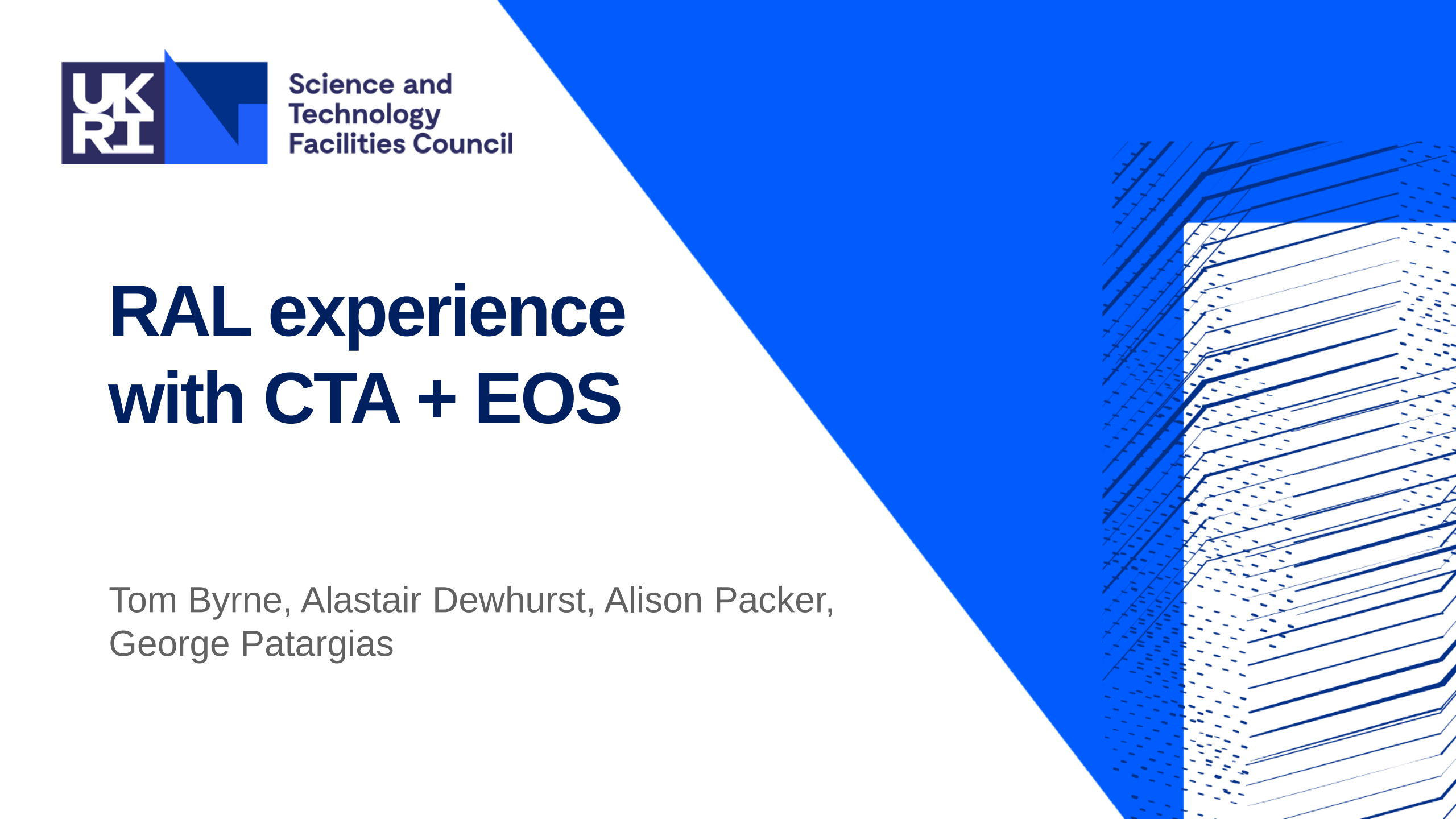


RAL experience with CTA + EOS

Tom Byrne, Alastair Dewhurst, Alison Packer,
George Patargias

Background

- Castor in use at RAL since 2006 to provide tape archival service – both Tier-1 and STFC Facilities
- CERN move to CTA and therefore support and collaboration with CERN would cease for Castor.
- RAL evaluated other commercial solutions to provide business case to all stakeholders in 2019
  - Commercial solutions less attractive than moving to CTA - migration process, development effort, funding model plus vendor lock-in all weaknesses of this approach
  - Contrast with a migration process with data in situ, development is part of ongoing collaborations within the community, CTA is open source and strong links with CERN - the team wanted to continue collaborating on tape archive solutions.
- **Antares chosen as name for production CTA at RAL service**
- *Side project 1: migrate from Oracle to Spectra tape libraries*

Setup and Support

Antares

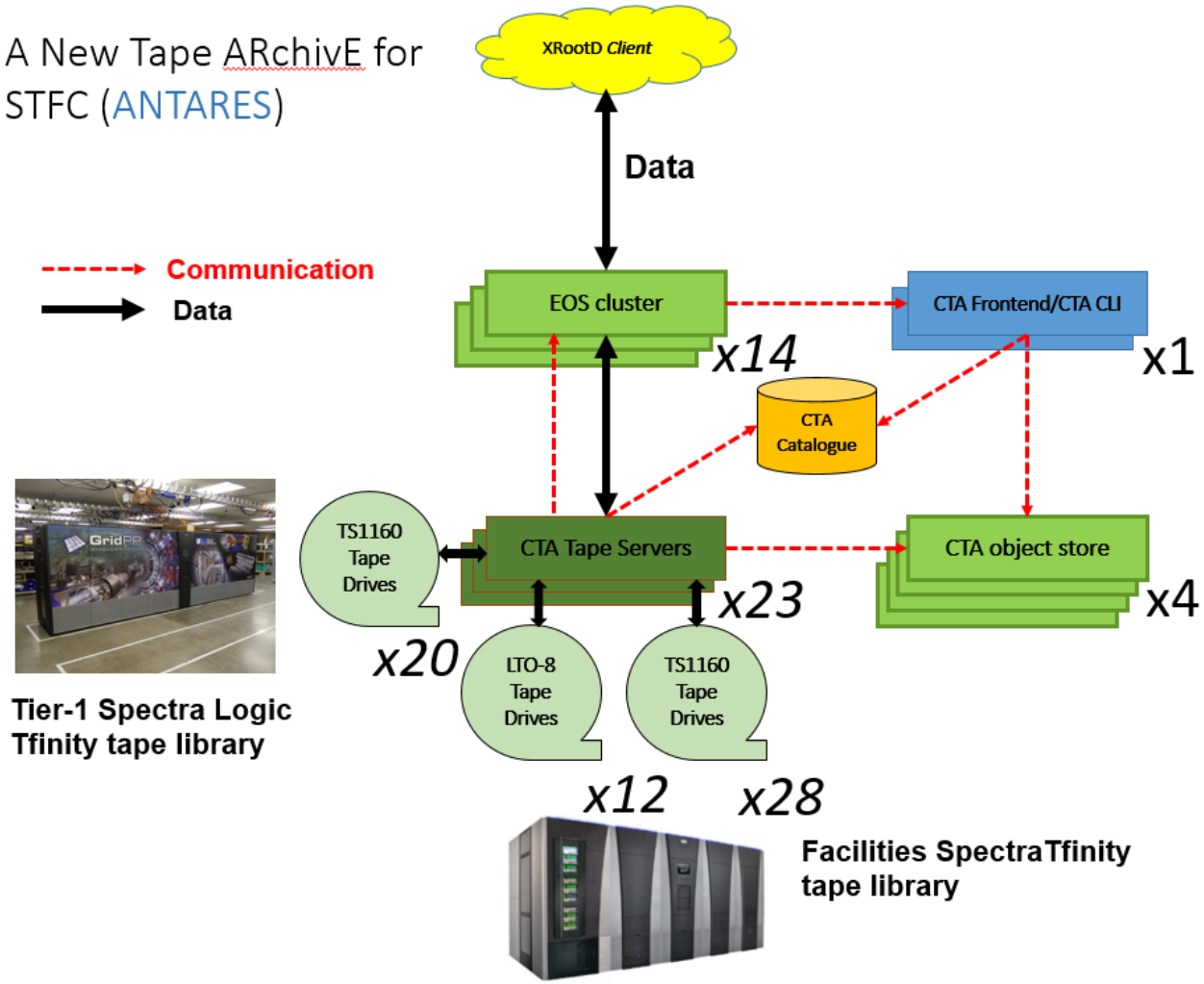
- CERN team generous with time to assist us getting setup. RAL/CERN meeting October 2019 to understand more about requirements for CTA – software, hardware & migration
  - Need to deploy EOS - CERN EOS team support and regular EOS workshops
  - Castor to CTA migration differs at RAL, far smaller archives to migrate but other challenges with 2 instances, but CERN migration tools can be used at RAL
  - Decision to copy CERN deployment model as closely as possible, tracking upgrades etc.
  - Documentation, resources, community forum etc. https://cta.web.cern.ch/cta/
- Pandemic challenges:
  - Hardware specced, ordered, delivered just in time for U.K. LOCKDOWN! Can not get it installed… (hardware info on backup slide for those who want it)
  - Go with advice to setup Kubernetes EOS-CTA instance
  - Hardware for CTA frontend servers, Ceph object store, EOS, databases and tape servers finally installed March, 2021
- *Side project 2: The RAL Tier-1 is replacing its network.*

Antares Setup

2 Storage Admins (George & Tom) - primary staff working on the project, George from 2020 and both since Dec. 2020, with support from:

- **hardware team who look after installs/networking/fabric/tape library for the Tier-1 and archives.**
- **DBA support running the Oracle Databases and now working on migration**
- **Support from VO Liaisons with testing**

Side project 3: Upgrade Castor to 2.1.19 in readiness for migration to CTA



CTA Setup and Testing

- EOS – most unfamiliar component, no prior experience running EOS at RAL. Also new hardware – EOS is all SSD nodes, benchmarking carried out to evaluate performance and any bottlenecks.
- Ceph object store – setup straightforward as one of a number of Ceph clusters run at RAL, configuration management, deployment, monitoring etc. can follow our standard setup.
- Databases – Oracle RAC, similar to Castor, known setup, install, config. and documentation etc. from CERN CTA team covers schema etc.
- Tape Servers – again, templates for installation, configuration, monitoring etc. very like existing tape servers.
- Functional testing carried out by storage admins, VOMS setup in place, testing by Atlas and CMS VO Liaisons based at RAL.
- Plan was to focus on migration next….

Tape Challenge – October, 2021

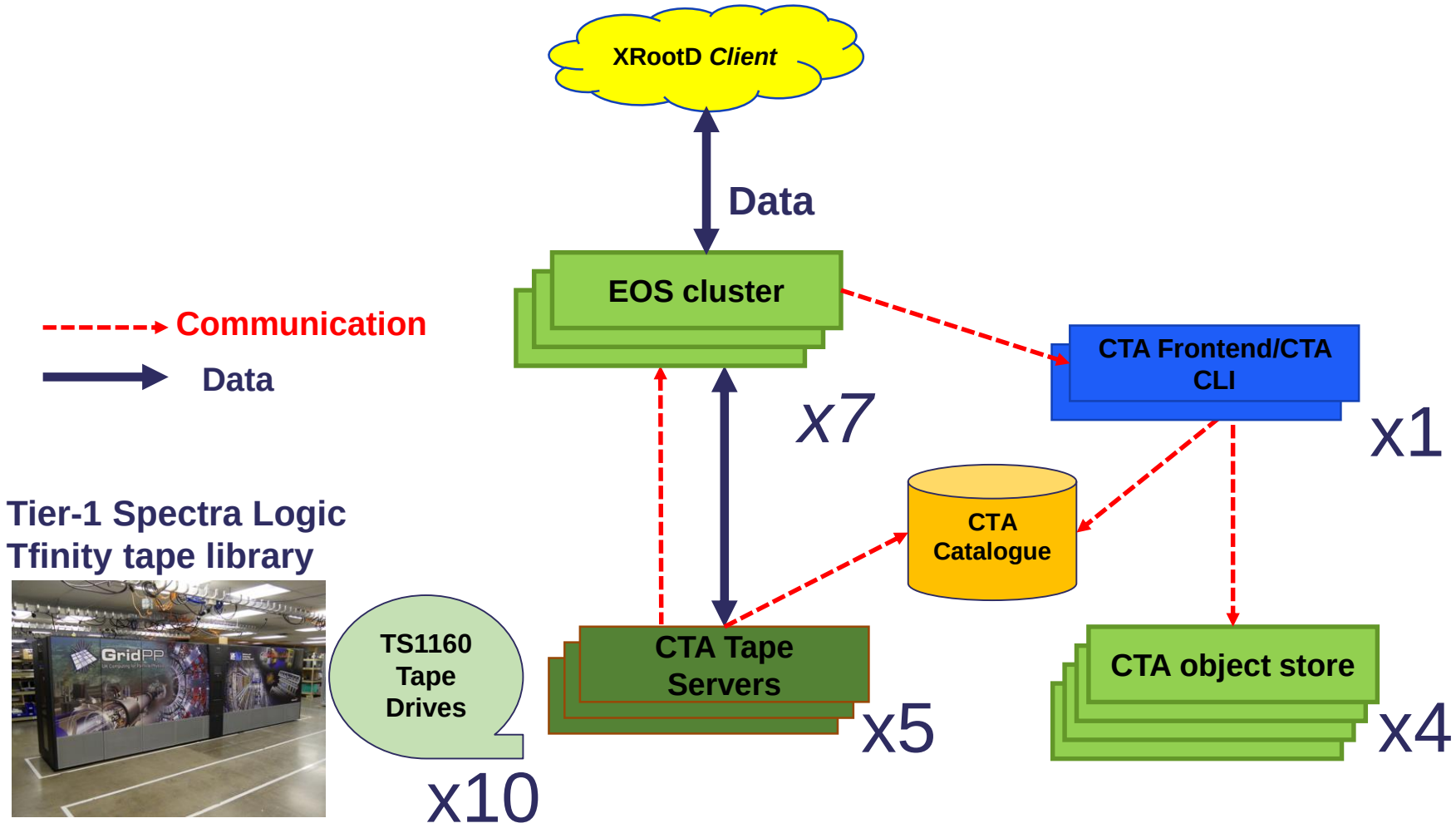
Antares

	Reads (DT) GB/s	Writes (DT) GB/s	Reads (A-DT) GB/s	Writes (A-DT) GB/s	Castor / Antares
ALICE	-	0.08	0.05	0.08	Castor
ATLAS	0.4	1.4	1.2	0.7	Antares
CMS	0.1	0.9	1.5	0.1	Antares
LHCb	-	2.92	1.12	-	Antares

ALICE test used Castor, as ALICE authz not yet tested with Antares.
Tape hardware split between Castor and Antares, 10 x TS1160 drives each.

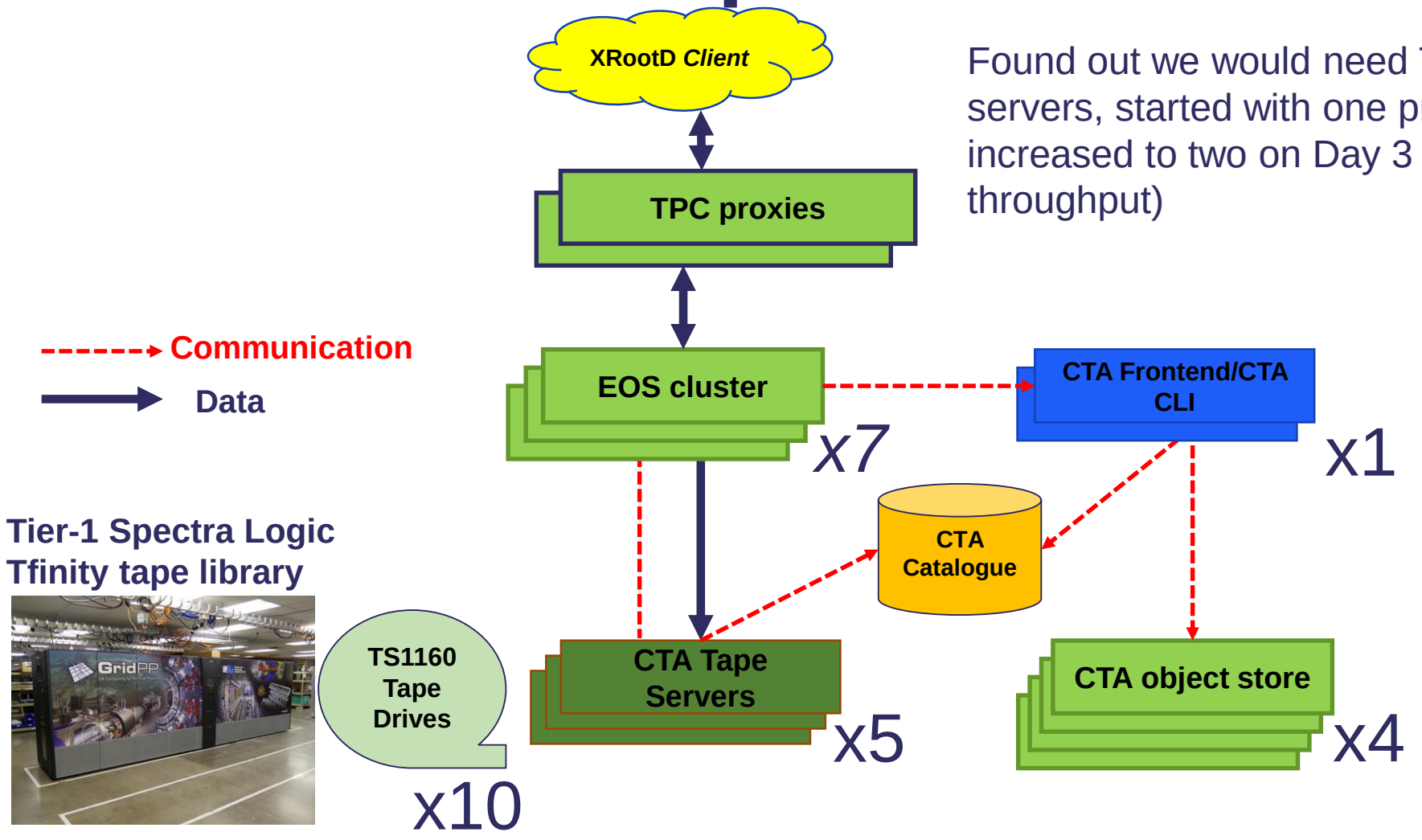
Tape Challenge Antares setup

Antares



Actual Antares setup

Antares



Found out we would need TPC proxy servers, started with one proxy (25Gb), increased to two on Day 3 (40Gb throughput)

Tape challenge outcomes

- Review with CERN CTA team after first week to understand some issues encountered:
  - Garbage collection not set up correctly and site name needed setting up in EOS config
  - Issue for CMS initially as files not on the archive (was a Rucio config issue) – corrected and further testing completed OK
  - LHCb had written but read tests still to do (were run by Chris Haen in November)
  - Confirm requirement for proxies

- Antares performed well for writes given the hardware available.
- Antares also performed well for reads, possibly slightly unrealistic scenario as all files were located on a small number of tapes.
- Antares should have almost 3 times as much tape hardware available at the start of Run 3.
- We still need to work with ALICE to migrate to using the Antares endpoint.

Tape challenge outcomes – EOS+CTA

Antares

	Required read rate GB/s *	Achieved read rate GB/s **	Required write rate GB/s *	Achieved write rate GB/s ***	Castor / Antares
ATLAS	0.4	1	1.4	1.1	Antares
CMS	0.1	2.7	0.9	3.5	Antares
LHCb	1.12	2	2.92	1.5 ****	Antares

*** The largest requested read/write rate from the VO**

**** Maximum sustained read rate from the EOS buffer seen from our monitoring in the past 90 days**

***** Maximum sustained write rate to the VO tape pool tape seen from our monitoring in the past 90 days**

****** A misunderstanding of the required rates lead to half the number of tape drives being allocated for LHCb during the tape test.**

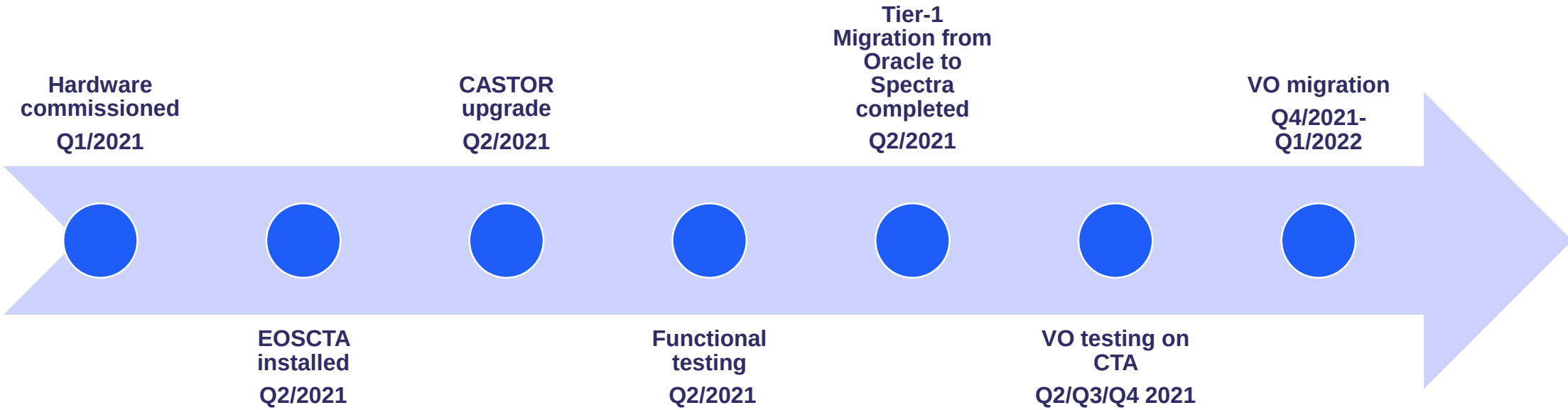
Next steps

- MIGRATION
  - Castor namespace migrates to EOS
  - Castor catalogue/tape data migrates to CTA catalogue.
  - Migration testing of Castor Tier-1 before Christmas
  - All VO testing (including ALICE) December/January on migrated test instance
  - All VO migrated simultaneously to production during long downtime at the start of 2022.

- HARDWARE
  - The number of tape drives will double when we migrate from Castor.
  - Additional drives and media are currently being procured.

- Network Upgrades ongoing – completion scheduled for January, 2022

A Year in the Life of CTA at RAL



Science and Technology Facilities Council

Antares

Questions?

GridPP
UK Computing for Particle Physics

Science and Technology Facilities Council

Antares

Backup Slides

GridPP
UK Computing for Particle Physics

Hardware

Node Type & Number	Function	Model	CPU	Memory	Disk	Network
EOS 12 x production 2 x test	Namespace management & disk cache	DELL R740XD	2 x Intel Xeon Gold 5218	192 GB	System + 1 NVMe + 16 x 2TB SSD	1 x Mellanox ConnectX-4 LX Dual Port 10/25GbE 1 x Intel Ethernet I350 Dual Port 1GbE BASE-T Adapter
Ceph 3 x production 2 x standby/dev	For transient data, queues and requests stored as objects in key-value store	DELL R6415	1 x AMD EPYC 7551	128GB	System + 8 x 4TB SSD	1 x Mellanox ConnectX-4 LX Dual Port 10/25GbE
Database 2 x Oracle RAC production 2 x Oracle RAC test	CTA catalogue	DELL PowerEdge R440	2 x Intel Xeon Gold 5222	192 GB	System + separate storage array (~90TB capacity)	1 x Broadcom 5720 Dual Port 1 GbE 1 x Dual-Port 1GbE On-Board LOM
Tape Server	RAL intend to allocate 1 tape server per 2 tape drives (initially)	DELL PowerEdge R640	2 x Intel Xeon Silver 4214	96 GB	2 x 240GB SSD SATA	1 x Mellanox ConnectX-4 LX Dual Port 10/25GbE
Frontend Servers (virtual)	Accepts archive/retrieve requests from EOS and send to CTA object store. Used for admin commands					

Tape library migrations

Antares

- **Support for Oracle tape ends mid-2020s**
- **Two Spectra TFinity libraries purchased in 2019 and 2020**
- **CTA is integrated with Spectra and IBM currently, but not Oracle**
- **Migrate 130PB of data from Oracle SL8500 to Spectra before CTA goes into prod:**
  - Tier-1 migration completed May, 2021
  - Facilities CEDA migration completed August, 2021
  - Diamond Archive migration scheduled to complete December, 2021

RAL CTA Talks

- Discussion with CERN over Tape adoption in October 2019: https://indico.cern.ch/event/848893/
- RAL & DESY CTA discussion December 2020: https://indico.cern.ch/event/981157/
- RAL Report at the Tape Evolution pre-GDB in February 2021: https://indico.cern.ch/event/876801/contributions/4211820/attachments/2186938/3695353/CTA-preGDB-Feb2021-final.pdf
- Tape Evolution pre-GDB report March 2021: https://indico.cern.ch/event/876787/contributions/4258900/attachments/2205380/3731235/TapePreGDBSummary20210310.pd
- CTA Update at GridPP46 meeting September 2021: https://indico.cern.ch/event/1054156/contributions/4491567/attachments/2302094/3915990/CTA-gridpp46.pdf
- Tape Challenge debrief with CERN, October 2021: https://indico.cern.ch/event/1089343/contributions/4579318/attachments/2332472/3975189/AntaresTapeChallengeRecap.pdf
- RAL Tape challenge Report November 2021: https://indico.cern.ch/event/1094310/contributions/4608204/attachments/2344213/3997376/Antares20211111.pdf

RAL Tier-1/Tape talks

Antares

Migration to Spectra Library:

- George Patargias - talk at HEPiX in October 2019 on the Facilities Spectra Robot: https://indico.cern.ch/event/810635/contributions/3593326/attachments/1927972/3192345/WLCGTape_HepixOct2019.pdf

- Martin Bly - site update at HEPiX in March 2021 on the completion of tape library migration: https://indico.cern.ch/event/995485/contributions/4263427/attachments/2207923/3736135/2021-03%20-%20HEPiX%20Spring%202021%20-%20RAL%20Site%20Report.pdf

RAL Tier-1 Network, paper from vCHEP 2021:

dx.doi.org/10.1051/epjconf/202125102074