LSST DRP Operations

for the LSST:UK DRP Team

Special thanks to George Beckett, Tim Noble and Mathew Sims

Overview

- Last year James gave a lovely update on the Science we hope to achieve with the data obtained during the run of LSST.
  - I am not an astronomer, so this will not be an exciting trip to the stars.
- Reminder of the tasks and scales involved.
- A closer look at the components beyond what we “normally” have at a typical grid site.
  - This could, in theory, be applied to other Astronomy groups in the future if they want to utilise grid resources.
  - Or perhaps other disciplines.

Vera C. Rubin Observatory, a reminder

- Observatory in Chile, almost complete.
- Wide field of view 8.4m Simonyi Optical Telescope
- Rubin Observatory Legacy Survey of Space and Time (LSST)
  - 10 year survey starting early 2025
  - Covering 20,000 square degrees in u,g,r,i,z,y bands
  - Each field will be observed at least 800 times during the survey
  - Operations expected to produce ~20TB of data each night.
  - Nightly alert stream, ~10M alerts per night.
  - (~Annual) Data Releases
    - Catalogues, Deep and Single-visit Images
    - Served by independent Data Access Centres
    - Produced at internationally located “Data Facilities”
      - That’s Lancaster and RAL in the UK

Acknowledgement: Rubin Obs/NSF/AURA

Rubin and UKDF Status, milestones

- On the summit, telescope commissioning is continuing well.
  - Components have been shipped to the summit and being install
  - “ComCam” data set to come later this year.
    - This processing will mostly be done in the USDF in the first instance.
- Within the USDF (SLAC), rolling out the necessary infrastructure.
- Most of the big infrastructure pieces are in place
  - On site Rucio deployment in the US.
  - Some parts are still “borrowed”, e.g. using RAL’s FTS
- At RAL and Lancaster we’re in a good state of readiness
  - All/most of the “extra bits” are up and running, if not quite fully polished.
    - There’s still dev work to be done to finalise them.
    - We’ll elaborate on these “extra bits” later.
    - Sites seeded with “DP0.2” test data.

Contributions to Monitoring

Tim has made notable efforts into providing DRP operations with extensive rucio monitoring via OpenSearch.

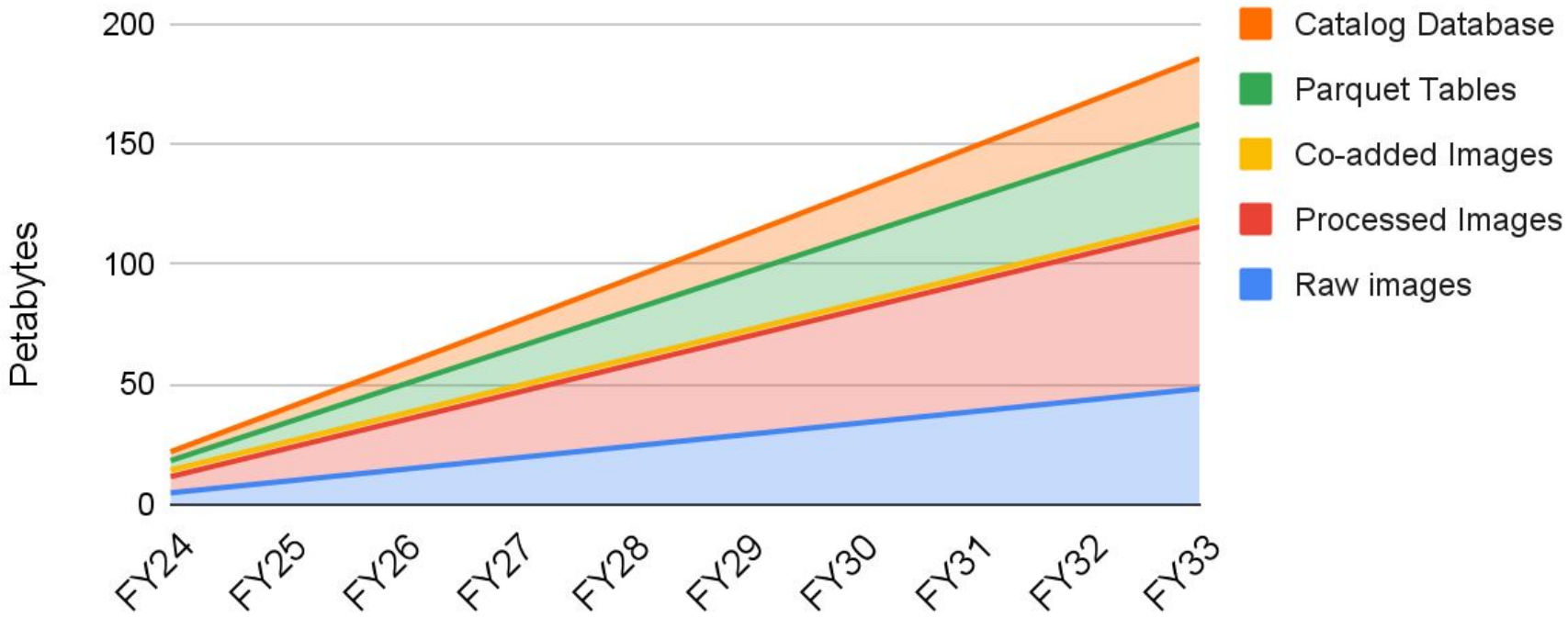
Rucio metrics are extracted from Rucio in 3 ways:

- Prometheus - Internal metrics for Service owners
- Hermes - External metrics useful for campaign managers
- Logstash - Database information for summarising site and user usage

Used in combination give a clear picture of the site usage, data flow (transfer efficiency, rates, and volume), and Rucio health.

Data Release Product “By Volume”.

To remind ourselves of the expected data volumes:



DRP Resource Projections

DRP	Preops	Survey Operations (based on commencement of survey in FY25)										Post-ops	
Capability	FY24	LOY1	LOY2	LOY3	LOY4	LOY5	LOY6	LOY7	LOY8	LOY9	LOY10	FY35	FY36
CPU (M core hrs)	6	11	21	30	40	50	63	73	83	93	103	100	50
Normal/ Object (PB)	8.0	9.0	16.0	23.0	30.0	37.0	44.0	51.0	58.0	65.0	72.0	61.4	0

As part of the UK's in-kind contributions we will provide Data Release Processing for ~25% of the survey - so roughly a quarter of the above numbers.

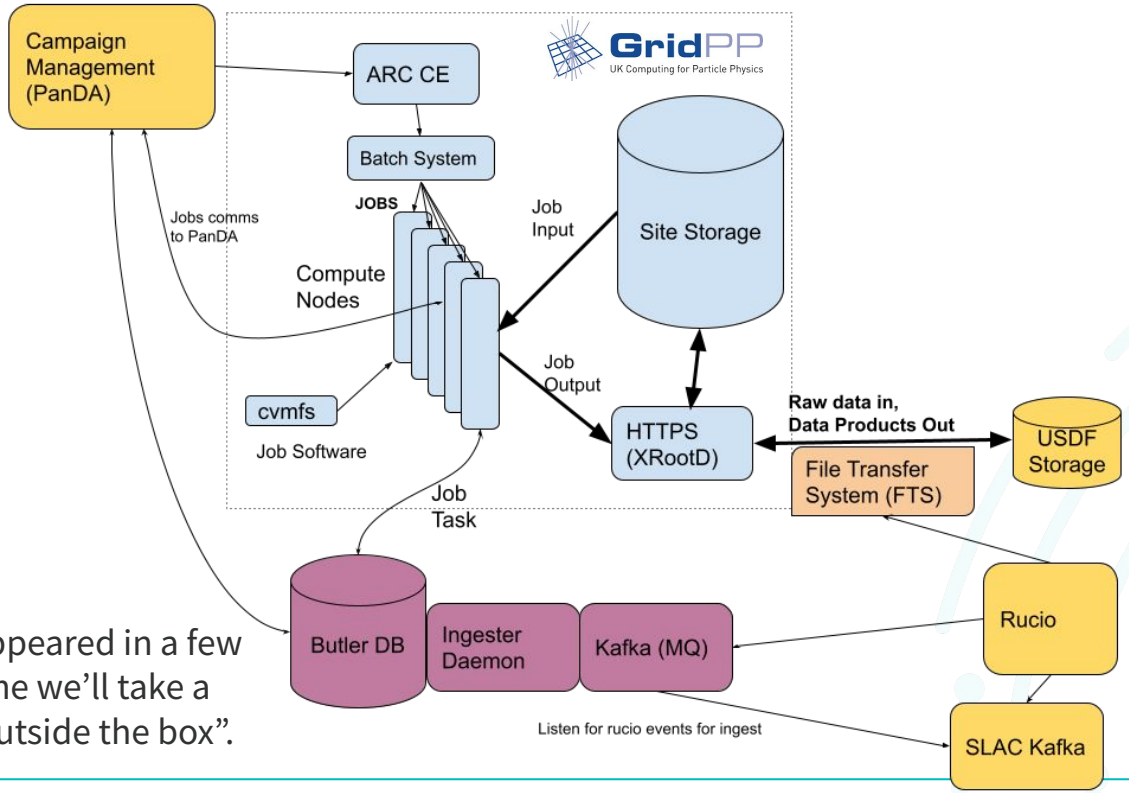
- Still some debate about how that quarter will be decided.

Each year's data release processing effort will be expected to be delivered in a similar timeframe each year, a "conveyor belt approach" to delivery.

- Each year's release will process all raw images taken by the survey to date, so processing needs will increase ~linearly year-on-year.

The DRP workflow is a good fit for "grid style" HTC - a lot of large scale serial (re-)processing over many many files.

Anatomy of a UK DRP Site



This image has appeared in a few DRP talks, this time we'll take a peek at what's "outside the box".

Not pictured: PerfSonar

LSST “Customisations”

@LANCS We had to implement a few tweaks to get LSST jobs running

- LSST run as a single user, rather than a pool
  - Extensively use shared area for configs, this provides a straightforward way to control that.
  - Looking at giving intermediate jobs write access to CEPHFS for outputs.
  - Considering running a dedicated xroot server as the LSST user
    - We considered multi-user xroot but only lsst have this requirement and the multi-user config is everyone-or-no-one.

@RAL

- LSST configs either downloaded to /tmp at job start from S3 storage using prolog script or stored in S3 and accessed when needed
- Using external Echo gateways for read and write as ‘shared’ storage between workers needed - Coordination between RAL and PanDA team needed to resolve a hidden file:// prepended by PanDA

What does the (Data) Butler do?

At its heart the Data Butler is a postgres database, but is a vital part of operations.

- Similar to a “VOBox”
- Stores information on the (meta-)data stored, as well as the tasks that need to be performed upon it.
- Vital to operations, and to maintain a backup of!
  - Kind of equivalent to the backend databases of our old SEs.
- Needs to be adequately provisioned, and well connected to the clients.
  - Found out the latter the hard way at Lancs, when we had clients trying to connect through the cluster NAT.
  - Also needs to be well connected to The Data.
- The “IngestD” is the service that connects the rucio based transfers of data with the “ingestion” of that data into the Butler.
  - Data Ingestion being an involved, non-trivial process.

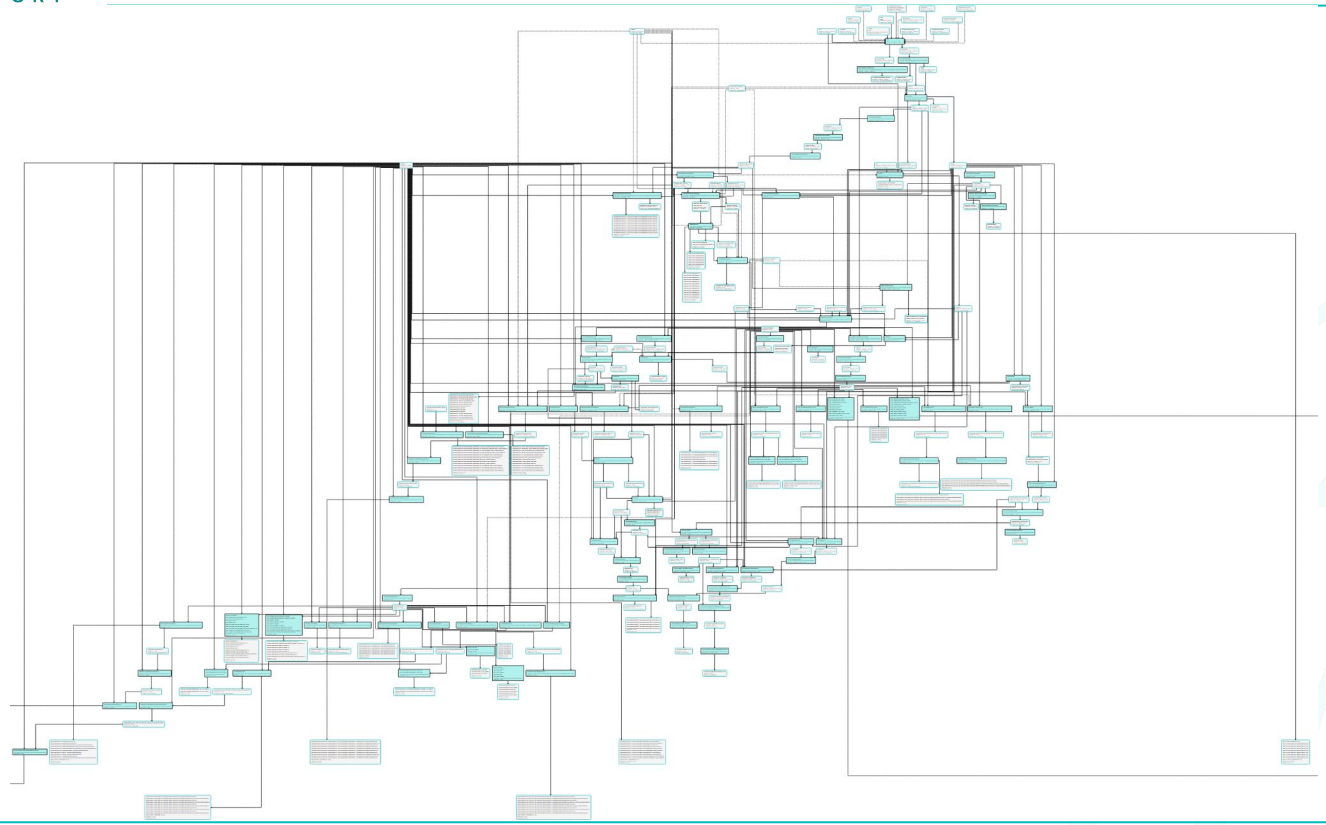
Messaging Services

- **Tl;dr: Kafka is a fancy message queue which moves metadata between DF's.**
- When data is transferred from USDF to UKDF (or FrDF), metadata must be transferred at the same time so each DF knows what data it has, and what it is.
- Rucio (Hermes) messages consumed into USDF Kafka cluster, sorted into topics by destination DF. Local MirrorMaker2 instances at Lancs., RAL, IN2P3 replicate relevant topic to a local Kafka cluster & produce to a custom daemon that imports to Butler.
- Authentication and security between Kafka clusters is a new, to us, challenge.
  - mTLS is of some benefit and (relatively) painless to configure with, e.g., eScience cert's.
  - Currently, authentication between Lancs. and USDF is not working. RAL as yet untested.
  - Only some parts of this messaging pipeline need to be externally available (e.g., MM2, but not Kafka).
- Scaling and throughput could be problematic depending on the cadence of data transfer. USDF use Strimzi on K8S, will we need to do the same?
- Will a similar pipeline operate in reverse to send DRP output back to the USDF?

DRP Cycle

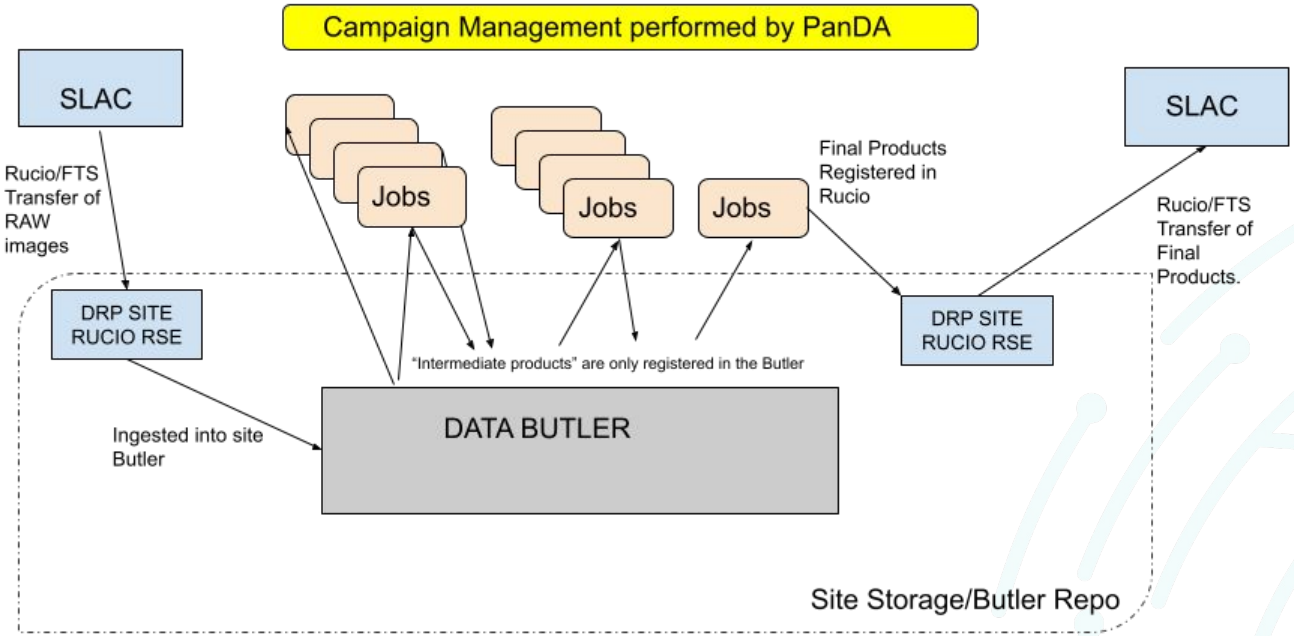
- For the DRP, the same eight(ish) processing jobs will run on all the data to produce each data release; six-monthly in year one, annually after that.
- DRP is effectively a single user, the Rubin-LSST project, with zero ‘end-user’ jobs.
- We should quickly be able to spot any issues with the code for the DRP jobs, and fix them. Limited changes to those jobs throughout the lifetime of the experiment *should* mean an efficient, largely pain-free DRP for most of the 10-years it will run.
- Each year is an independent dataset processed in isolation, so changes can be made between each DRP processing run to fix problems and implement optimisations.
- The final step in the DRP is a series of checks to ensure the output is scientifically sensible. We have capability within the UK to investigate any issues found here - effectively to act as the ‘voice of the user’, but in the same time zone!
- End-user science jobs which consume the data products DRP generates are not part of the DRP’s current planning (they come under the LSST:UK IDAC run out of Edinburgh University IfA/EPCC) though we may be asked to support them in years to come.

DRP Pipeline - it’s complicated.



Source

(Over)simplified look at what DRP sites will see.



Wrap Up

- LSST jobs require more non-optional extras to run on a grid site then I (and maybe others) naively hoped
  - Not just a case of creating a user and adding the VO details to your servers.
- Work on the summit is progressing well.
  - Sky pictures will soon be on their way.
    - At least to SLAC.
- Need to make sure that we're ready, ensuring:
  - That we have sufficiently provisioned services.
  - That all systems are well tested.
  - That we're confident in the processes.
- Keeping an eye out for opportunities
  - Lessons learned that can be applied to other groups
  - Potential utilisation of the DRP data we have access to by UK scientists
  - Experience gained dealing with the challenges.