Centre de Calcul
de l'Institut National de Physique Nucléaire
et de Physique des Particules

Some news on computing for LSST

dominique boutigny, fabio hernandez

Journées Rubin LSST-France, Paris, November 27th, 2024

doc.lsst.eu

DATA RELEASE SCHEDULE

- RTN-011 Rubin Observatory Plans for an Early Science Program

Rubin Early Data Release Scenario	Jun 2021	Jun 2022	Aug 2023	Jul 2025 - Aug 2025	#N/A	Mar 2026 - Jun 2026	Sep 2026 - Feb 2027	Sep 2027 - Feb 2028	Sep 2028 - Dec 2028	Sep 2029 - Dec 2029
	DP0.1	**DP0.2**	**DP0.3**	**DP1**	**FLD**	**DP2**	**DR1**	**DR2**	**DR3**	**DR4**
Data Product	**DC2 Simulated Sky Survey**	**Reprocessed DC2 Survey**	**Solar System PPDB Simulation**	**ComCam Data**	**#N/A**	**LSSTCam Science Validation Data**	**LSST First 6 Months Data**	**LSST Year 1 Data**	**LSST Year 2 Data**	**LSST Year 3 Data**
Raw images	☑	☑	☐	☑	☐	☑	☑	☑	☑	☑
DRP Processed Visit Images and Visit Catalogs	☑	☑	☐	☑	☑	☑	☑	☑	☑	☑
DRP Coadded Images	☑	☑	☐	☑	☐	☑	☑	☑	☑	☑
DRP Object and ForcedSource Catalogs	☑	☑	☐	☐	☐	☑	☑	☑	☑	☑
DRP Difference Images and DIASources	☐	☑	☐	☐	☐	☑	☑	☑	☑	☑
DRP ForcedSource Catalogs including DIA outputs	☐	☑	☐	☐	☐	☑	☑	☑	☑	☑
PP Processed Visit Images	☐	☐	☐	☐	☐	☐	☑	☑	☑	☑
PP Difference Images	☐	☐	☐	☐	☐	☐	☑	☑	☑	☑
PP Catalogs (DIASources, DIAObjects, DIAForcedSources)	☐	☐	☐	☐	☐	☑	☑	☑	☑	☑
PP SSP Catalogs	☐	☐	☑	☐	☐	☑	☑	☑	☑	☑
DRP SSP Catalogs	☐	☐	☐	☐	☐	☐	☑	☑	☑	☑

DP: Data Preview

DR: Data Release

DP1: ComCam Data, 5-6 months after System First Light

LSST Survey Start, 8-11 months after System First Light

DR1: LSST First 6 Months Data, 20-25 months after System First Light

next data processing milestone

SIZING OF COMPUTING FOR ANALYSIS (DESC)

- Initial estimates of resources needed at CC-IN2P3 for **science analysis** in the framework of the DESC collaboration

  study conducted by D. Boutigny with inputs from science coordinators

  science use cases included in the study: 3x2pt + cluster analysis, simulations, synthetic source injection, supernovae studies

  *goals: determining the **budget**, making **contribution statements** to the collaboration and ultimately **purchasing** and **provisioning** the equipment*

  estimates include compute (mostly CPU) and disk storage

  needs of GPU equipment acknowledged but not yet fully understood: inputs welcome

- Next step: to inform those estimates with observations from execution of prototypes of some analysis tools

  ongoing work by E. Barroso (LAPP), S. Elles (LAPP) and M. Ricci (APC)

- Budget for equipment for science analysis not yet secured

INFRASTRUCTURE

- ## Software and hardware infrastructure at CC-IN2P3 up to date

  CentOS 7 decommissioned for all production services: all user-visible services now run RedHat v9

- ## Compute capacity

  `lsst` *partition in the Slurm batch farm devoted to Rubin: this partition is used by both production campaigns and end user workloads*

  compute nodes in this partition have a hardware configuration specific for the needs of Rubin workloads, in particular in terms of RAM per CPU thread, which are higher than typical

  for details on the recommended practice for using the batch farm see the [documentation]

- ## Disk storage

  increased storage capacity of [dCache] which we use for raw data storage and data products resulting from organised production campaigns

  imminent increase of storage capacity of `/sps/lsst`*, user-visible storage on top of CephFS*

INFRASTRUCTURE (CONT.)

- ## Storage: databases

  upgrade and increase of capacity of PostgreSQL databases: back end for Butler registry databases

  two instances: one for Butler repos of organised production campaigns and another for end user's repos

  see documentation

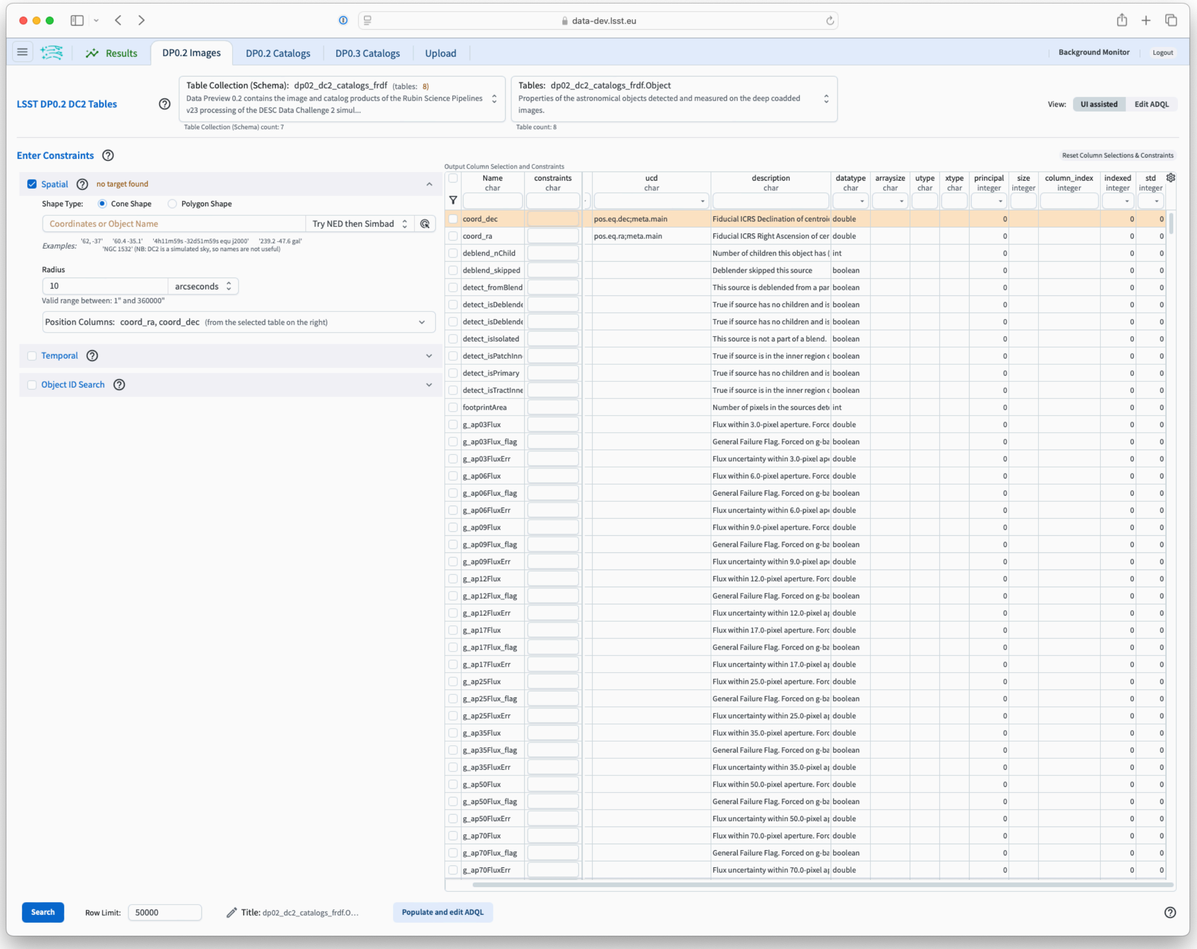
- ## Storage: monitoring

  instance of OpenSearch devoted to Rubin to enter production, as part of the ELIAS service

  to be used for collecting application-level metrics and logs of services used by Rubin (dCache, ARC+Slurm, PostgreSQL, etc.) for monitoring purposes

Rubin Science Platform

- Local instance of RSP up to date
  - *integrated with CC-IN2P3 single sign-on system*
  - *loaded with several catalogs and Data Preview 0.2 data (both images and catalogs)*
  - *runnable example notebooks available in your individual space*
- Experiment of configuration of IN2P3's instance to interact with a remote catalog database
  - *uses IVOA's TAP protocol*
  - *conducted by Gabriele Mainetti with valuable help from colleagues from LSST UK data access center at Edinburgh*
  - *results documented in [DMTN-298]*
  - *no noticeable penalty compared to a purely local instance when executing similar queries*
  - *relatively low network latency among the two sites likely helps*
- However, test conditions may not be realistic enough
  - *not many active users at remote Qserv instance competing with our test*
  - *queries used in this test may not fully reflect how human users will use the system when real data will be available*



https://data-dev.lsst.eu

Batch Production Service

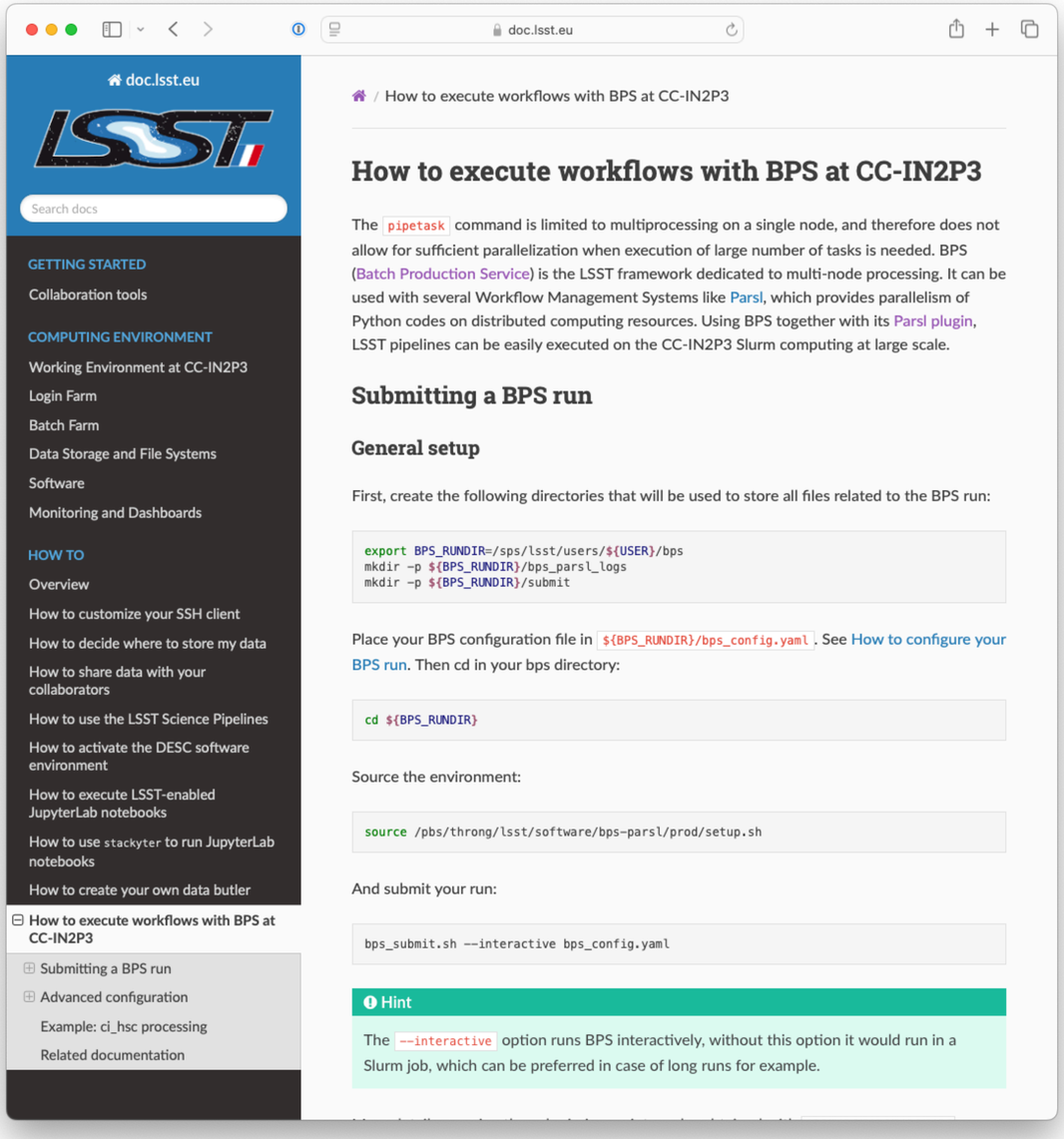
- Properly integrated CC-IN2P3's Slurm batch farm with Rubin's Batch Production Service (BPS)

  this allows for large-scale workflows to exploit the capacity of the batch farm in a well-managed fashion

  e.g. to avoid overwhelming the PostgreSQL database servers which host Butler repos

  uses Parsl for orchestration of Slurm jobs

- Details in the tutorial



NOTEBOOK SERVICE

- Implemented several software and hardware upgrades

  your notebook server now runs in a RedHat v9 container

  Python v3.11 is the new default interpreter

  JupyterLab v4.1 (with widgets) is the new notebook execution environment

  default amount of RAM for your notebook server increased to 16 GB (shared by all of your active notebooks)

  several users have more RAM: configurable on demand

  see the documentation on how to use this service

- Planned hardware upgrade

  addition of 7 nodes, each configured with 768 GB RAM and 4 Nvidia L40S GPUs (48 GB per GPU)

DASK

- Memory limits adjusted, in particular for the Dask scheduler

  you need to install the latest release of the `dask4in2p3` *Python package*

  details on how to use this service in the documentation

Organised processing campaigns

- **Principle of operations**

  input data spatially partitioned and distributed to the three data facilities

  quantum graphs centrally created and submitted for execution at the facilities

  selected intermediate data products and all final products transferred to USDF

- **Organisation**

  campaigns conducted by the Campaign Management team with support from many other teams (PanDA, middleware, data replication, etc.)

- **Tools**

  Rubin's Batch Processing System and PanDA for centralised job submission, Rucio+FTS for inter-site data movement, batch farms, storage systems and local butlers at each facility

Organised processing campaigns (cont.)

- Status

processing of two tracts of HSC public data release 2: 400 visits assigned for processing at each data facility

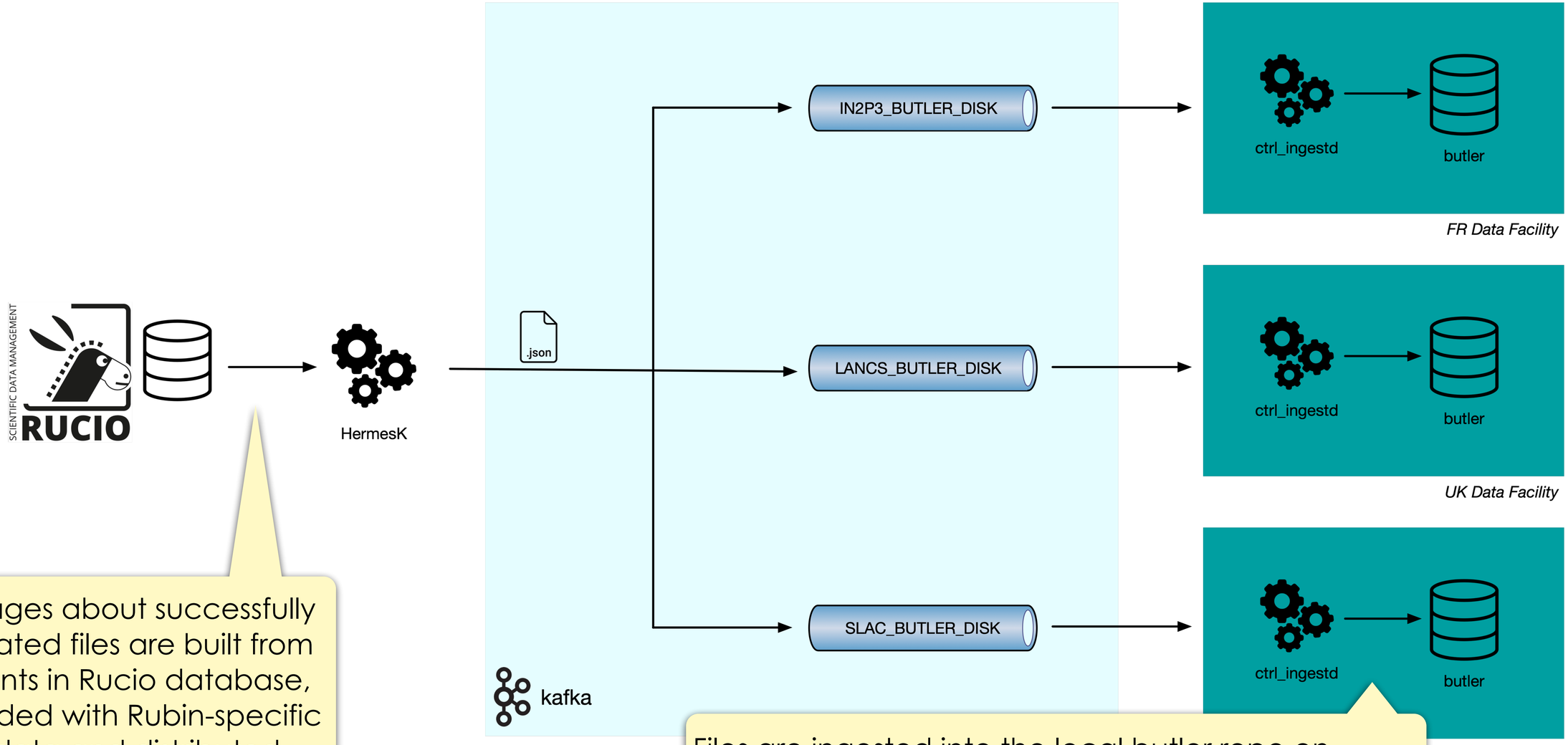
campaign tracking JIRA issue: DM-40654

inter-site replication via Rucio and FTS is working: files requiring replication are correctly registered into Rucio which triggers replication

FTS interacts with source and destination storage endpoints at the facilities to actually copy the file contents

*currently working on **reliably** ingesting the replicated data into the local butler repos at reception, at the required scale*

Organised processing campaigns (cont.)



Messages about successfully replicated files are built from contents in Rucio database, extended with Rubin-specific metadata and distributed via a multi-site Kafka cluster.

Files are ingested into the local butler repo on reception at the destination facility.

Data Replication

- We initially intended to replicate data of the electro-optical tests of the focal plane

  but replication and ingestion system was not well tested and was not ready on time (PREOPS-5511, DM-46402, DM-46654)

  these data is stored and the camera team uses USDF for processing

- We intend to import LSSTComCam data to FrDF

  for exercising the replication system and for contributing to Data Preview 1

  we have been holding this to avoid interfering with the ongoing multi-site processing campaign

- We aim to start importing LSSTCam data when appropriate as it becomes available

  this is part of our intention to archive a copy of raw data

  do we need in-dome calibration data at FrDF as soon as it can be exported?

  reminder: embargo of on-sky data during commissioning is 30 days

Data Facilities Meeting

- Data Management & System Performance joint meeting

  focused on data facilities and multi-site processing

- To be held at CC-IN2P3 Feb. 10-13, 2025

  preliminary agenda: https://indico.in2p3.fr/e/rubin-jtm

Questions & Comments