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

News on computing for LSST at FrDF

fabio hernandez, quentin le boulc'h, gabriele mainetti

Journées Rubin LSST-France, Orsay, June 11th, 2025

doc.lsst.eu

DATA RELEASE SCHEDULE

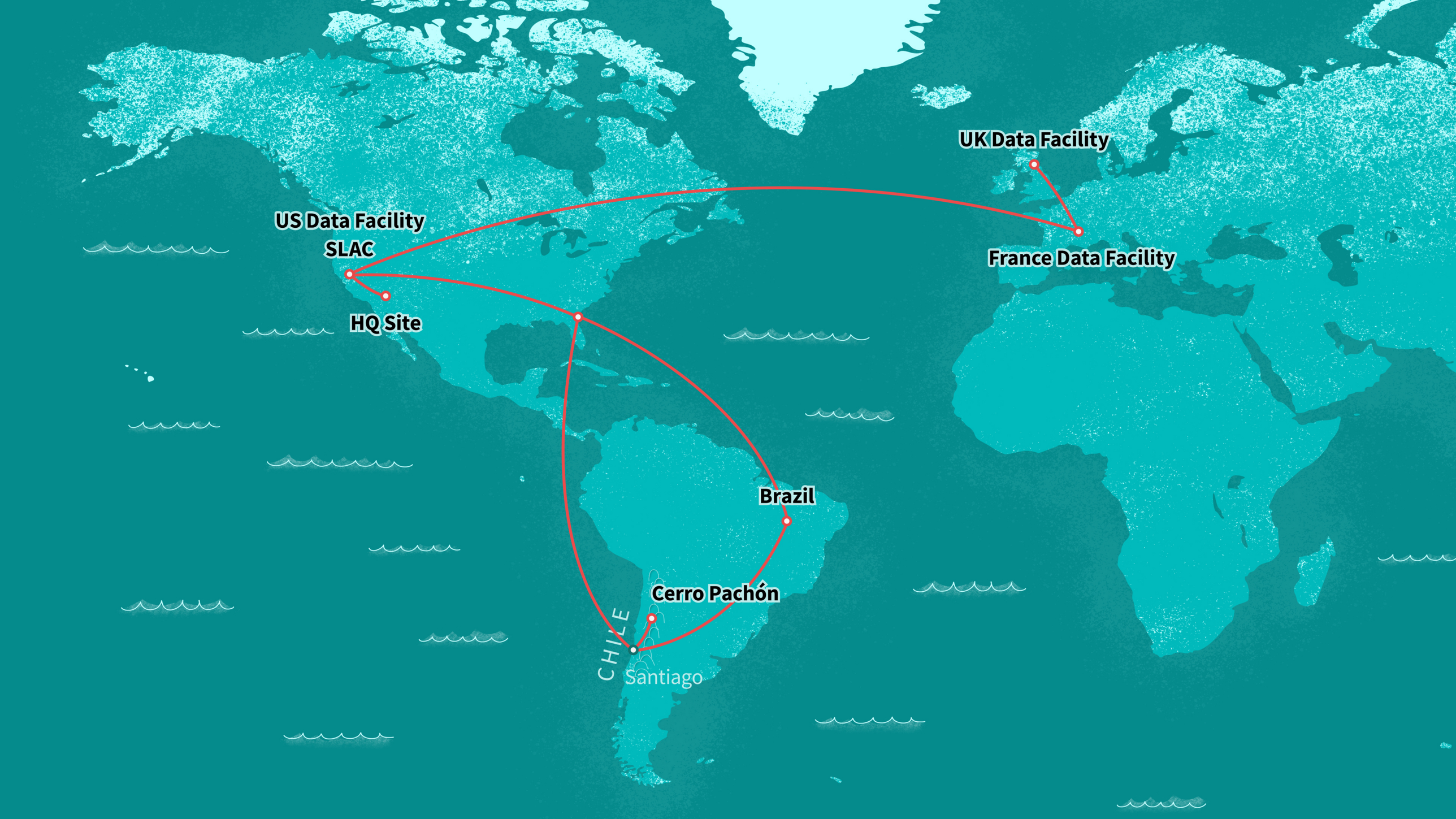
- RTN-011 "Rubin Observatory Plans for an Early Science Program"

Rubin Early Science – Data Release Scenario

Data Product	Jun 2021 **DP0.1** DC2 Simulated Sky Survey	Jun 2022 **DP0.2** Reprocessed DC2 Survey	Jun 2023 **DP0.3** Solar System PPDB Simulation	Jun 2025 – Jul 2025 **DP1** ComCam Data	Mar 2026 – May 2026 **DP2** LSSTCam Science Validation Data	Sep 2026 – Jan 2027 **DR1** LSST First 6 Months Data	Sep 2027 – Jan 2028 **DR2** LSST Year 1 Data	Sep 2028 – Nov 2028 **DR3** LSST Year 2 Data
Raw Images	●	●	–	●	●	●	●	●
DRP Processed Visit Images and Source Catalogs	●	●	–	●	●	●	●	●
DRP Coadded Images and Object Catalogs	●	●	–	●	●	●	●	●
DRP Cell-based Coadded Images and ShearObject Catalog	–	–	–	–	●	●	●	●
DRP ForcedSource Catalogs	●	●	–	●	●	●	●	●
DRP Difference Images and DIA Catalogs	–	●	–	●	●	●	●	●
DRP SSP Catalogs	–	–	●	●	●	●	●	●

Table 1: Summary of the main data products expected in each data preview and early LSST data releases. A dark teal dot denotes confirmed data products whereas a gray dot denotes data products that currently remain a stretch goal.

Data Preview 1 (DP1) to be released on June 30th, 2025
See announcement in Community
Access to data via the Rubin Science Platform at https://data.lsst.cloud


UK Data Facility
US Data Facility
SLAC
France Data Facility
HQ Site
Brazil
Cerro Pachón
CHILE
Santiago

DATA PROCESSING CAMPAIGNS

- Coordinated by the Campaign Management team

  execution at FrDF lead by Quentin Le Boulc'h

  a run every two weeks with the latest available version of the LSST Science Pipelines

  inputs: HSC and simulated DESC data

- Currently preparing a campaign to generate products similar to Data Preview 1

  LSSTComCam data

  identifying the right calibration data more difficult than initially anticipated (DM-48746)

STORAGE

- ## Client for interacting with dCache completely rewritten

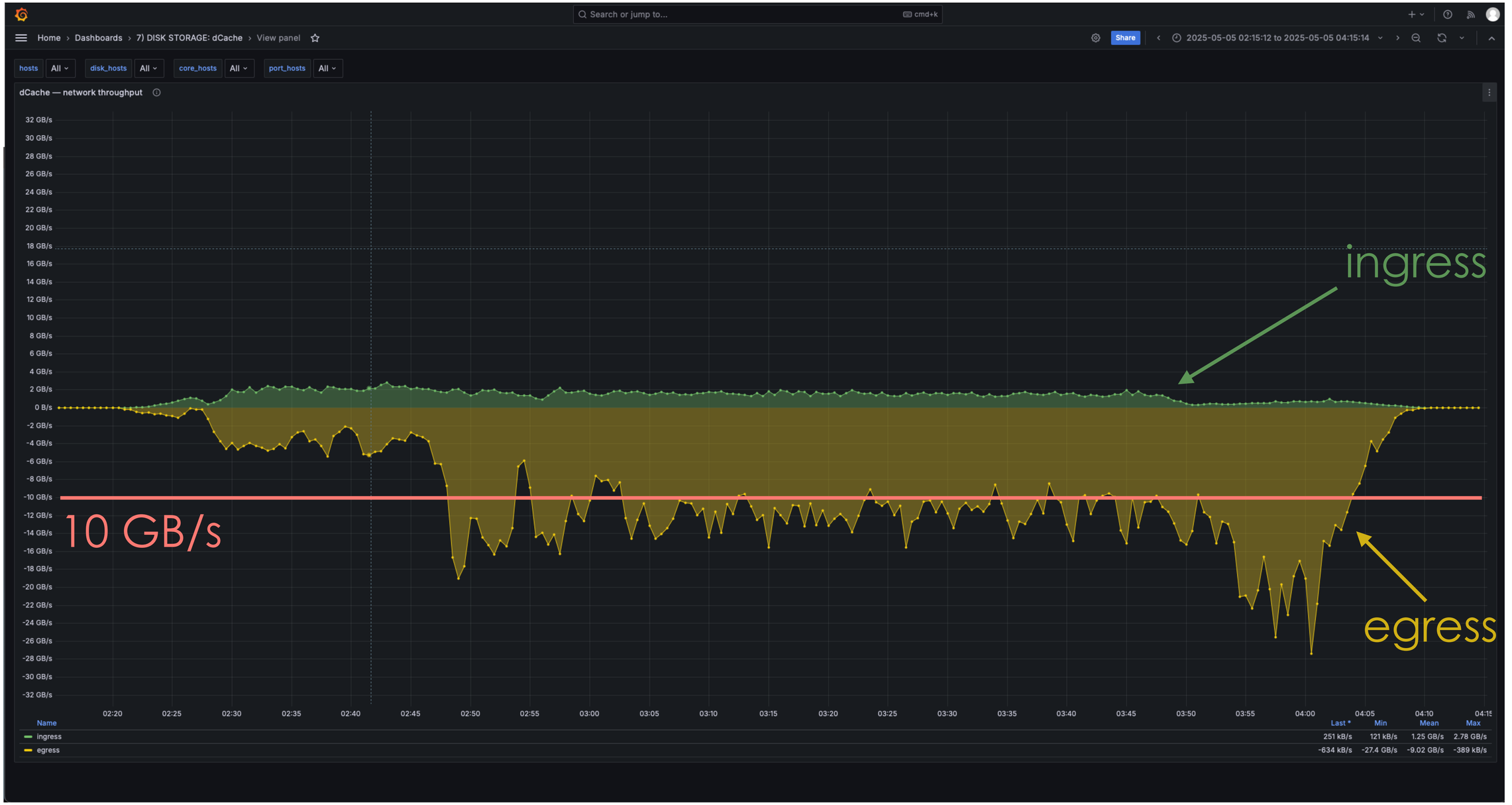
  Butler uses this client to upload and download data, to manage metadata

  takes into account the specifics of dCache behavior for improving performance and to work at the expected scale

  integrated to the LSST Science Pipelines since release `w_2025_20`

  JIRA: DM-49784

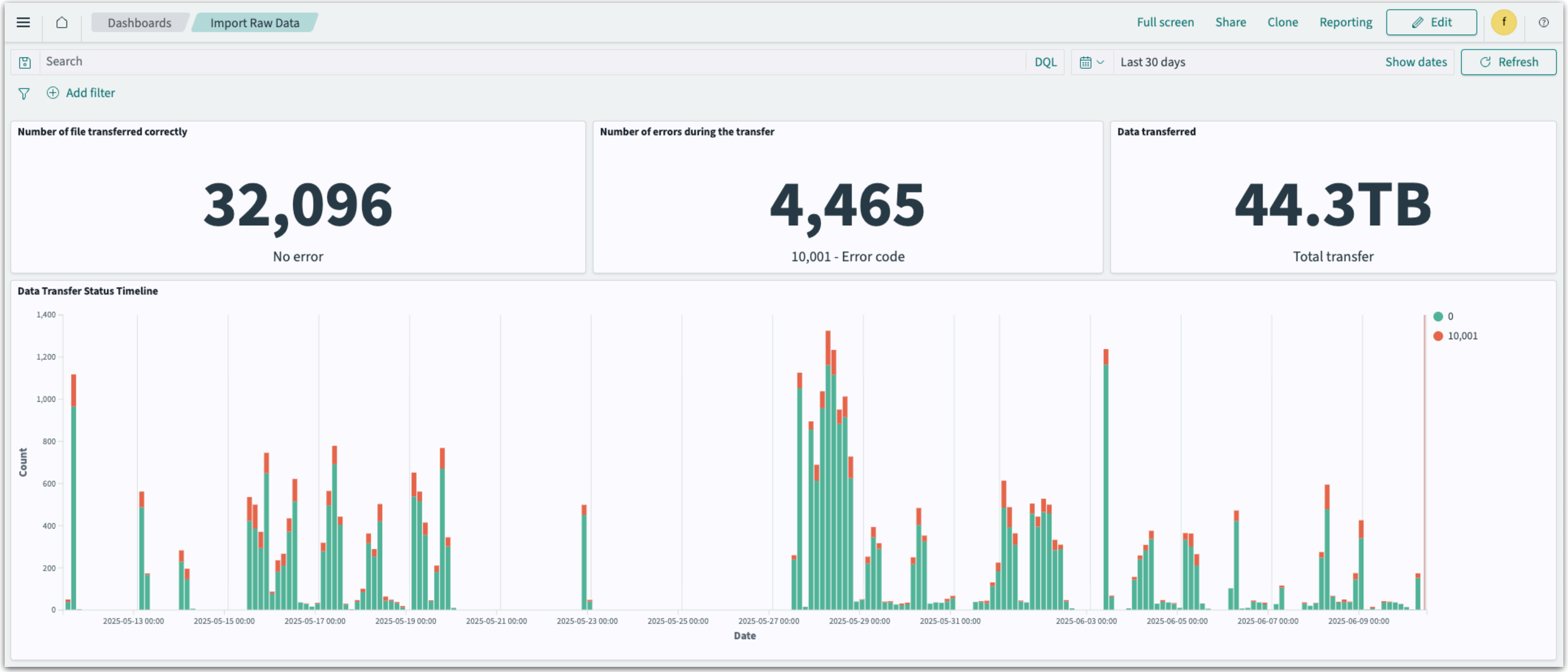
STORAGE (CONT.)



DATA REPLICATION

- Full set of LSSTComCam raw data replicated from USDF
- LSSTCam raw data replicated daily to FrDF since mid April
  - *off-sky raw exposures replicated without delay, typically by the end of the observation day*
  - *on-sky raw exposures replicated after the end of their embargo period (30 days during commissioning)*
  - *aggregated: 18k exposures, 52 TB*
- Transfer efficiency to be improved
  - *first-try transfer failure rate of approximately 15%*
  - *failure scenario understood but root cause not yet identified*
  - *details DM-50535*
- Tools for verification and reconciliation to be developed
  - *details DM-50982*

DATA REPLICATION (CONT.)



Raw exposures replicated from USDF to FrDF over the period 2025-05-13 to 2025-06-09

ARCHIVAL OF RAW DATA

- FrDF committed to save on tape a copy of the entire raw data set

  mainly for the purposes of recovery in case of disaster

- Configuration of dCache prepared and tested

  to write on tape every raw exposure

  to keep on disk only a copy of the exposures to be processed at FrDF

- Final details of the configuration of HPSS being worked out
- Automation and activation of the system will follow

Identity And Access Management System

- An evaluation instance of INDIGO IAM deployed

  Rubin data rights holders, members of the LSST France community who need access to raw data to be enrolled in this system

  IAM delivers access tokens to storage conveniently scoped to prevent accidental deletion of precious data

- Currently in the final stages of tests before opening for users to enrol

SCIENCE PLATFORM

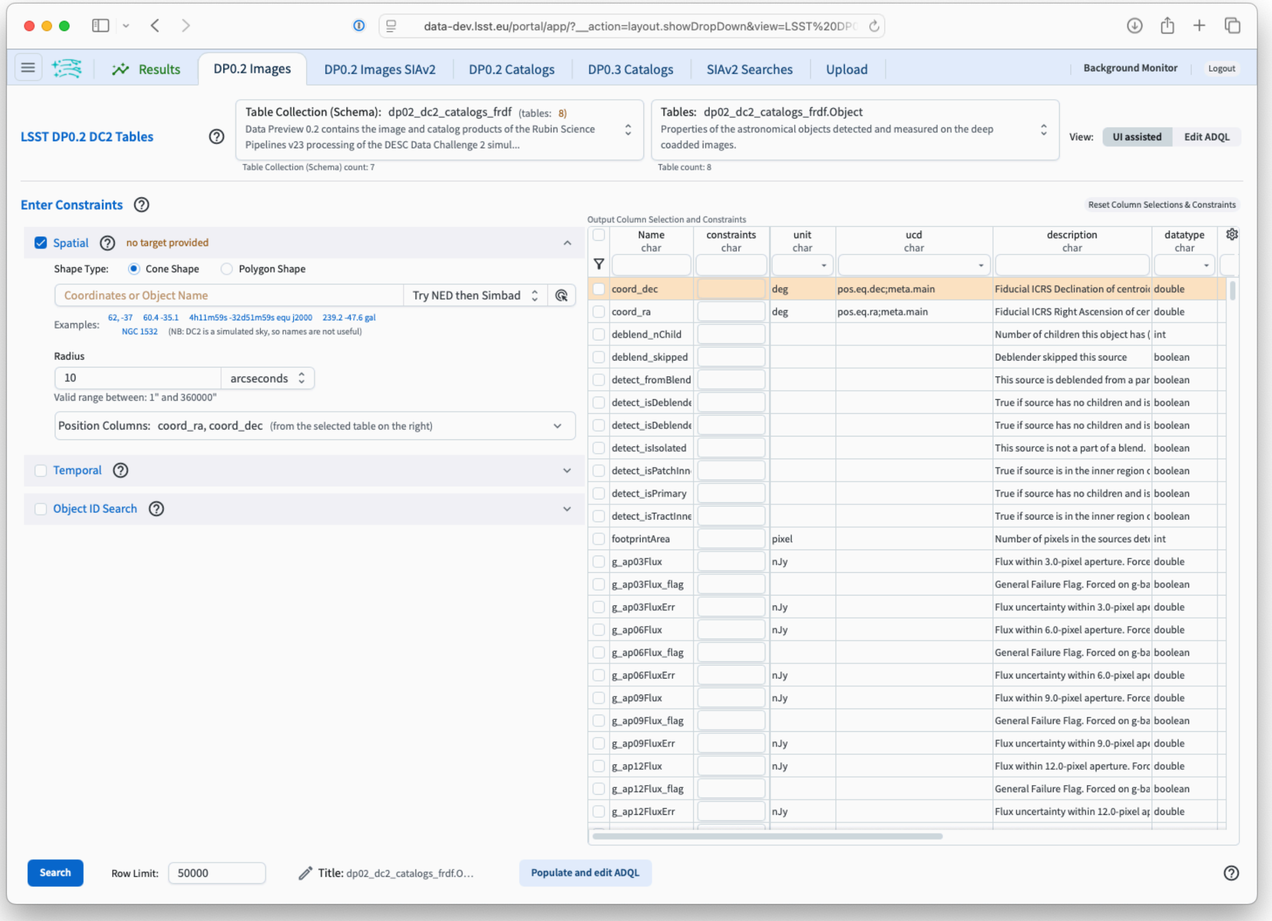
- FrDF's Rubin Science Platform up to date

  including the underlying operating system version

- Available catalogs

  `cosmoDC2_v1_1_4_image`
  `dp01_dc2_catalogs`
  `dp02_dc2_catalogs`
  `dp02_dc2_catalogs_frdf`
  `skysim5000_v1_1_1_parquet`

- DP0.2 image data set



https://data-dev.lsst.eu

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

EVENTS

- Rubin Data Facilities & Multisite Processing (Feb ’25)
  https://indico.in2p3.fr/event/34450/
- 19th International dCache Workshop (May ’25)
  https://indico.desy.de/event/48191/

Questions & Comments