CTA at RAL

In production since 4th March 2022!

Tom Byrne, Alastair Dewhurst, Alison Packer,
George Patargias

EOS Workshop 09/03/2022

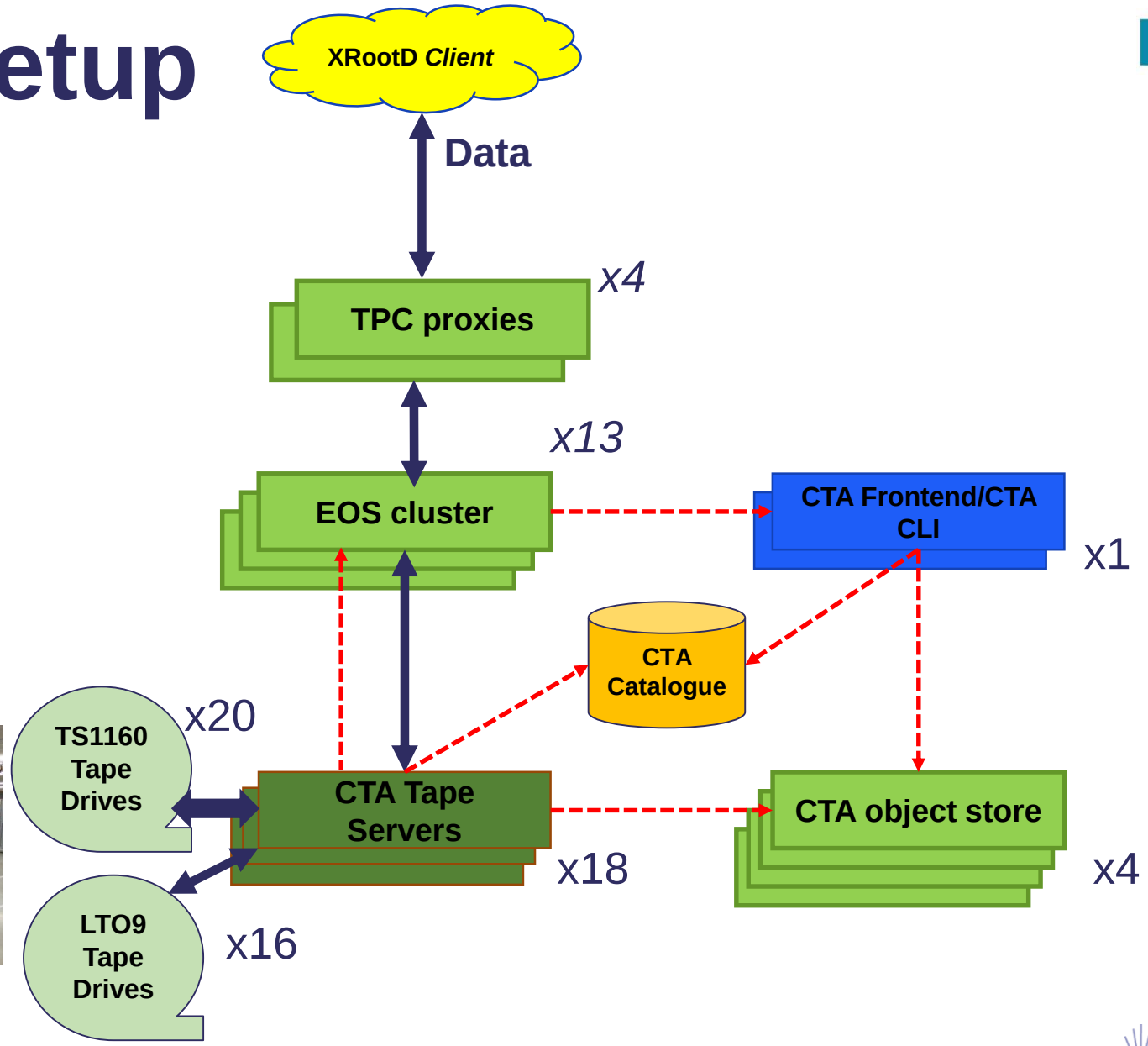
Background

- 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
- CTA chosen, many advantages:
  - provided for a migration process with data in situ from Castor
  - 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 - name for production CTA at RAL service

CTA operational status timeline

- Q1 2021: Hardware networked
- Q2 2021: EOSCTA installed
- Q3 2021: CASTOR upgrade
- Q3 2021: Antares instance created
- Q4 2021: LHC Tape Challenge
- Q1 2022: Tier-1 CASTOR to Antares migration
- Q1 2022: Functional testing/CMS Tape Challenge

Antares Setup



Antares Team

Antares

2 Storage Admins (George & Tom) - primary staff working on the project, George from Feb. 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 on the migration PL/SQL scripts
- Support from VO Liaisons with testing

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 – configuration, monitoring etc. very like existing CASTOR tape servers.
- Functional testing carried out by storage admins, VOMS setup in place, testing by Atlas and CMS VO Liaisons based at RAL.

Tier-1 Castor to Antares Migration

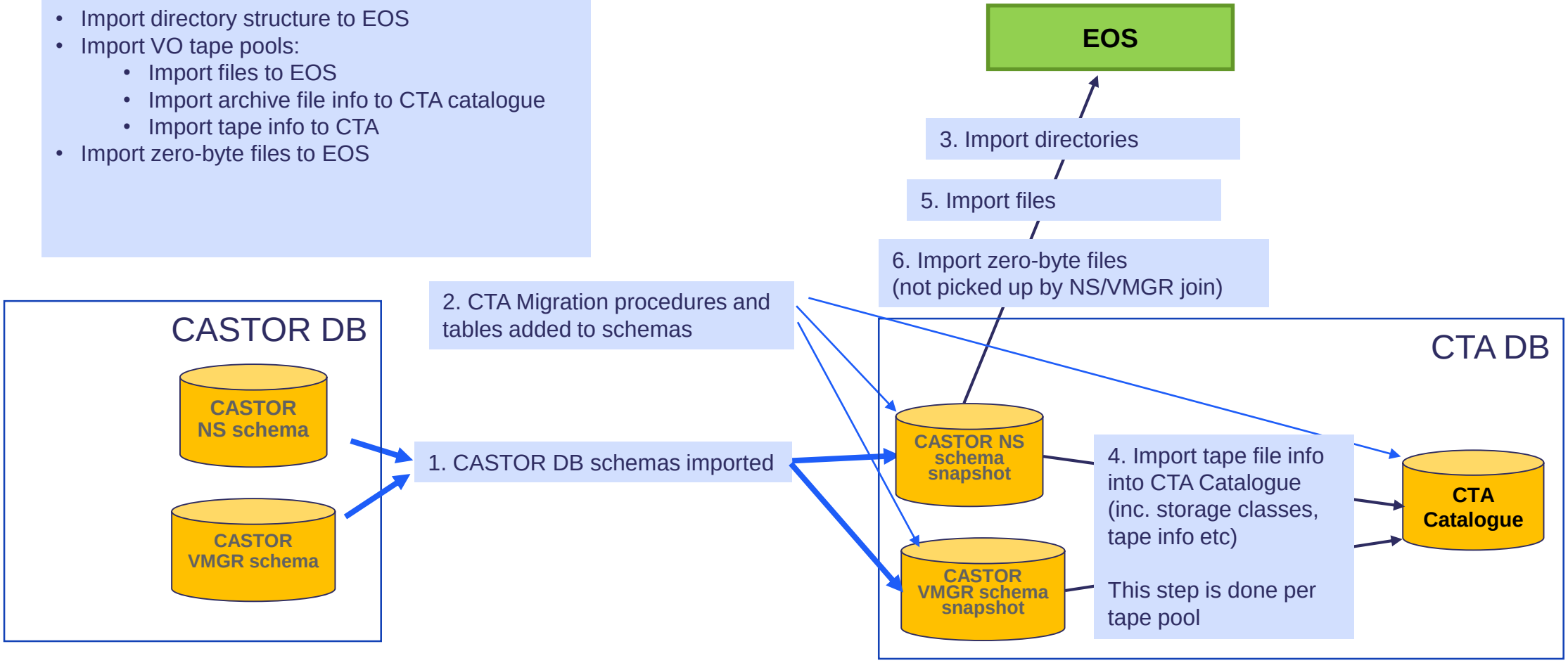
Antares

- **Migration method:**
  - ➢ **Castor namespace injection to EOS**
  - ➢ **Castor tape metadata migration to CTA DB**

- **Migration pre-requisites:**
  - ✓ **Upgrade to CASTOR to 2.1.19-3**
  - ✓ **Import CASTOR DB schemas (NS,VMGR,STAGER) snapshot to the CTA DB**
  - ✓ **Review/modify PL/SQL scripts to be run on the imported CASTOR DB schemas**
  - ✓ **Further namespace clean up (repack files to the right tape pools)**
  - ✓ **Set up a migration node to run the migration client tools**
  - ✓ **Estimate timings to be scheduled in the intervention plan**

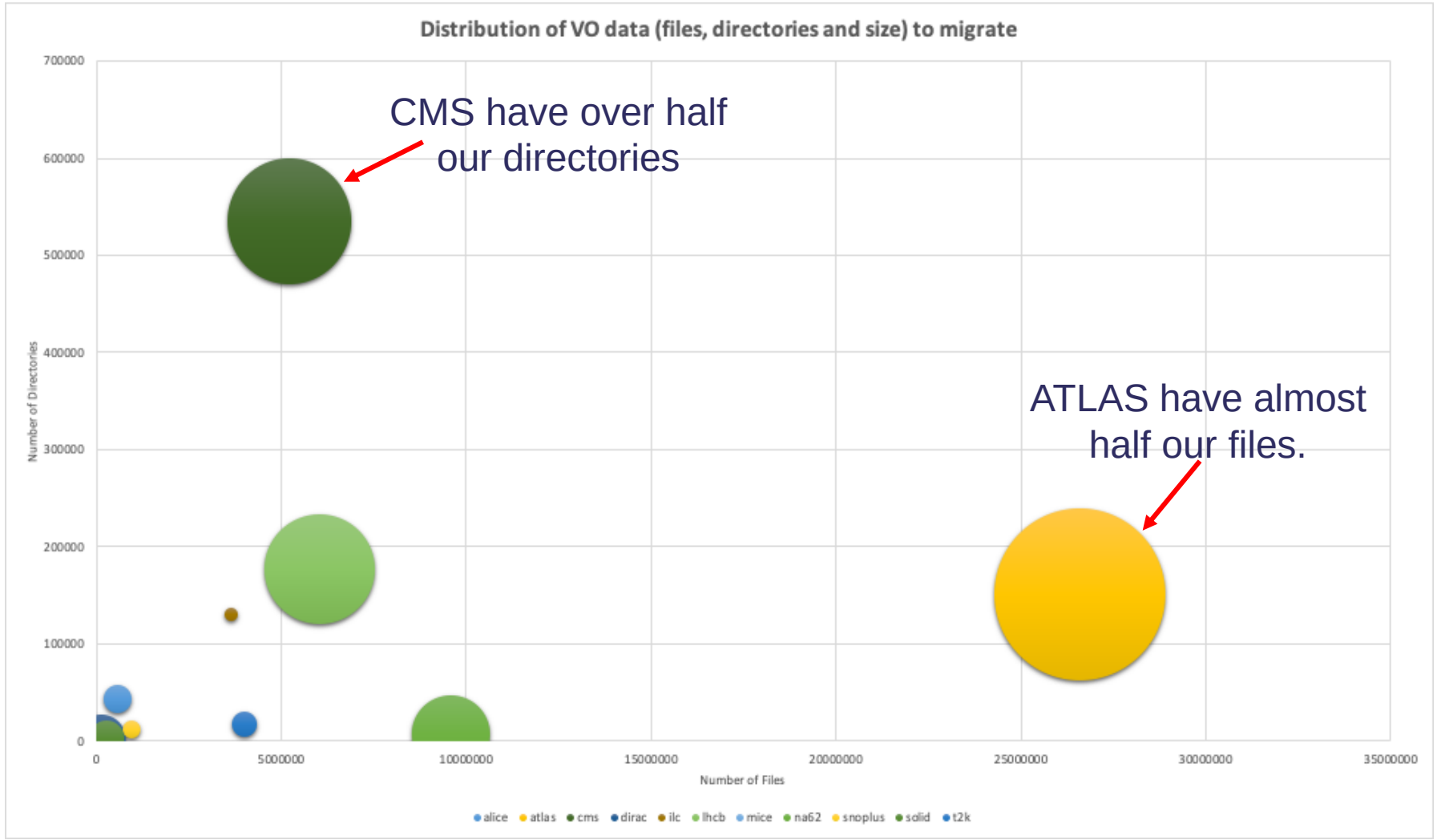
Tier-1 Castor to Antares Migration

Steps needed for each VO:
- Create VO in CTA
- Import directory structure to EOS
- Import VO tape pools:
  - Import files to EOS
  - Import archive file info to CTA catalogue
  - Import tape info to CTA
- Import zero-byte files to EOS



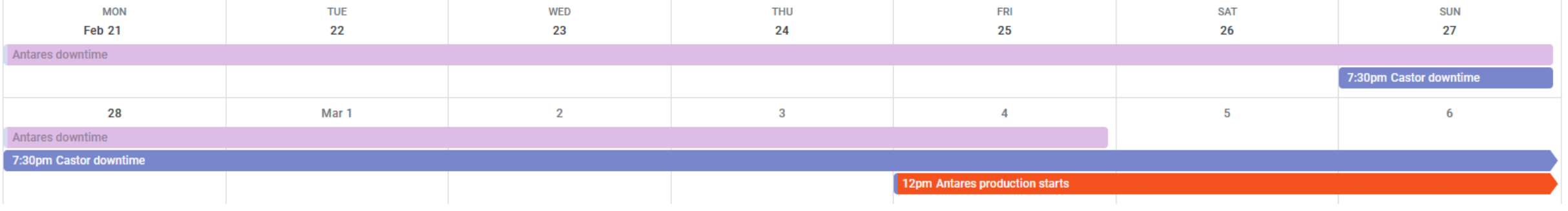
Tier-1 CASTOR to Antares migration

Total:
1,079,217 dirs,
57,011,928 files,
70.5PB



Castor to Antares Migration

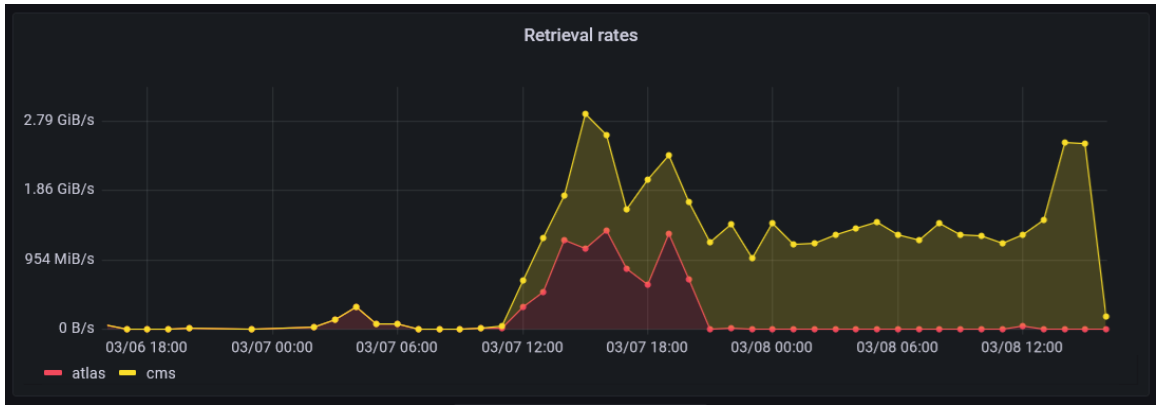
- Actual migration – downtime Sunday pm to allow Castor to drain - to midday Thursday
- Backups of Castor taken and file lists created for VOs pre-migration – rollback checkpoint
- Two team members migrating all the VOs one at a time. File lists in EOS produced for each VO to compare with the Castor file list --> small number of anomalies!
- Had to apply dir extended attributes (ACLs and CTA workflows) on the ***whole*** dir structure after migration
- Scripted applying across the whole namespace – ATLAS: 150,000 dirs, CMS: 535,000 dirs, LHCb: 177,000 dirs – and found that 70,000 directories was the maximum namespace size to apply the attributes without hitting the timeout limit



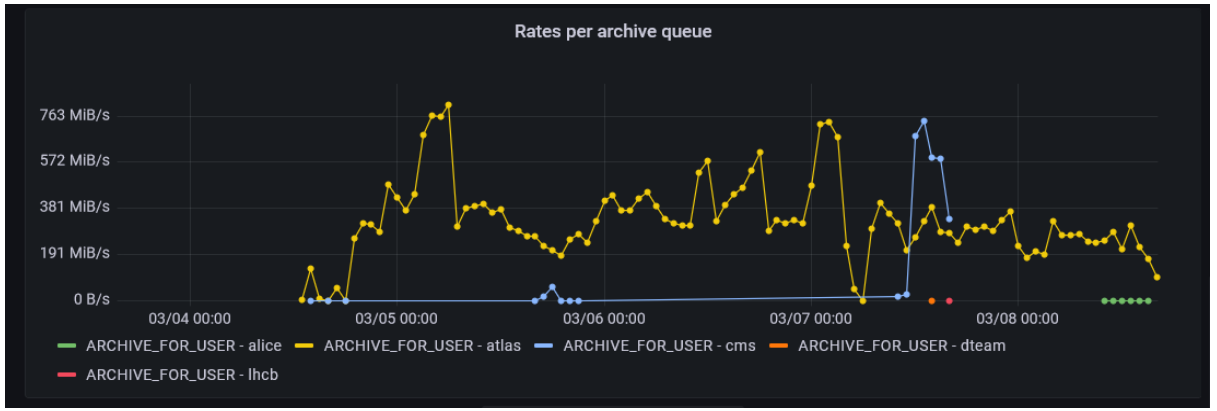
Antares Testing

Test	ATLAS	CMS	LHCb	ALICE
End-to-end client archive/retrieval (internal/external)	Done	Done	Done	Done
FTS XrootD transfer between CERN and Antares	Done	Done	N/A	N/A
WebDAV and FTS WebDAV/TPC	N/A	N/A	Writes confirmed, testing continues	N/A
FTS multihop transfer between Antares and offsite via Echo	Done	Done	N/A	N/A

Antares Performance



Retrieval rate last 2 days



Archival rate last 5 days

Last 5 days tape server and EOS node network

Next steps

- Tape Challenges in March pre LHC RUN 3 (In progress for CMS)
- Enable access to non LHC VO's
- Upgrade EOS/CTA to the versions deployed at CERN
- Upgrade to EOS5
- Prepare and execute the migration of CASTOR Facilities
- Migrate the CTA Catalogue from Oracle to PostgreSQL

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

LHC VO Run3 requirements

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

2021 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 ***	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.**

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