Science and Technology Facilities Council

RAL CTA Architecture and Procurement

Alison Packer

9th December, 2020

Agenda

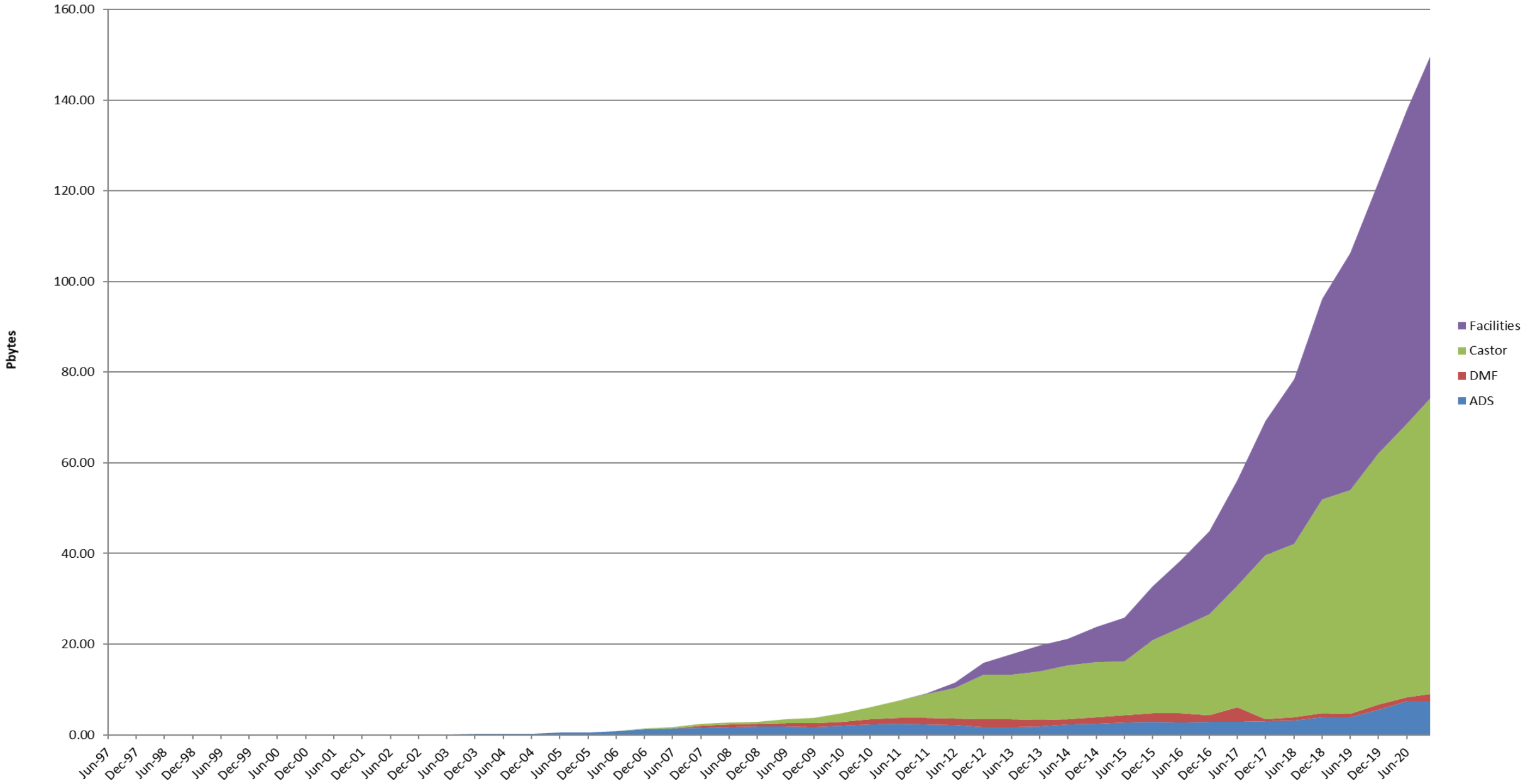
- Castor Replacement
- CTA Architecture
- Test/CI setup
- Hardware Procurement
- Commissioning

Science and Technology Facilities Council

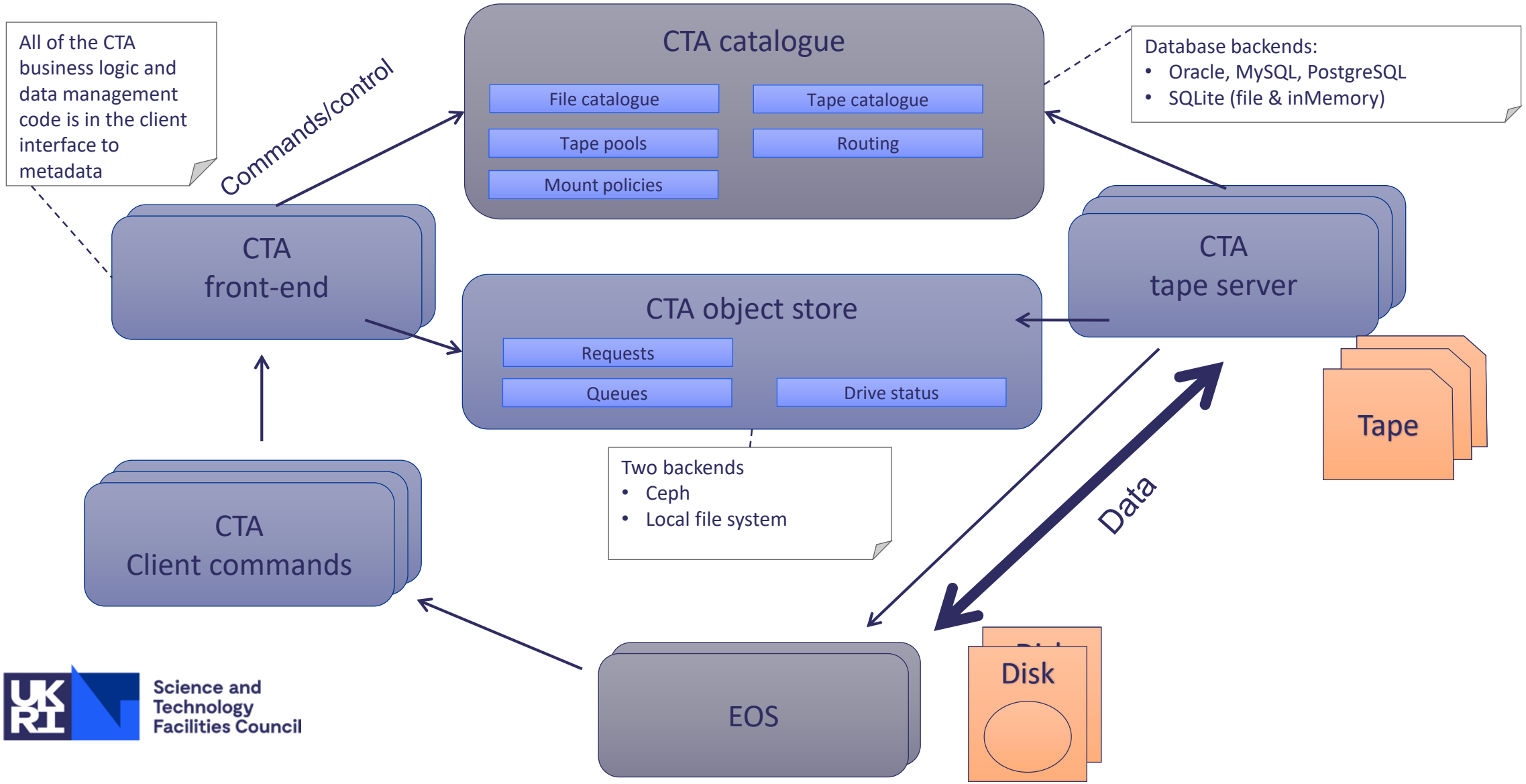
Castor Replacement

- Castor running at RAL for ~14 years
- Two instances, one for Tier-1 (66PB/65M files) and one for our local Facilities (Diamond Archive, CEDA) (60PB/75M files)
- Tier-1 had been disk and tape storage, moved disk storage to Ceph object store in 2017
- CERN announced EOL for Castor and move to CTA, needed to look at options for replacing other legacy backup technology
- Review of commercial alternatives, including HPSS, DMF etc. and then CTA in 2019.

Total Holdings



CTA Architecture (CERN Team's diagram)



Science and Technology Facilities Council

CTA Testing

- CERN team requested that the continuous integration environment be set up on a VM at RAL
- Advised to extrapolate the hardware from this testing
- CERN expect to provide support based on reports which can demonstrate the issue in the CI environment
- BUT…. We had hardware money to spend immediately!

From architecture to hardware list…

- Frontend servers
- CTA catalogue servers – a database management system
- A filesystem or Ceph
- Tape servers
- Network and management switches

Frontend Servers

- Expect to be lightly loaded, CERN team using OpenStack VMs for testing and expected to continue in production
- The frontend server's job is to queue requests for archive and retrieve from the client running XrootD commands
- Also, responds to admin queries, querying queues, listing files
- Purchased an additional hypervisor to be added to our existing VMWare setup for these – expect to start with a pair for resilience

CTA Catalogue

- Planning to model our initial implementation on CERN's as our project also includes migration from Castor
- CERN have 2 x 2-node RAC implementation on Oracle 19c (one production, one standby)
- RAL running Castor on Oracle RAC and will continue with Oracle RAC for CTA – advice from CERN was to spec. as for Castor
- CTA is written in C++ with SQL and does not include Oracle PL/SQL which would complicate the provision of alternative database backends
- RAL plan to move to PostgreSQL instead of Oracle once migration is completed
- Test instance has PostgreSQL backend

Ceph Object Store

- RAL Data Storage Group run multiple Ceph clusters (storage for OpenStack plus CephFS cluster) and the largest, now running for nearly 4 years in production is “Echo” – Ceph object store providing LHC experiment disk storage
- Will install a small separate Ceph cluster for CTA – with the mons and storage on the same nodes
- CTA object store holds transient data, queues and requests stored as objects in key-value store

Tape servers (& tape library)

- Spec. provided by RAL Fabric Team based on experience with Castor
- CERN have implemented one tape server per tape drive
  - advised testing this first before moving to 1 server for 2 tape drives (RAL production plan)
  - CTA testing with several tape drives per tape server when running on virtual tape drives (MHVTL) so high level of confidence
- RAL migrating all data from Oracle SL8500 libraries to Spectra Tfinity before migrating from Castor to CTA
- CTA support for Spectra and IBM tape libraries (not Oracle at present)

Sundry items

- Network switches purchased for the CTA nodes – Mellanox SN2100 and SN2410
- Some “general purpose” hosts which will provide for the PostgreSQL servers (test DB is set up in a VM), InfluxDB, Perfsonar etc.
- Icinga and TIG stack for monitoring etc. already in place for other services at RAL and so will be used for CTA too.

Hardware

Node Type & Number	Model	CPU	Memory	Disk	Network
EOS x 14 (12 prod/2 test)	DELL R740XD	2 x Intel Xeon Gold 5218	192 GB	System + 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 x 5 (4 prod/1 test)	DELL R6415	1 x AMD EPYC 7551	128GB	System + 8 x 4TB SSD	1 x Mellanox ConnectX-4 LX Dual Port 10/25GbE
Database x 4 (2 x 2 Oracle RAC) (prod & test)	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 x 12 (10 prod/2 test)	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

Hardware commissioning

- Hardware arrived in March, 2020 as planned and was racked
- Commissioning delayed by Covid-19
- Networking involves multiple groups as connection with other services, e.g.

  Tier-1, Facilities archives etc. required - being put in place now
- So cabling, operating system installation, testing and benchmarking still to do

Science and Technology Facilities Council

Questions?