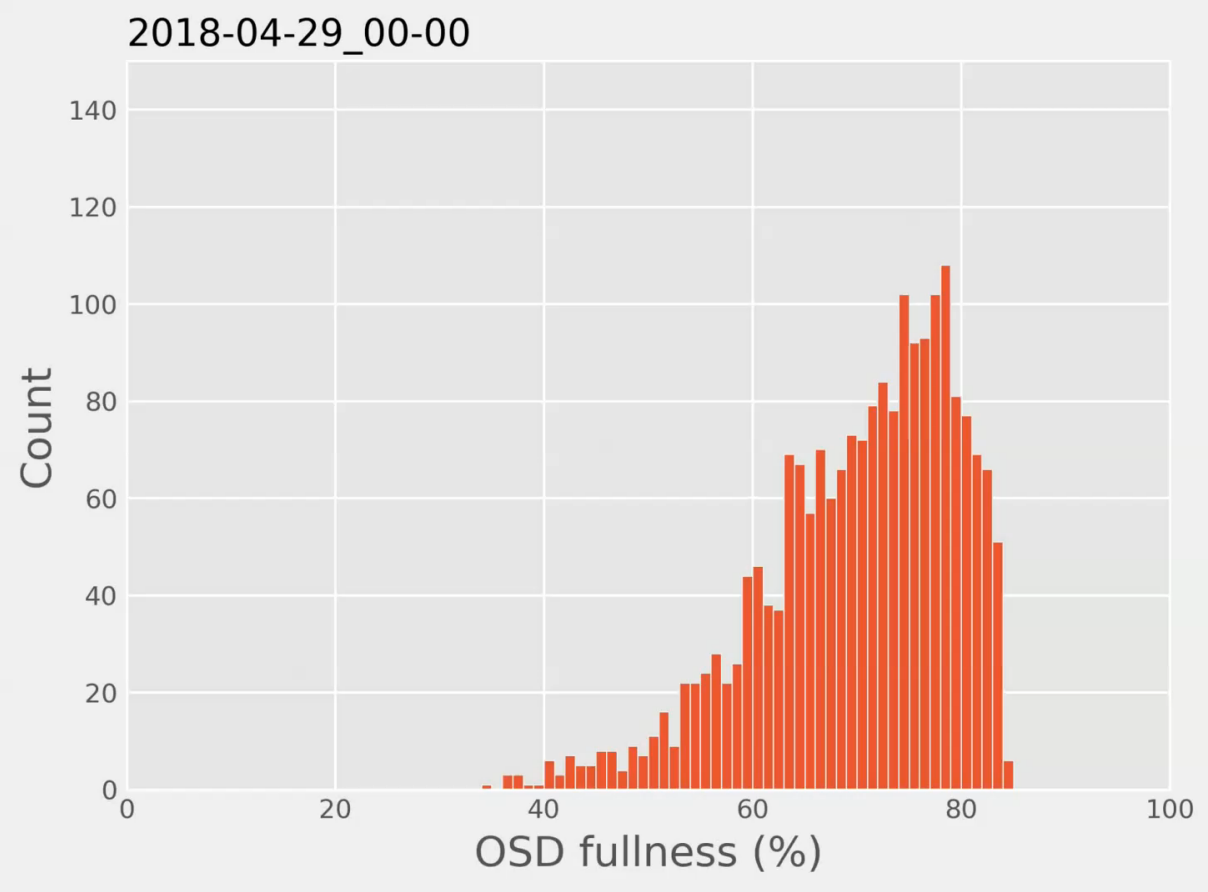
- Ceph is very sensitive to disk read errors
  - Read errors are passed to Ceph by OS (no RAID card)
  - Highest error state (‘HEALTH_ERROR’) if found during scrubbing
    - Easy fix
    - Appropriate behaviour for replicated pools, less so for EC
- Dedicated procedural pipeline for dealing with problem disks
  - Establish vendor tolerance for errors on returned disks
- Now very smooth

Data distribution

- Large variations in the amount of data* that Ceph's placement algorithm chooses to put in each disk (OSD)
  - No protection against overfilling
  - 1 disk fills, cluster enters error state, problems start
- Mid 2018: Big nuisance; hard to add hardware, constant micromanagement of full disks
  - Manually change reweight of full disks
  - Difficult to fill cluster beyond ~70%

*Actually, number of placement groups

Data Distribution



UK Research and Innovation

The Upmap Balancer (1)

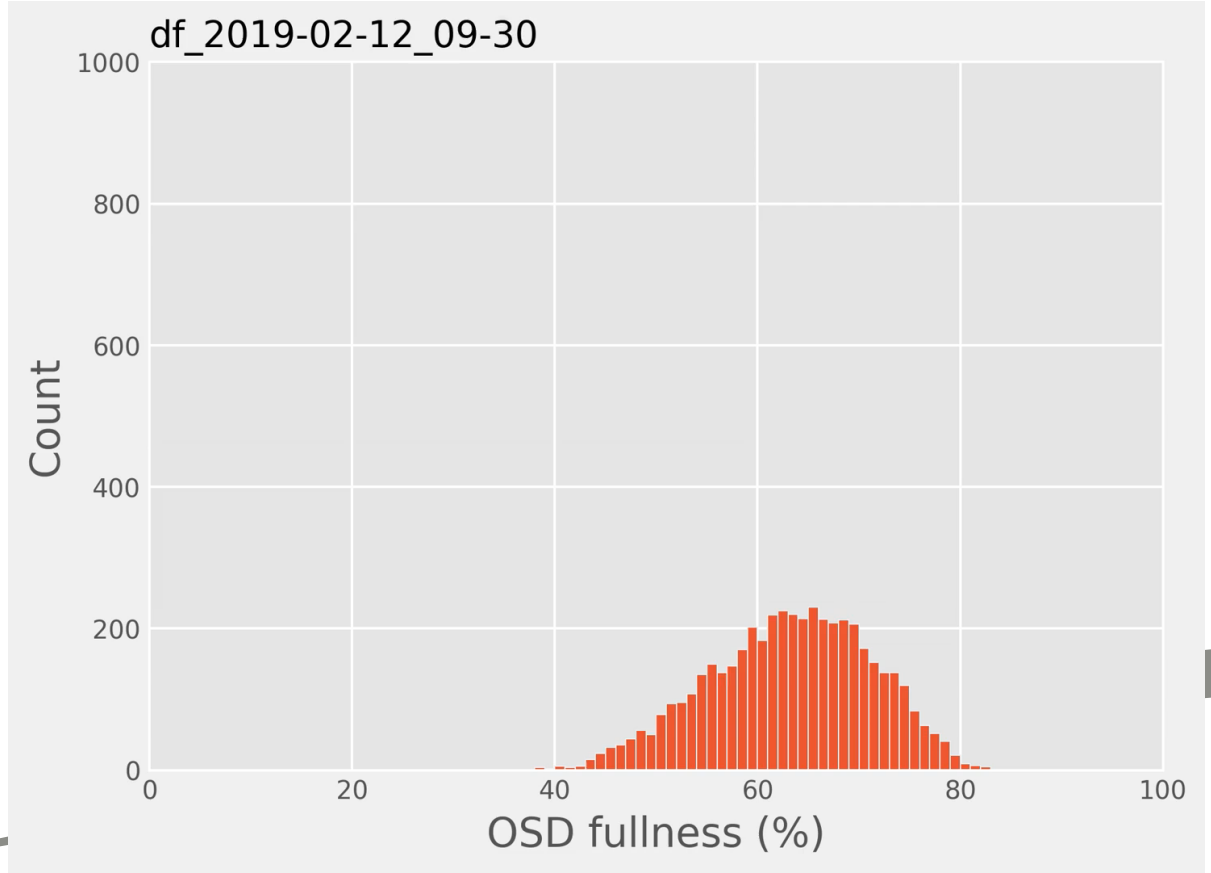
- New feature in Ceph Luminous v12.2 – ‘Upmap’
- Explicitly map a specific placement group to a specific disk
  - Mapping stored with the CRUSH map
- Used to implement automatic balancing
  - Difficult to move from a reweight-balanced cluster to an upmap-balanced one without mass data movement
  - Dan van der Ster wrote a script to avoid this by freezing PGs*
- Greatly improves data distribution

Science & Technology Facilities Council

UK Research and Innovation

*https://gitlab.cern.ch/ceph/ceph-scripts/blob/master/tools/upmap/upmap-remapped.py

The Upmap balancer in action



UK Research and Innovation

The Upmap Balancer (2)

- Principled: “Aren’t you undermining the point of Ceph by adding a static lookup on top of CRUSH?”
  - True, up to a point
  - Balancer works to minimise upmaps and deletes unneeded ones
- Practical: Upmap adds complexity to core monitor operations
  - Internal cluster management tasks take longer, more RAM
  - Not yet solved; upgraded Ceph version to reduce memory consumption by other processes
- Nonetheless, big improvement!

Cluster Scale

- Many challenges come down to Echo being really big
- Big cluster means:
  - Monitor load
  - CRUSH map size
  - Routine need for disk error management
- Mitigations found – some recently:
  - osdmaps creation too long -> increase mon_lease timeout
  - RAM usage on storage while rebuilding -> add more RAM

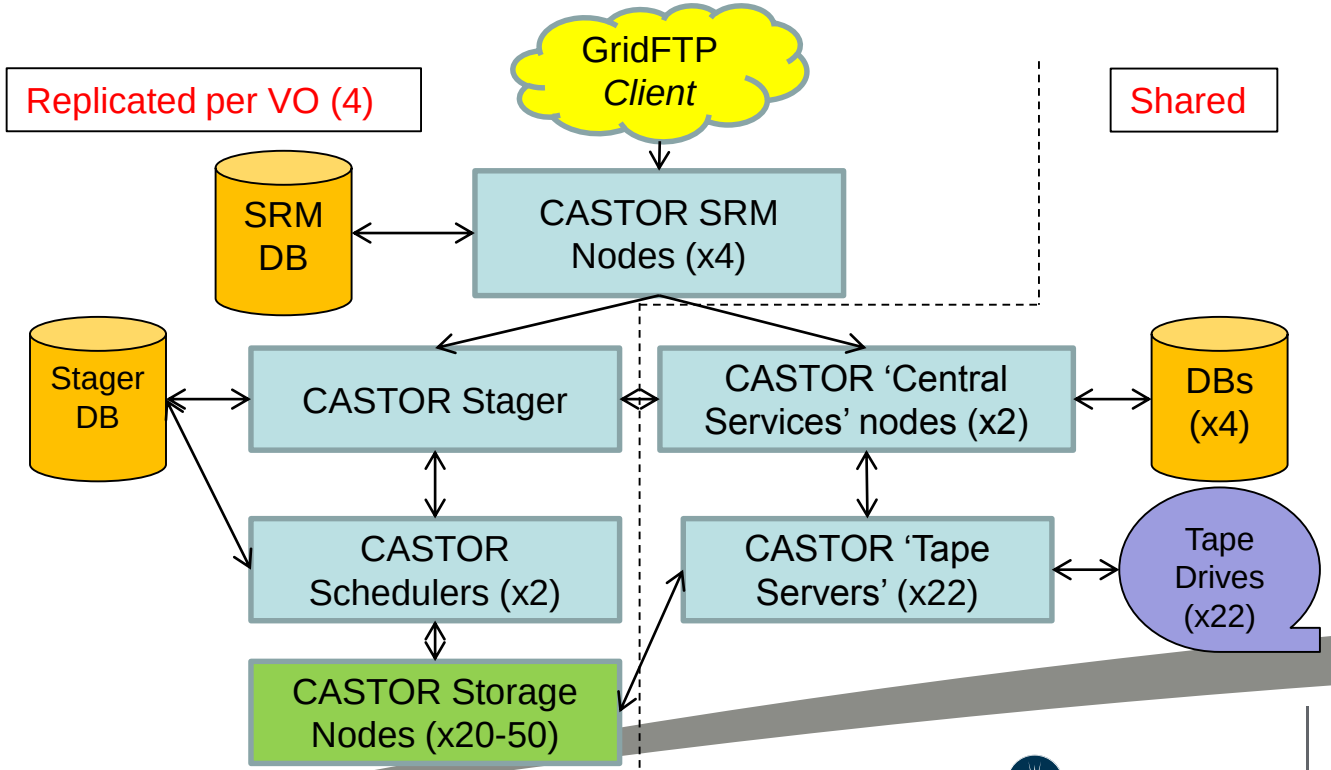
CASTOR

UK Research
and Innovation

CASTOR Rationalisation

- CASTOR infrastructure before Echo deployment
  - 13PB storage, across 137 nodes
    - 108 used for ‘disk-only’
  - 29 management/interface nodes
    - All ‘pets’ with specific roles
  - 2 oracle RACs plus 2 standbys
- New system only needs 1 pool, ~2PB storage
  - Need to reimplement to reduce overhead

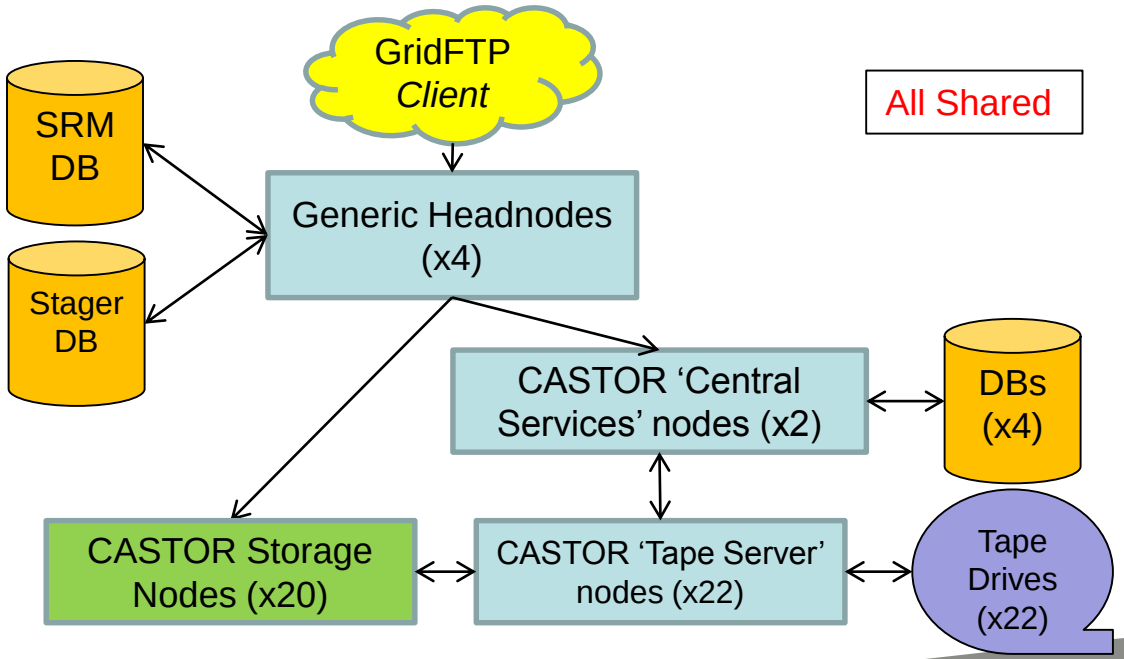
A Picture of CASTOR - GridFTP



Goals

- Principles
  - Minimise node count
  - Minimise configuration complexity
- 1 RAC, 1 stager DB, 1 species of headnode, 1 storage pool
  - Reimplemented headnodes as generic ‘cattle’ machines

A Picture of CASTOR - GridFTP



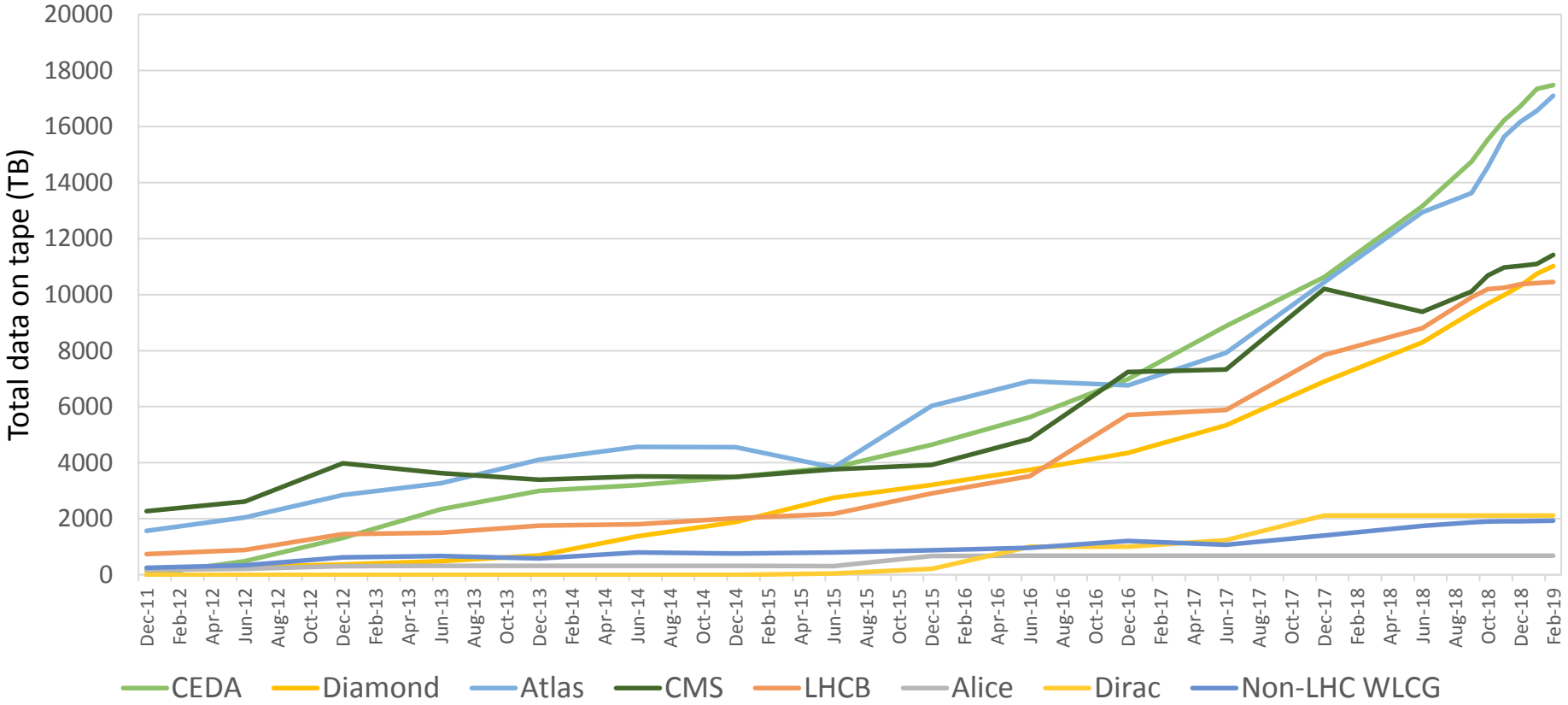
Result

- Plan implemented
  - New storage system in production Nov 2018
  - VOs migrated as they started using Echo
  - All done except for LHCb and ALICE's disk
- New headnodes hosted by VMWare
  - 'Cattle' where all nodes host all management processes

Non-WLCG CASTOR Users (1)

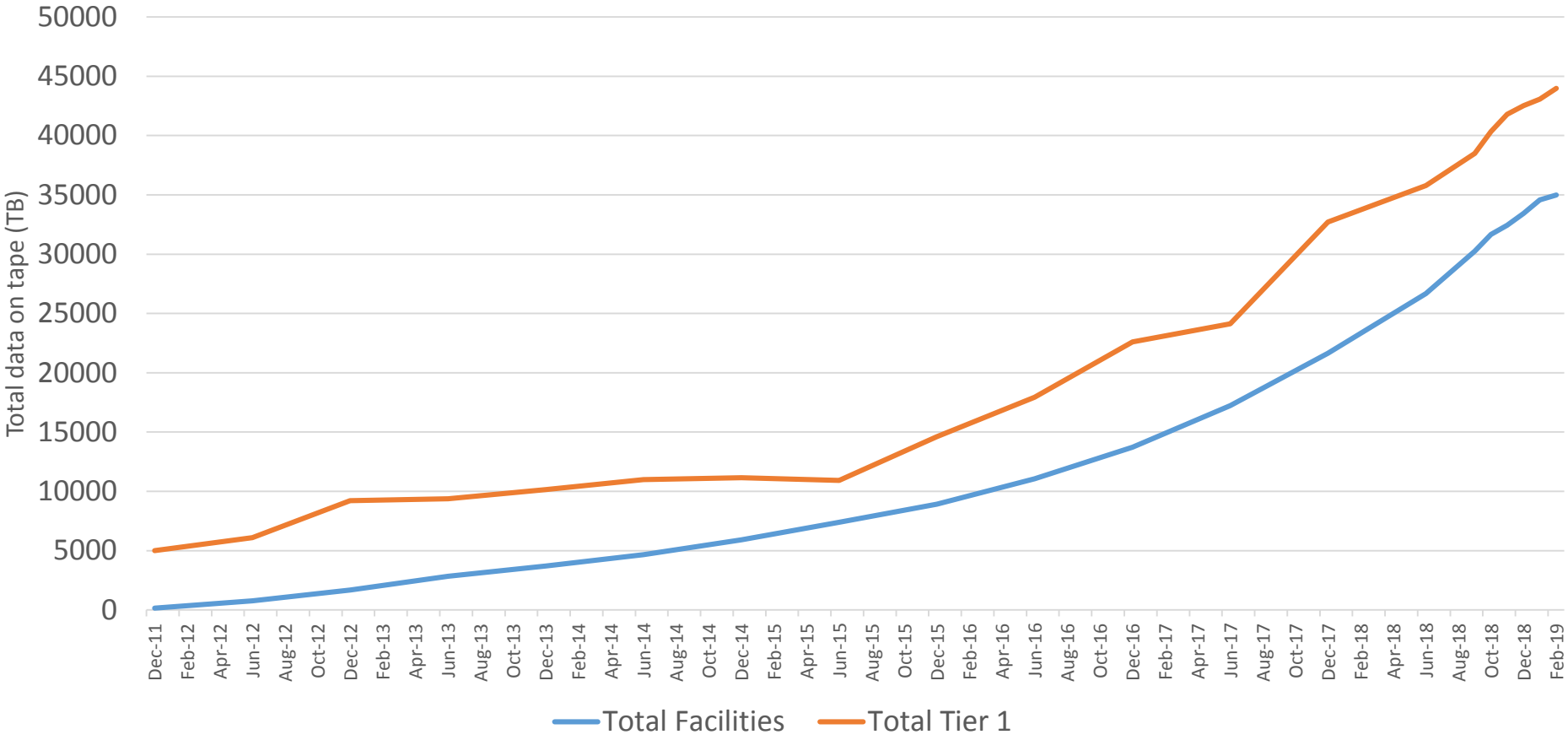
- CASTOR also used to support local RAL facilities
  - Diamond Light Source
  - Centre for Environmental Data Analysis
- Access Patterns
  - Much closer to WORN tape ideal then WLCG
  - 40 PB stored on tape
- Entirely separate CASTOR instance

Tape Data Holding by user



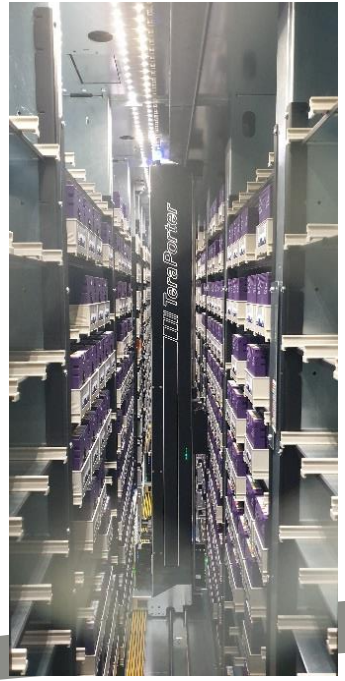
Dirac: https://dirac.ac.uk/category/home/ Diamond: https://www.diamond.ac.uk/Home.html

Tape Data Holding by User Type



Non-WLCG CASTOR Users (2)

- New robot – Spectra Logic Tfinity
- Initially used for non-WLCG communities
- Switch from T10K to LTO/TS1160
- See Martin Bly's talk for more information



Science & Technology Facilities Council

UK Research and Innovation

CASTOR Future

- CERN CASTOR service is scheduled to be discontinued ~ mid 2019
  - CERN moving to CTA[1]
- RAL has a tender out for a replacement
- Motivations for divergence
  - Non-HEP users growing
  - Already diverged from CERN for disk

1: **An efficient, modular and simple tape archiving solution for LHC Run-3, S Murray, et al** http://iopscience.iop.org/article/10.1088/1742-6596/898/6/062013/pdf

UK Research
and Innovation

Thank you very much

Spare sides

Abstract (for ref)

- RAL's Ceph-based Echo storage system is now the primary disk storage system running at the Tier 1, replacing a legacy CASTOR system that will be retained for tape. This talk will give an update on Echo's recent development, in particular the adaptations needed to support the ALICE experiment and the challenges of scaling an erasure-coded Ceph cluster past the 30PB mark. These include the smoothing of data distribution, managing disk errors, and dealing with a very full cluster.
- In addition, I will discuss the completed project to remodel RAL's CASTOR service from a combined disk and tape endpoint to a low-maintenance system only providing access to tape.