Input from Facilities:
UK

Alastair Dewhurst
Introduction

• The current project to fund LHC computing within the UK is known as GridPP5.

• On Wednesday 6th June the GridPP project management board will begin discussing proposal for GridPP6.
  • This should fund LHC computing in the UK from 2020 to 2024.
  • What I describe here is therefore subject to change.

• I believe that GridPP will need to appeal to a larger number of users in order to maintain the same level of service for our existing users.
  • These new users are likely to be significantly less effort to spend on their computing operations.
IRIS

• Initiative to bring STFC computing interests together
  • Formed bottom up by the science communities and compute providers

• Association of peer interests (PPAN + National Facilities + Friends)
  • Particle Physics: LHC + other PP experiments
  • DiRAC
  • National Facilities: Diamond Light Source, ISIS
  • Astro: LOFAR, LSST, EUCLID, SKA, ....
  • Astro-particle: LZ, Advanced-LIGO
  • STFC Scientific Computing Dept (SCD)
  • Nuclear
  • CCFE (Culham Fusion)
  • ALC/GridPP

• It is not a project seeking to find users - it is users wanting to work together
  • Avoid duplication, share experience, pool resources,...

• Make case for investment in e-Infrastructure commensurate with STFC investments in scientists, facilities, instruments and experiments
IRIS

- Initiative to bring STFC computing interests together
  - Formed bottom up by the science communities and compute providers
  - Association of peer interests (PPAN + National Facilities + Friends)
    - Particle Physics: LHC
    - DiRAC
    - National Facilities: Diamond Light Source, ISIS
    - Astro: LOFAR, LSST, EUCLID, SKA, ...
    - Astro - particle: LZ, Advanced - LIGO
  - STFC Scientific Computing Dept (SCD)
  - Nuclear
    - CCFE (Culham Fusion)
  - ALC/GridPP

- It is not a project seeking to find users - it is users wanting to work together
- Avoid duplication, share experience, pool resources, ...
- Make case for investment in e-Infrastructure commensurate with STFC investments in scientists, facilities, instruments and experiments

IRIS has already secured £1.5 million of capital to UK sites: RAL, Manchester, Edinburgh
This is being used opportunistically by LHC experiments.

Alastair Dewhurst, 4th June 2018
Tier-1 Tape

- Castor has worked very well for us as a Tape system.
  - Due to CERN running Castor, experiments know how to use it.
  - Due to generous provisioning performance never been an issue either.

- Currently run 4 Castor instances for LHC.
  - With the retirement of the disk only Castor endpoints, it will be possible to consolidate these instances.

- By September this year, we intend to have a consolidated Castor instance

- But in the medium term Castor is being replaced by CTA at CERN....
**New Tape service**

- RAL’s Scientific Computing Department has been allocated ~£1 million to replace the Castor tape system.
  - Money must be spent by April 2019.

- LHC tape usage makes up about 50% of the tape system.
  - Many requirements from other users.

- We are currently evaluating tape solutions (e.g. HPSS).
  - Still deciding on what APIs are required.
  - I would really like to not have to provide an SRM endpoint.
    - My preferred solution would be recalls to S3 buckets with lifetime set. Much easier for users to understand.
Echo

- Echo is our new disk only storage service.
  - Production for ATLAS for over a year (5.3PB Quota).
  - Production for CMS for ~6 months (3PB Quota).
  - Migration has started for LHCb (300TB transferred so far).

- Echo has delivered
  - Erasure Coding as a replacement for hardware RAID.
  - XRootD access to Object Stores.

- Getting VOs to use cloud APIs (S3/Swift) has been harder but will keep trying.
Erasure Coding

• We don’t have enough money to have multiple replicas of the data.

• Hardware RAID6 was reaching its limit:
  • RAID card limited throughput.
  • Ever larger disks (12TB this year) mean that rebuild times are frighteningly long.

• Echo uses 8 + 3 Erasure Coding.
  • Each file is spread across 11 disk servers.
  • 3 entire machines can be out of production with no loss of data availability.
  • We have observed clear improvements in quality of service offered.
WN Caches

- The Ceph Object Store backend is very simple.
- Every WN at RAL runs an XrootD proxy Cache in a container.
  - This pulls blocks of data from the Ceph backend.
  - Lots of ongoing tweaking to improve performance but it has been working very effectively.
- I think this is actually really important as it proves that pretty much ANY type of storage can provide efficient access to data.
S3/Swift API

• Our current strategy is to provide industry standard APIs to new users.

• Integrating S3/Swift API with existing tools is going slowly.
  • Lack of WebDav support for third party transfers at other sites is a significant barrier.
  • Developing our own tools takes time.

• Looking at both DynaFed and Rucio.
DynaFed / Rucio

• Using DynaFed to provide an authentication and authorization layer for S3.
  • Recently finished tool providing LDAP authentication for DynaFed for local users.
  • Not using any of the Dynamic or Federated bits...
  • User friendly web browser access to files.
    • Also provide filesystem structure.

• Since end of April 2018, we have been running a Rucio instance for SKA to test.
  • Rucio is closest to being able to integrate S3 endpoints.
  • Considering running this as a service for multiple VOs.
Tier-2s

- Most UK Tier-2 providing storage run DPM.
  - We also have dCache, Lustre+StoRM and EOS.
- Aim has been to consolidate storage at fewer sites.
- UK universities have provided huge amounts of beyond pledge resources.
  - Want to keep allowing them to provide resources.

Plot from December 2016, shows deployed disk and CPU capacities at UK Tier-2 sites.
Caching development

• UK is small and relatively well connected.
  • Should be easy to allow remote access to resources!!

• Several different approaches:
  • “Direct Access” - WN at one site talk directly to another sites storage. E.g. UCL WN and QMUL.
  • ARC Cache - If a site has an ARC CE and a shared file system, it should be able to do what NDGF have been doing successfully for years. E.g. Durham
  • Xcache - For sites with a few hundred TB of space, we have been looking at ways we could use XCache to improve job performance and allow improved data access for users.
Conclusion

• **UK strategy to:**
  - Engage with a wider community
  - Lower barrier to access.

• **Rather Tier-1 focused talk:**
  - New Tape system on the horizon and would like to replace SRM.
  - Important things to be learned from Echo.

• **Tier-2s are evolving towards fewer endpoint.**
  - Exploring Caching solutions.