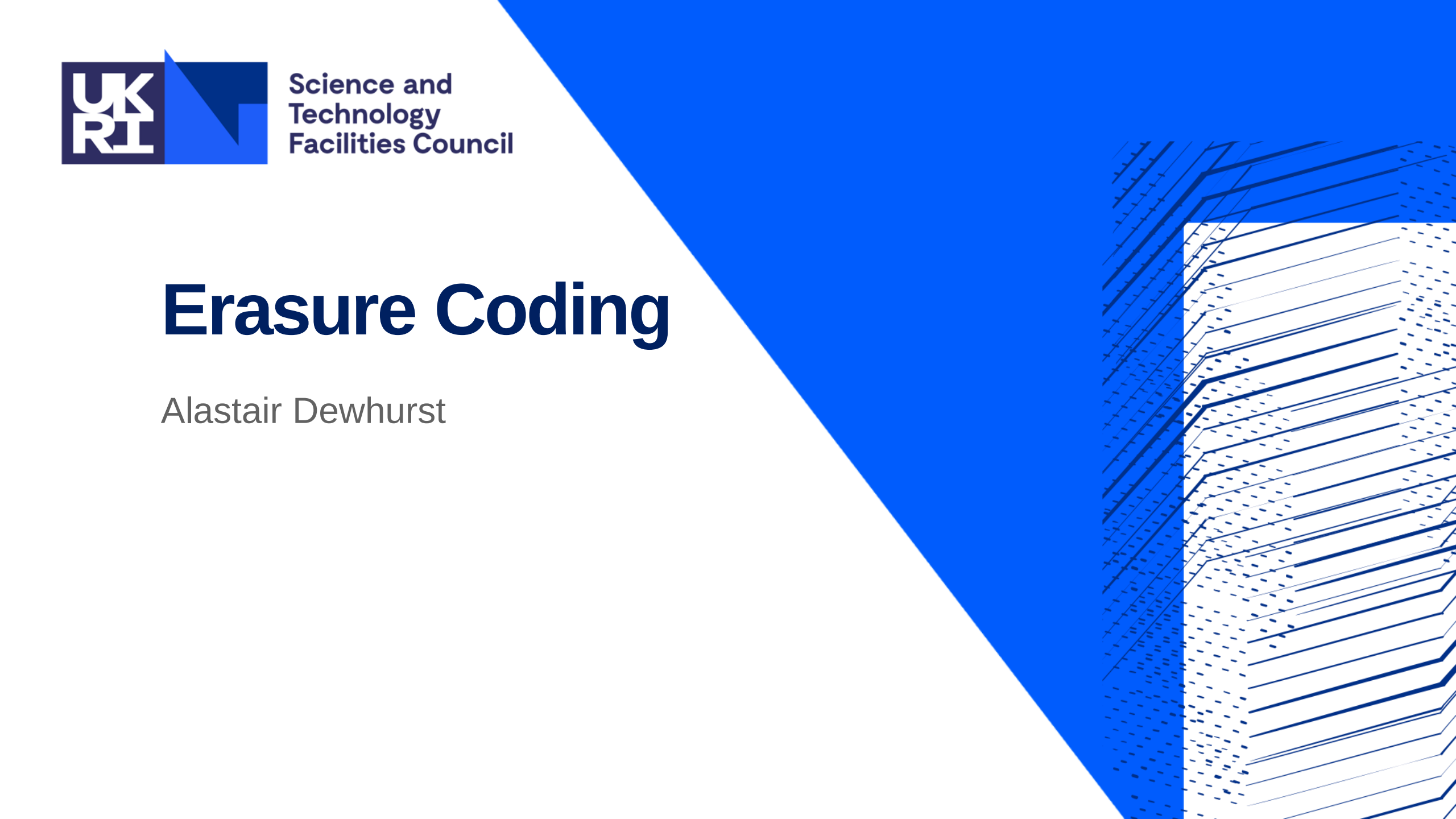
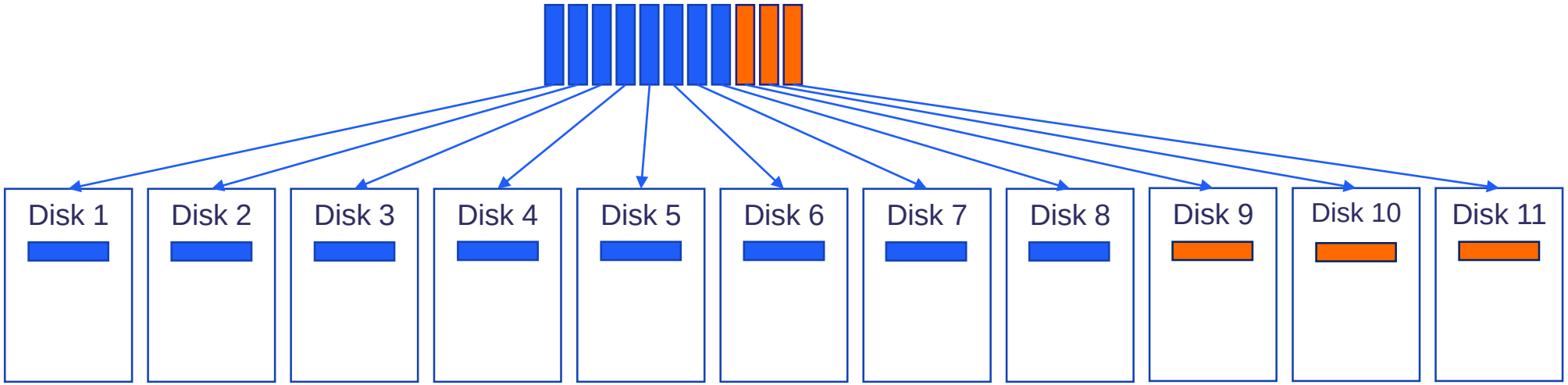


Erasure Coding

Alastair Dewhurst

What do I mean by Erasure Coding?

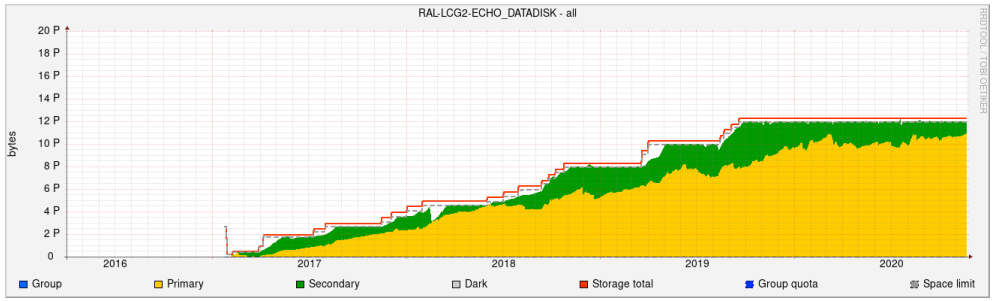
- Erasure Coding takes a file and splits into k chunks.
- It generates m additional chunks such that the original file can be reconstructed from any k out of the k+m chunks.
- Each of the chunks are stored on different storage nodes.

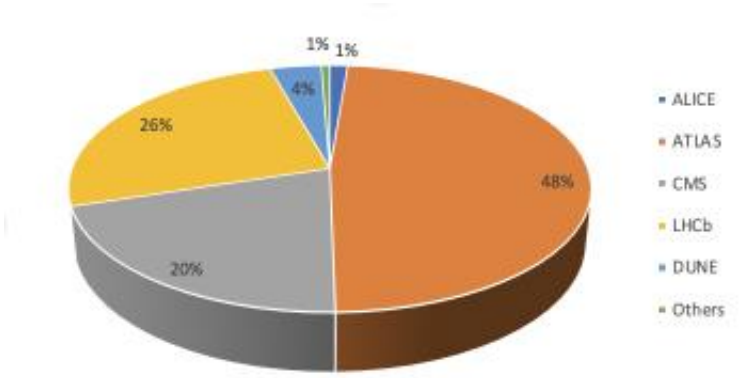


Erasure Coding at RAL

- The RAL Tier-1 provides disk storage to ALICE, ATLAS, CMS, LHCb and Dune as well as other small communities.
- All the data is stored on an Erasure Coded Ceph cluster.
  - Entered production in 2017 with 40 servers providing 5PB usable storage.
  - Now provides 34PB usable storage across ~6000 disk in 240 storage nodes.
  - Hardware has been purchased to increase capacity to 51PB usable.
- Single data loss incident (so far) in 2017 caused by “human error”.
  - In-experienced sysadmin taking wrong action when dealing with a bug.

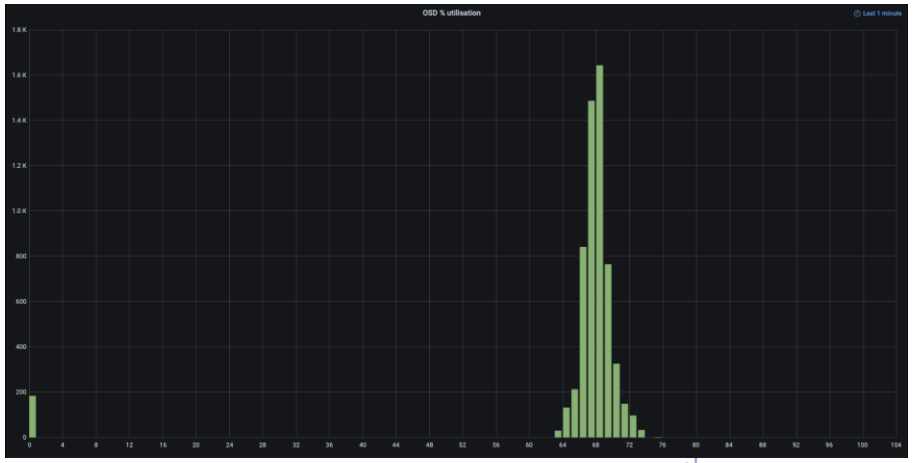
ATLAS usage of Echo





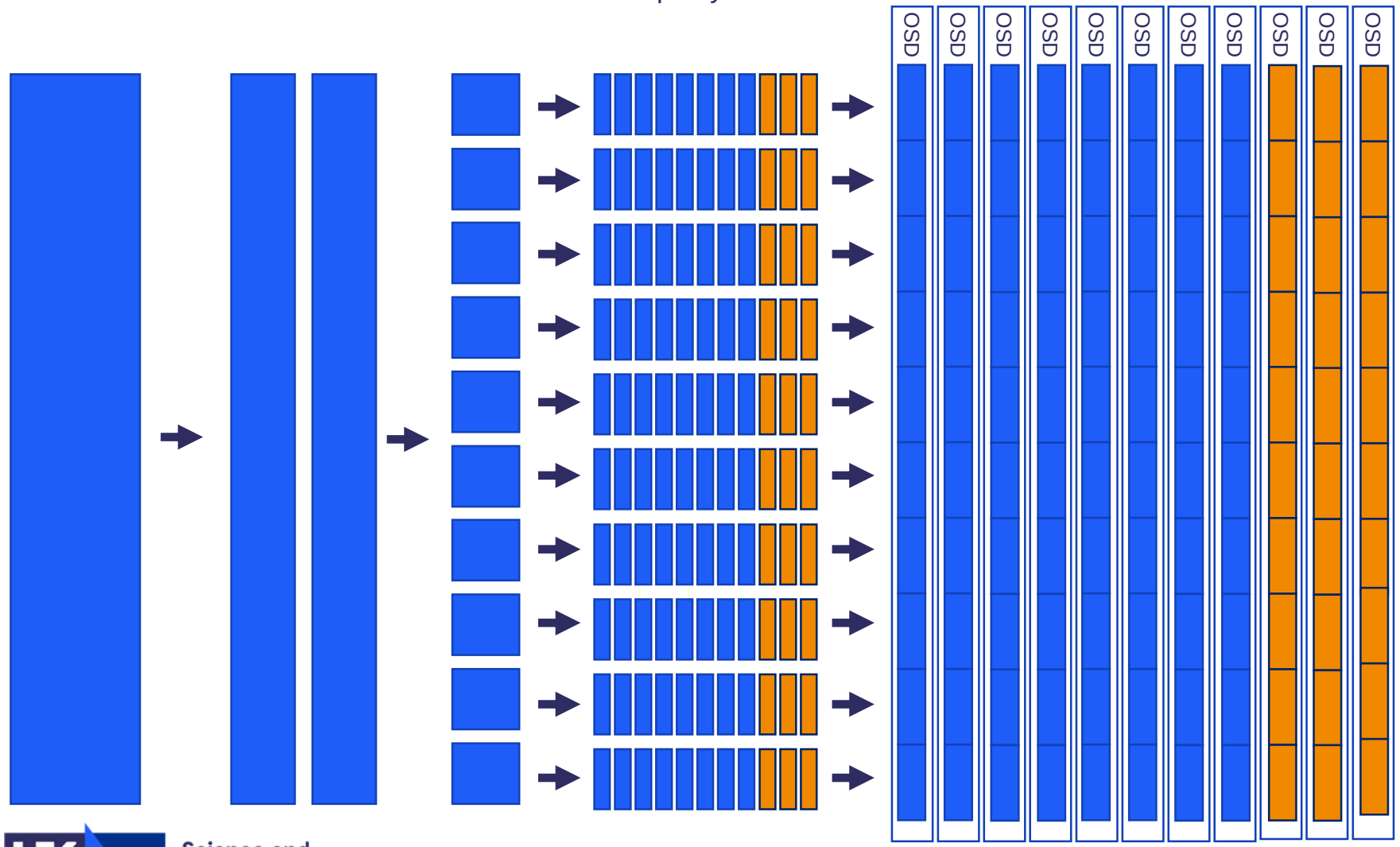
Ceph

- Ceph is not the only storage service to offer Erasure Coding.
- Erasure Coding your data allows it to survive more failure scenarios than with hardware RAID or replication.
  - It doesn't solve your problem.
- Ceph has self healing / automatic data rebalancing features that will restore the data elsewhere after a hardware failure.
- Ceph automatically scrubs disks and has internal check summing to actively look for potential data loss.
- Replacement of failed hardware is therefore separated from maintaining data resilience.
  - Carried out by Students / Apprentices.
  - Can be done months after the event.



Writing a File

File → Objects → Stripes → Stripe chunks + parity calcs → On-disk objects



The on-disk data objects are made up of 512 byte continuous ranges of the object with 4096 byte gaps

The subdivision, parity calculations and distribution of chunks are handled by the primary OSD after it receives the object from the client.

As long as all data OSDs are reachable, the parity shard OSDs are not contacted for reads

Performance

- When reading a file the data is sent from the “primary” disk in each Stripe.
  - Additional latency compared to replication.
  - For sparse reads, smallest amount is a stripe chunk which is 32kb.
- The way data is split up and how it is written to the hardware is all configurable.
  - RAL has gone for high throughput over latency.
- In general the performance of Echo is similar to other Tier-1s.
  - There are some current known problems but, these are not caused by any limitations due to Erasure Coding. Tom Byrne will go into more detail in his talk on Monday[1].

[1] https://indico.cern.ch/event/941278/contributions/4088015/

Operational experience

- Most common failure is a bad sector on a drive.
  - Daily occurrence.
- Complete disk failures are the next most common type of failure.
  - 1 - 2 a week. ~1.5% failure rate.
  - We have had a bad batch of drives. ~200 failures in a year out of 2200 total. ~10% failure rate.
- An entire node needing an intervention ~once a month.
  - Being able to easily remove hardware allows regular rolling upgrades, which will provide better reliability.
- As a site you don't want to be operating with no resilience. i.e. you want to be able to survive another failure.
  - M = 3 has double the comfort zone compared to M = 2 (or RAID6)
- We have comfortably managed failures that would have resulted in significant data loss with hardware RAID.

Cost?

- First order:
  - EC 8 + 3 means 72.7% usable space.
  - 2 x Replication means 50% usable space.

  } Need to purchase 45% more capacity with Replication
- Erasure Coding has higher CPU and Memory requirements compared to Replication. Assume:
  - 2 x CPU
  - 1.5 x Memory

  } Adds 3 - 5% to cost of hardware for Erasure Coding
- Assume Erasure Coding requires larger overhead ~5%
- Upfront costs for same amount of usable storage with Replication ~25% more than Erasure Coding.
- Power cost over 5 years ~40% upfront cost.
  - Additional ~20% cost over the lifetime of the hardware for Replication .
- For RAL, Total Cost of ownership: Erasure Coding is about 70% the cost of Replication.

Summary

- It does take significant effort to get large (k+m > 10) Erasure Coding algorithms to work well.
  - It is achievable as a community.

- Erasure Coding can provide:
  - Lower upfront and ongoing costs compared to replication.
  - Smaller physical presence in data centres (greener, more sustainable)
  - More reliable service (run hardware longer)
  - Enables site admins to manage more hardware.

- Don’t use Erasure Coding to solve LHC Run 2 problems.
  - Its not “just” a replacement for hardware RAID.

Questions?