

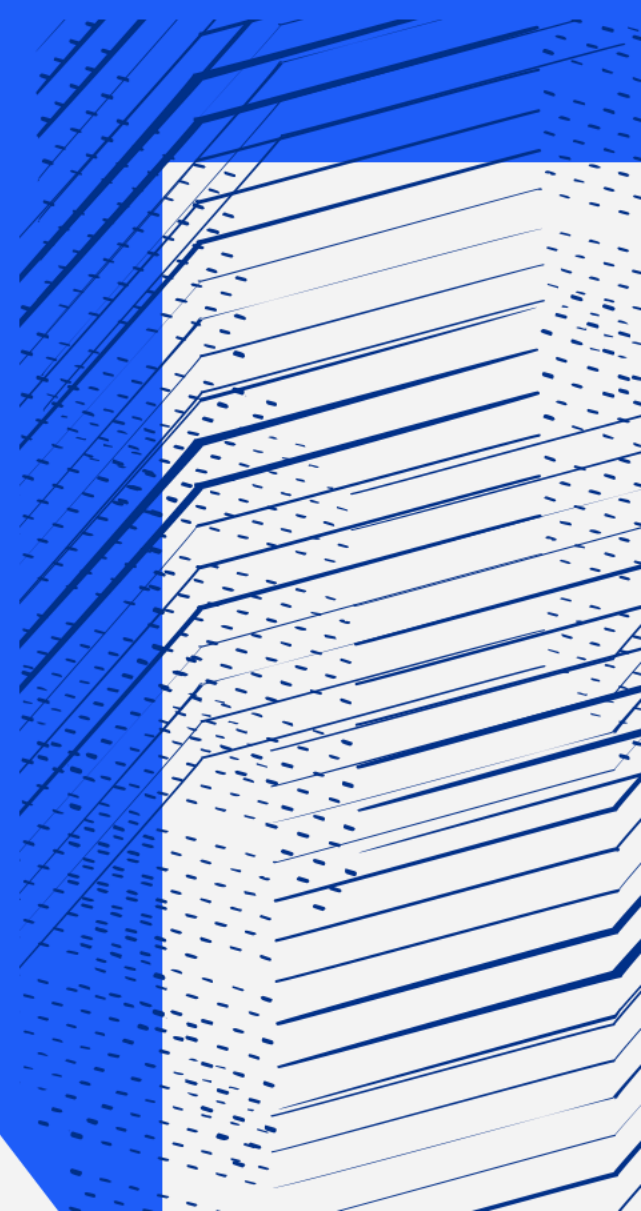


Science and
Technology
Facilities Council

Scientific Computing

RAL Site Report

HEPiX Spring 2024, Paris
Martin Bly et al.
April 2024



Outline

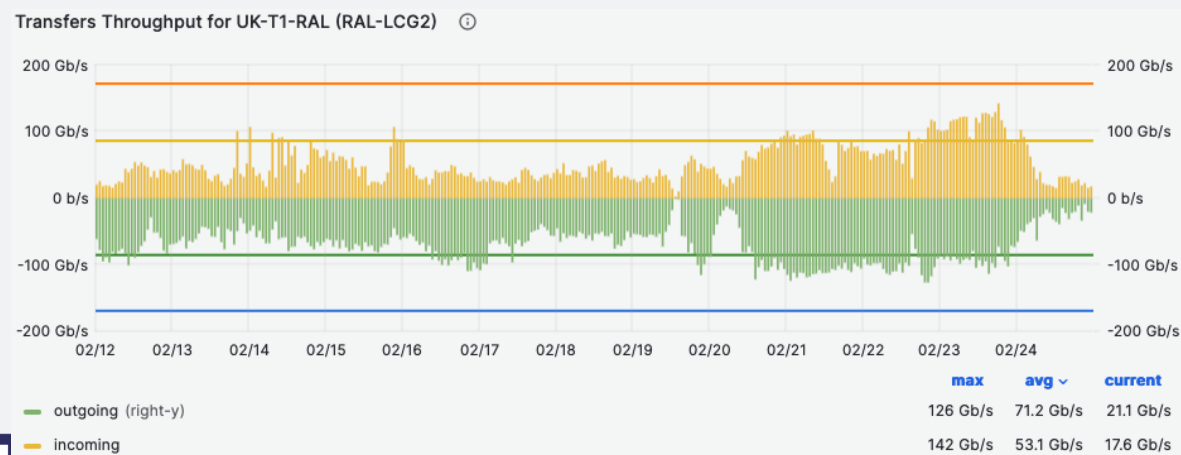
- DC24
- Batch/Tape/Echo
- Procurement
- Time series data
- Space

Thanks to colleagues for their input

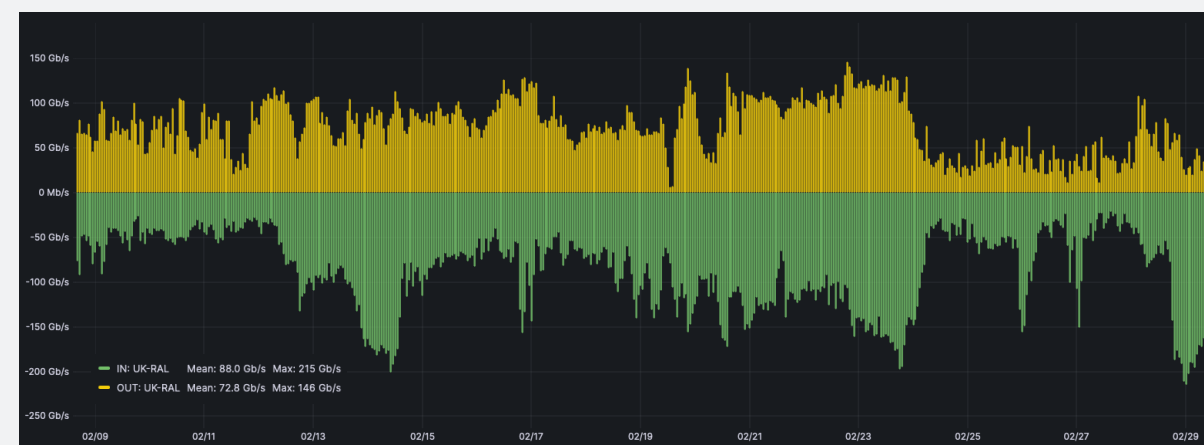
DC24 - Overview

- Overall DC24 was very successful for RAL.
 - Significant network problems at the start of challenge resulted in low initial throughput.
- At the end, throughput was 2 – 3 times higher than we ever saw before:
 - We found a new set of bottlenecks well above Run 3 normal load that can be worked on over the next few years.
 - Fantastic effort from the Tier-1 to debug, fix/optimize in real time.

WLCG DC24 Monitoring

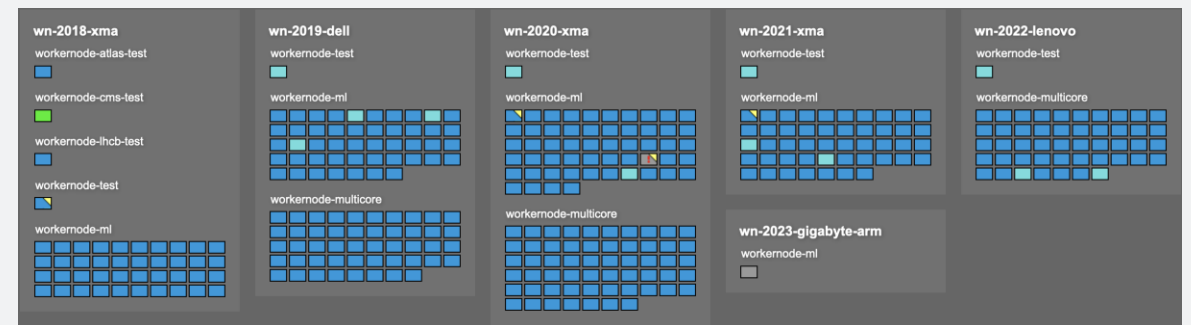
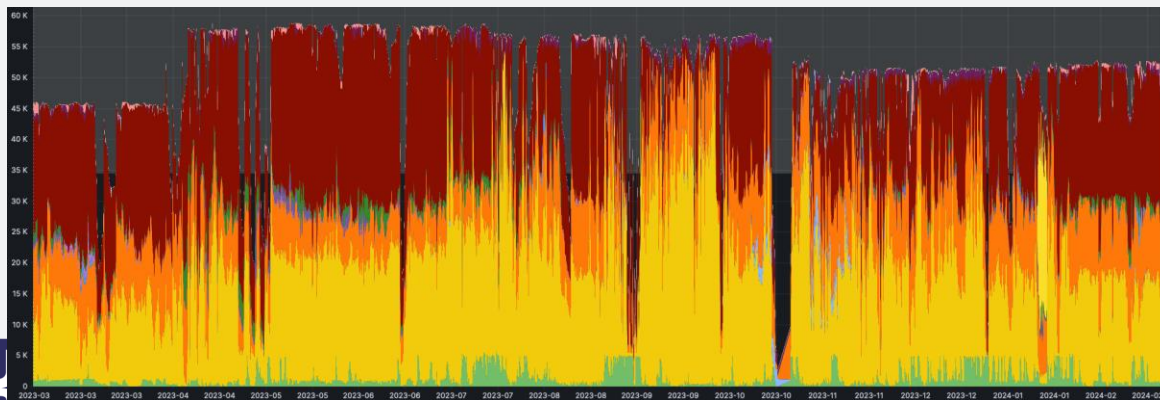


Site Link Monitoring



Batch Farm

- HTCondor and ARC CEs continue to perform very well for the Tier-1.
 - The utilization of the resources very impressive when compared to other science areas.
- The most frequent operational problem is with VOs not always receiving their desired fairshare.
 - This problem will become more complex as we add different architectures.
- Tom Birkett has setup a pre-prod cluster (which runs actual work and is accounted, and can be used to test the latest patches).
 - We have a single node dual stack which is being tested.



Tape Libraries



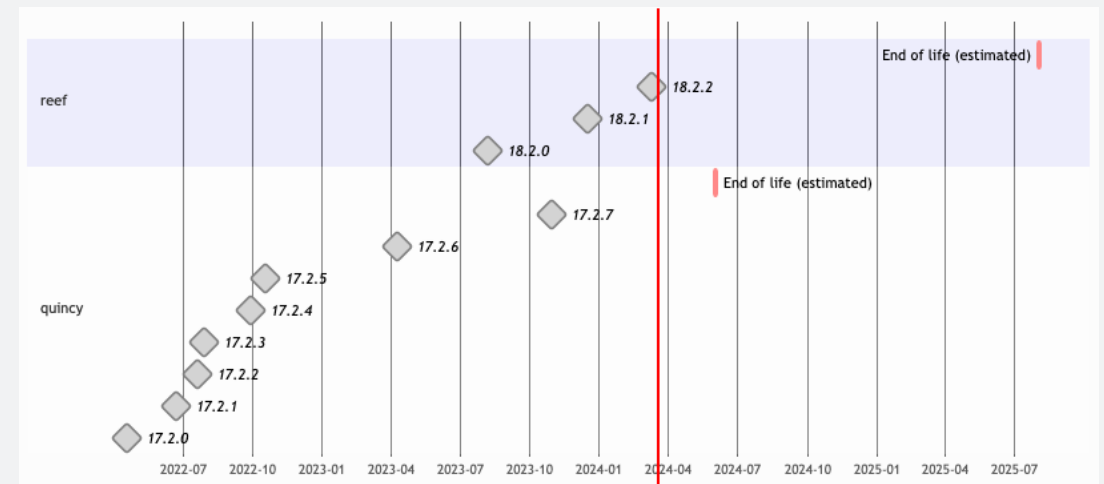
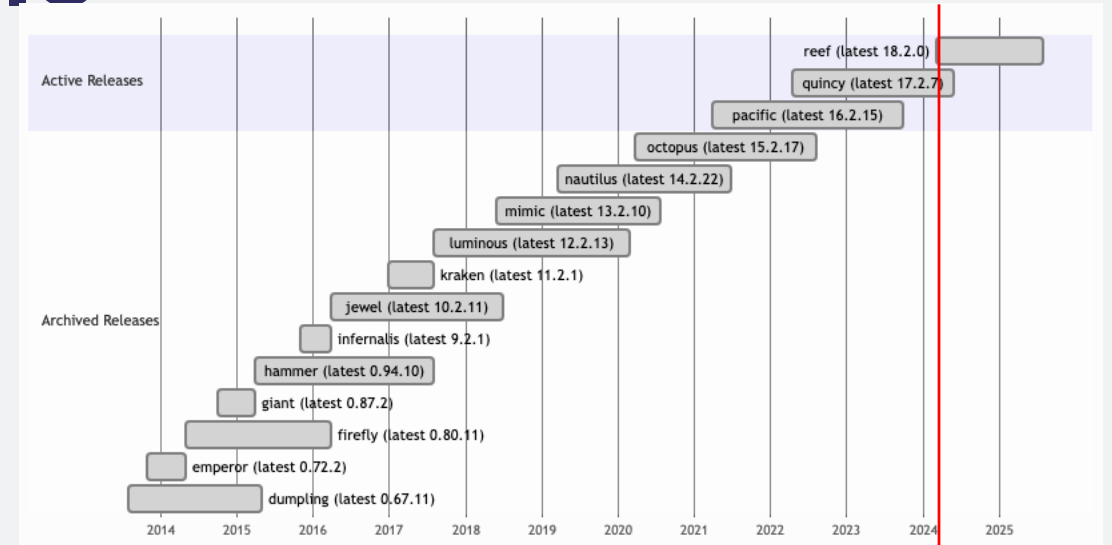
- DMF service now migrated to new server/storage hardware
 - RHEL 8.8
 - Service using Obelix Library
 - Data on tape migrating From Oracle library to Obelix as service runs
- 2 x SpectraLogic tFinity libraries
 - Asterix (WLCG) – 15 Frames
 - Obelix (Others) - 13 Frames
- 1 x Oracle SL8500
 - System backups
 - To be phased out when system backups replacement solution identified.

Echo status and plans

- Echo recently exceed 100PiB of raw capacity.
- Backend storage working very well.
 - Many operational changes can be done transparently, and the team is trying to automate them where possible.
- 2024 Q2:
 - SL7 → Rocky 8
 - Decommission 2017/18 generations of hardware
 - Continue deployment/weighting up of 2022 hardware
- Q3
 - Complete deployment and decommissioning
 - Upgrade Ceph from Nautilus to Pacific
 - Start deployment of 2023 Storage
- Longer term:
 - implement rack-level failure domains
 - Plan to upgrade Ceph Pacific -> Reef in ~ April 2025

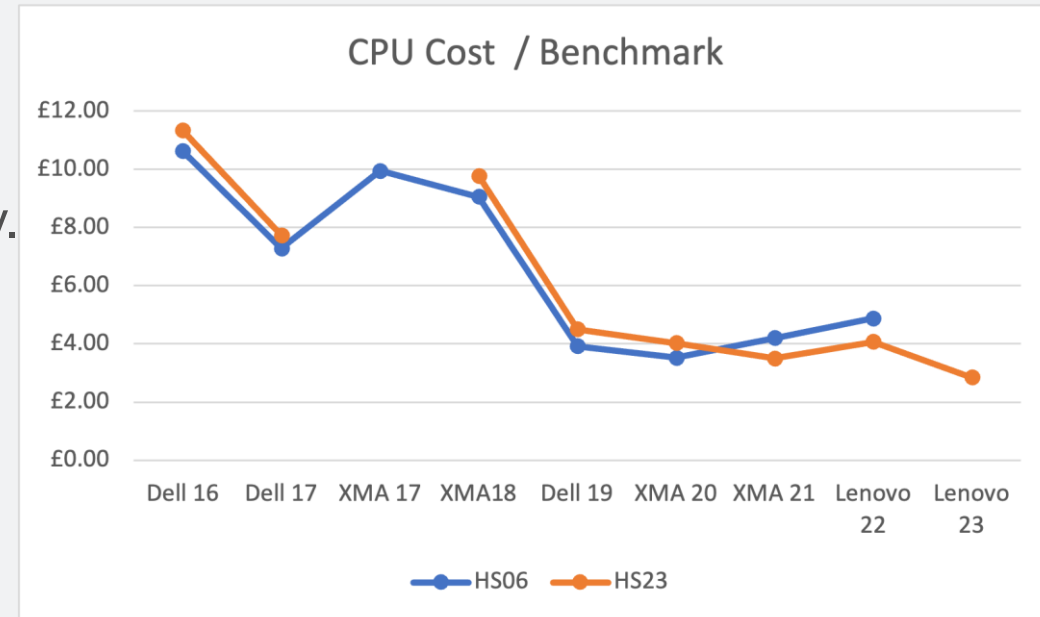
Ceph release schedule

- Our last major Ceph upgrade was in May 2020!
- Currently we are on Nautilus.
- Double upgrade to Pacific by September 2024.
- Double upgrade to Reef for April 2025.
- In GridPP7 want to aim to be looking to upgrade to latest release once it has been out for 6+ months.



CPU procurement

- This year SCD did a joint CPU procurement for JASMIN, SCARF, Tier-1 and the STFC Cloud.
 - Compute, Memory and OS Disk were identical, differences in networking and local storage.
- 407 servers in total, 36 for Tier-1, with:
 - Dual AMD EPYC 9654
 - ~ 6000 HS23 each).
 - 24 x 64GB memory (= 1.5TiB memory).
 - JASMIN purchased some with 6TiB memory.
 - 480GB SSD OS drive.
 - Mellanox ConnectX-6 25Gb/s NICs.
 - Bluefield version for the Cloud

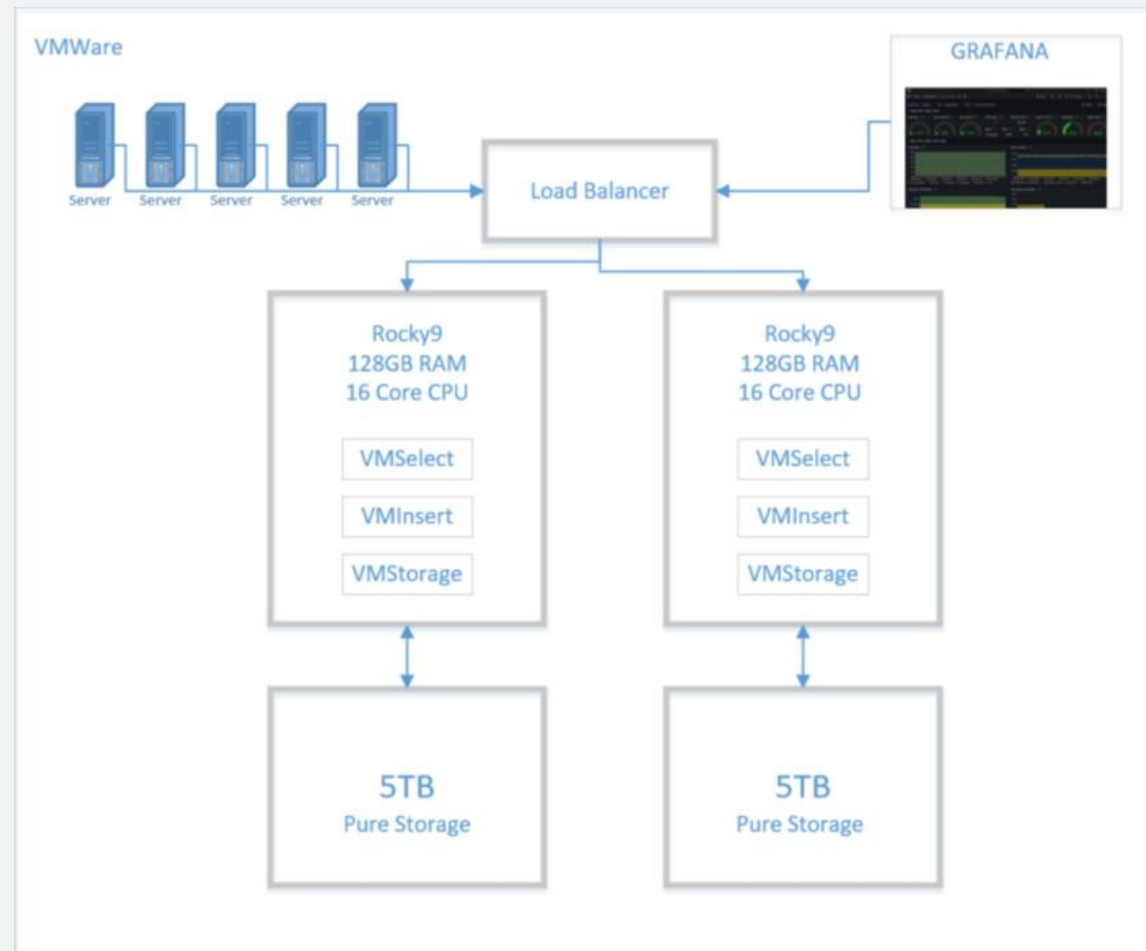


Disk and other procurement

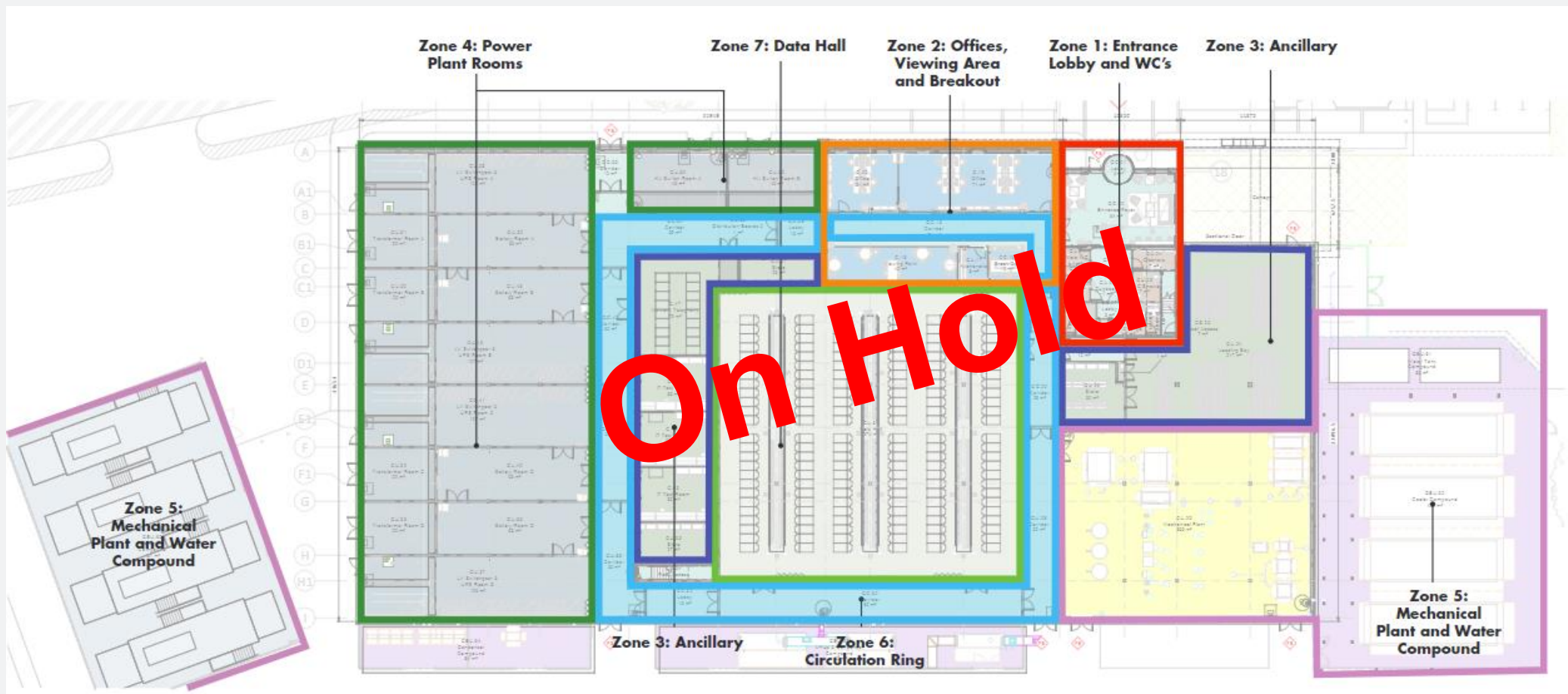
- We purchased 32 Dell Storage nodes for Echo T1 storage
 - 28 x 22TB drives.
 - 2 x 16C/32T CPUs
 - 480GB SSD OS Disk
 - 256GB memory
 - 25Gb/s Mellanox NIC.
 - Installed, almost ready for handover to Storage Team
- Echo gateways
 - 4 with 25Gb/s NICs
 - 1 with 100Gb/s NICs

Victoria Metrics

- We currently use InfluxDB for our time series monitoring.
 - Very old version – Security is OK because not publicly accessible.
- We are running a (joint) project to replace it.
 - New service will be called Timon = Time MONitor
- We currently plan to move to Victoria Metrics.
 - I am negotiating to get the Enterprise version for a 3 year contract at ~£20k per year for 0.5 million data points per second ingest.



RCC – R130



While we wait... New DC Update

- Converting one of the operations areas in the existing DC
 - Progressing...
 - 16 x 750mm wide racks, water cooled doors, rack power depends on size of coolers (~37kW rack)
 - Services, power etc., - top feed
 - No UPS, compute only





Science and
Technology
Facilities Council

Scientific Computing

A decorative graphic consisting of numerous thin, light blue lines that form a complex, abstract pattern of overlapping shapes and lines, primarily concentrated on the left side of the slide and extending towards the center.

Questions?