Rubin UK activity

Matt Doidge, Peter Love
GridPP48 Ambleside

Introduction

LSST - Legacy Survey of Space and Time, a 10-year mission

LSST will deliver a 500 petabyte set of images and data products

Undertaken at the Rubin Observatory Summit Facility, located in north-central Chile

Data Processed at 3 Data Facilities

    FrDF, UKDF, USDF

Corresponding to IN2P3, LANCS, SLAC

UK organisation

Active participation from Edinburgh, Lancaster, RAL.

- Edinburgh project management, operations
  - George Beckett, Pete Clark, Terry Sloan, Wenlong Yuan
- Lancaster main hardware resource, PanDA expertise, operations
  - Matt Doidge, Peter Love
- RAL hardware resource, Rucio expertise
  - Tim Noble
- JISC networking
  - Duncan Rand
- Tend to have fortnightly meetings, focussed on Rucio and PanDA commissioning
- Daily communication is via SLACK
- Project management via Confluence and JIRA

Data storage

- Camera raw data taken at observation and immediately transferred to USDF
    - 100% stored at USDF
    - 100% transferred to FrDF
    - 25% transferred to UKDF
- Master calibration products calculated at USDF
    - replicas made at both FrDF, UKDF
    - Likewise for template data science these needed for Alerts
- Data Release Processing performed at all DF
    - 25% at USDF
    - 25% at UKDF
    - 50% at FrDF
- Need 200,000 files/hr, file sizes on order 100MB raw image files, but exact mix of sizes isn't clear
- Estimated global "normal storage" 2023 6.6PB, 2024 24PB
- Data transfers are managed by Rucio using FTS
- Workflow management is via PanDA

Grid Resource

- Site requirements
  - CVMFS repo /cvmfs/sw.lsst.eu/
  - CE endpoint compatible with Harvester; HTcondorCE, ARC-CE
  - XROOTD/HTTPS storage endpoints
  - Singularity/Apptainer (prob via cvmfs)
- Services
  - Data Butler
  - Message Broker
    - These are the "elephants in the machine room" for regular-style grid operations.
    - Rubin Operations are centered around the concept of a small number of resource centres, not clouds.
    - Will likely need to concentrate on one or two UK sites for the start.

Activities

Focus is on DP0.2 which is to commission the data processing and transfer tools in preparation for Data Processing DP1

1. Hardware deployment at SLAC
    a. hosting Rucio and PanDA instances, plus other auxiliary services
2. Rucio Monitoring
    a. RAL/Edin designing a suitable system to monitor Rucio transfers
3. Job submission
    a. Lancs developing Rubin-specific wrapper and testing of sites
    b. basic jobs ok for all sites, including LANCS/RAL/QMUL/MAN
4. File transfer performance being quantified
    a. concern about smaller file efficiency

Round Up

- Strong and increasing involvement with Rubin across the UK.
- Grid sites could provide an almost ready made LSST resource (just add a Butler?).
- Experience from the Grid invaluable.
  - Discourage wheel-reinvention.
  - Encourage good practice.
  - Ready-made endpoints and infrastructure.