

Operating the **Brazilian IDAC**: LineA's Contribution to the Rubin Ecosystem

Julia Gschwend

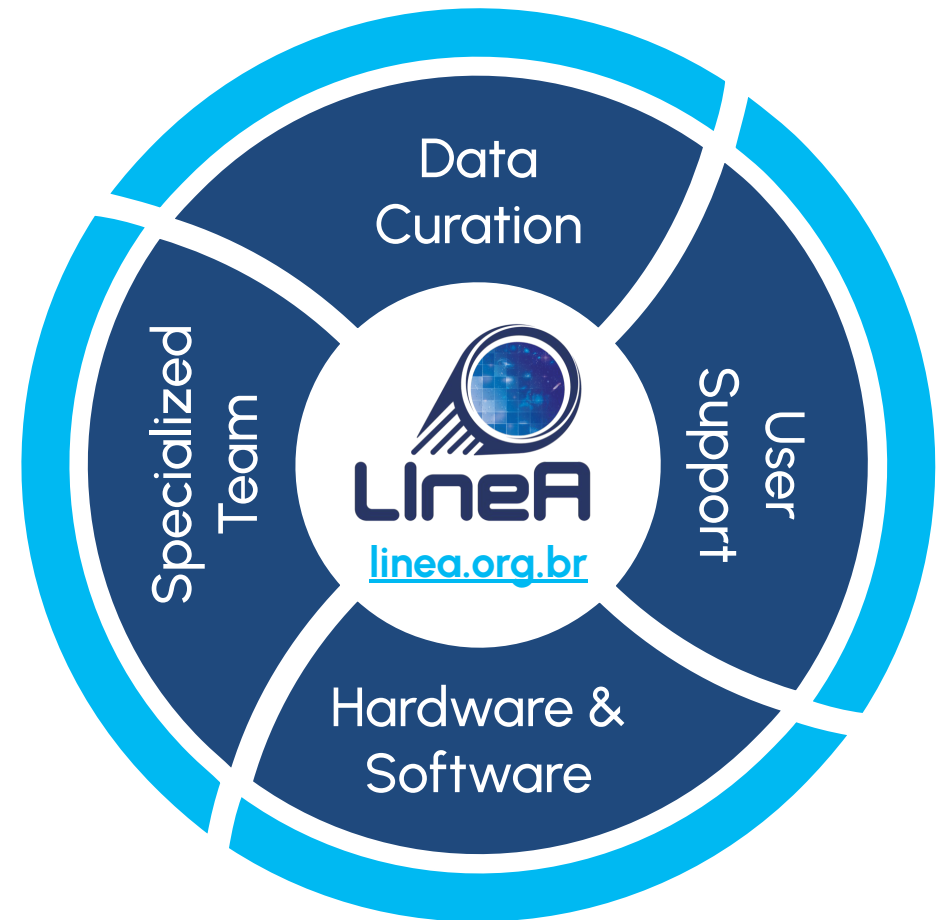
Software Lead and BRA-LIN Program Manager



LineA: Next-to-Data Science Support

- **Mission:** support data-driven astronomy
- **Paradigm Shift:** next-to-data computing
- **International integration:** enabling the participation in large surveys
- **Legacy Support:** seamlessly transitioning to the Rubin Era

DES → SDSS → DESI → LSST



BRA-LIN: Strategic Cooperation



In-kind contributions

- **Lite-IDAC**
- Software development
- Advanced data products



Optimization

- Reducing international data movement
- Processing of large-scale catalogs



Autonomy

Access to:

- proprietary and public data
- next-to-data processing power

Serving an **International** Community

120

BPG RESEARCHERS

26

INSTITUTIONS

3,000+

INTL. USERS POTENTIAL

Scaling infrastructure to handle concurrent requests from both national BPG members and the global astronomical community.

The **IDAC-BR** Engine Room (Current Status)

- **Network:** 100 Gbps network capacity deployed; RNP external link upgrade needed (RJ-Petrópolis).
- **Storage:** ~4 PB total (Lustre & NAS).
- **DB:** 150 TB Postgres with high-volume ingestion optimization; expansion to 500 TB soon.
- **Compute:** 500 Cores (JupyterHub) + ~1,200 Cores (HPC).



Software Infrastructure



Github org.: [linea-it](https://github.com/linea-it)

Apps

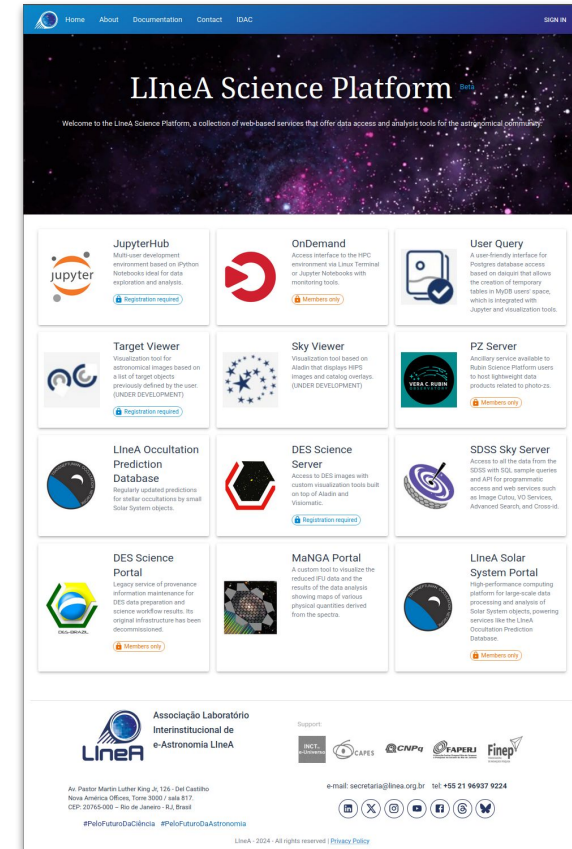
Jupyter notebook, Sky Viewer, Target Viewer, User Query, PZ Server, Solar System Portal, Occultation Prediction Database, etc

Support

Jupyter Hub, LSDB, Aladin, Daiquiri, Grafana, Nifi, TAP Service, etc.

Base

Rucio, Docker, Slurm, K8S, PostgreSQL, LDAP, etc.



Data Curation and Mirroring

The screenshot shows the Linea Data website interface. The top navigation bar includes the Linea logo, the word "Data", a search bar, and a user profile icon. The left sidebar contains a navigation menu with categories like "Data", "Home", "LSDB Surveys", "Rubin DP1", "2MASS", "DELVE DR2", "DES", "DESI", "eRASS1 Main", "Euclid Q1", "Gaia", "Pan-STARRS1", "SDSS DR7 Spectra", "WISE", "Databases", "HPC", and "Scientific Products". The main content area is titled "Rubin Data Preview 1 (DP1) Object (lite)" and contains a "Table of contents" section with "Catalog metadata". The text describes the Legacy Survey of Space and Time (LSST) and the Data Preview 1 (DP1) release. Below the text is a plot titled "Angular density of catalog object_ic" showing a distribution of objects in a galactic coordinate system with a color scale for "count / deg²".

data.linea.org.br

Current numbers:

- 12 catalogs in Postgres DB
- 11 HATS collections
- 5+ raw catalogs in Parquet (HPC)
- 30+ catalogs in PZ Server

Data Curation and Mirroring

Query interface

Database status

Using personal table space

New query job

SQL query

Upload VOTable

Job list

Run id: des

Run id: dp1_bpg

Run id: latam2025

Run id: test

Run id: tutorial

No run id

2026-03-02-17-21-20-804798

2026-02-23-13-17-28-397826

LSST DP1
lsst_dp1.object

DES DR2
des_dr2.main

Gaia DR3
gaia_dr3.source

2MASS
psc

MPC SBN
mpc_sbn.obs_sbn
mpc_sbn.mpc_orbits
mpc_sbn.comet_names
mpc_sbn.current_identifications
mpc_sbn.minor_planet_names
mpc_sbn.numbered_identifications
mpc_sbn.obscodes
mpc_sbn.primary_objects

Columns

Simbad

VizieR

Examples

statement directly in the text area below and submit your request using the button. You can use the dropdown menus to get out the database, to query external services or to access example queries.

can specify the name of the resulting database table and/or assign a run id, to group associated queries in the job list. Please choose suitable timeout for your query.

ms that are indexed and should be used whenever possible for faster filtering and improved performance.

Maximum rows per query: 5000000.

Table name	Run id	Query language	Queue
		ADQL	30 Seconds

Table Access Protocol (TAP) Query

Window TAP Authentication Registry Edit Interop Help

Select Service Use Service Resume Job Running Jobs

Metadata

Find:

Sort: Service Alphanumeric

TAP Service (2)

- des_dr2 (1)
 - des_dr2.main
- gaia_dr3 (1)
 - gaia_dr3.source
- twomass (0)

Short Name:

Title:

IVO ID:

Service URL:
https://userquery.linea.org.br/tap

Authentication:
(anonymous)

Reference URL:

Examples URL:
https://userquery.linea.org.br/tap/examples

Size:
3 schemas, 2 tables

Publisher:

Service Capabilities

Query Language: adql-2.0

Max Rows:

Uploads: unavailable

Log In/Out

ADQL Text

Query Mode: Synchronous

1

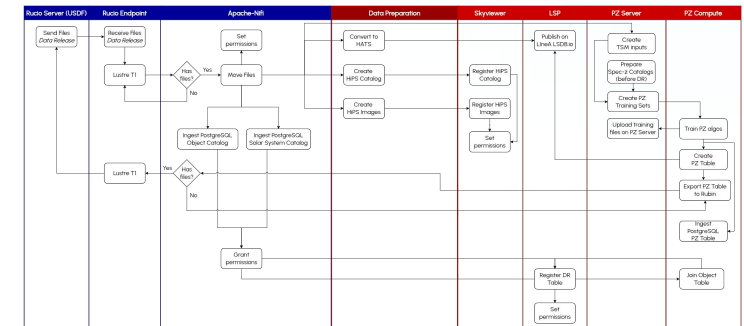
Run Query



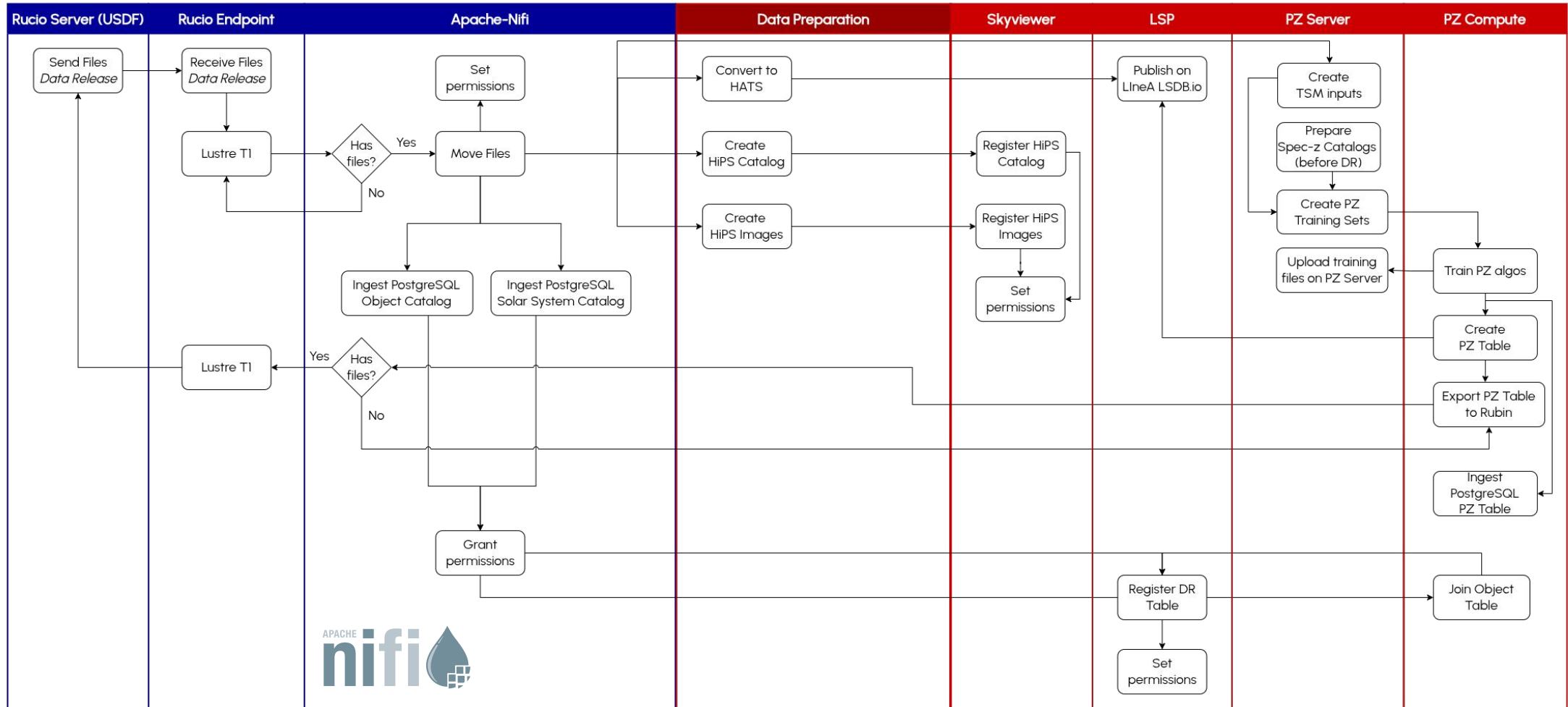
Data flow and Science Readiness

- **Rucio Transfer:** bulk dataset movement from USDF to IDAC-BR.
- **Internal data flow:** automated with Apache Nifi.
- **DB Ingestion:** high-speed ingestion into PostgreSQL.
- **Data Preparation:** converting files for LSP consumption; production of advanced data products.

spoiler alert!!! (full view on next slide)

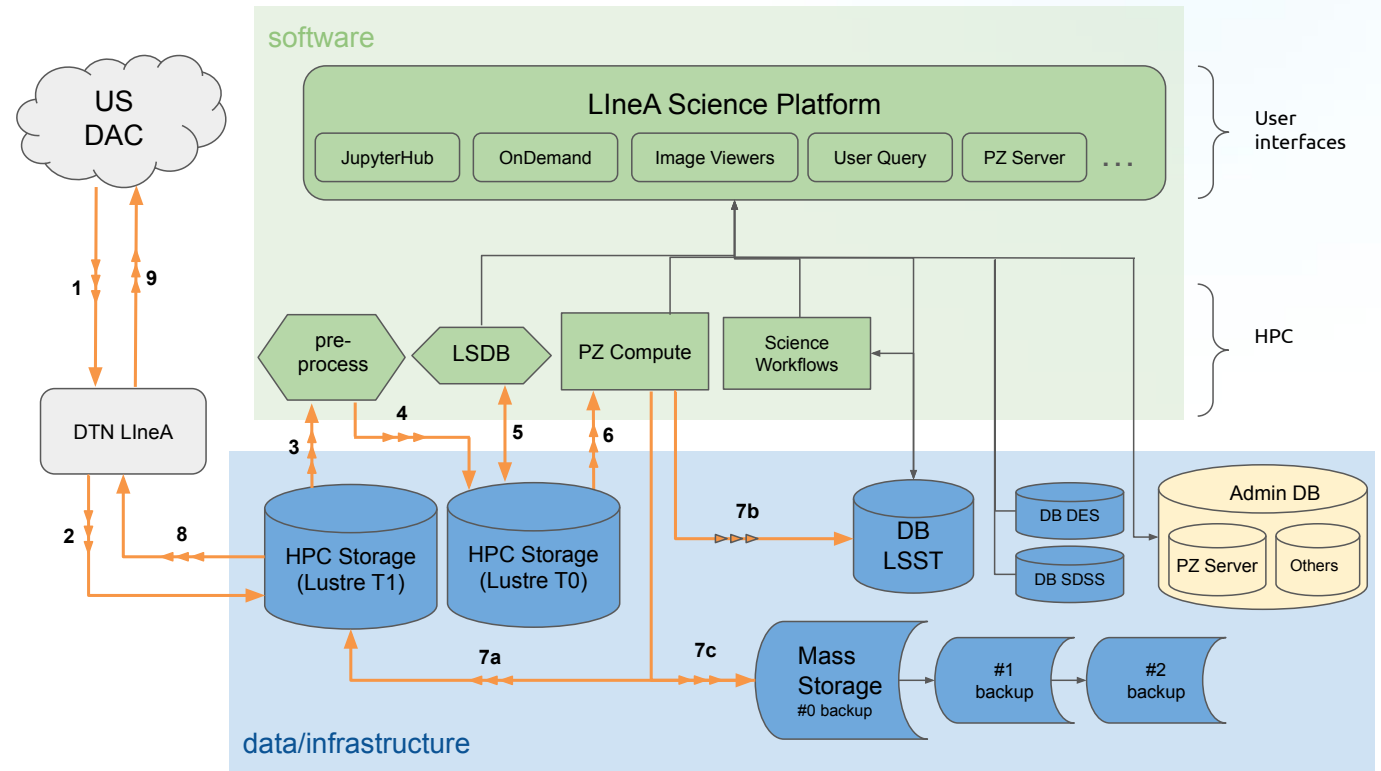


Data flow and Science Readiness



The Return Challenge

- **PZ in-kind Contributions:** software, commissioning activities and advanced data products.
- **PZ Compute:** intensive local processing workloads generating massive high-volume output products
- **Network Requirement:** bidirectional pipelines to transfer computed results **back** to USDF
- **Critical deadlines:** strict delivery windows; results must be ingested at USDF **prior** to official global data release dates



From DP1 to the DP2 Avalanche



DP1: Validation

Rucio and Ingestion validated;
Science platform functional.



DP2: Scaling

Preparing for massive
ingestion and production of
photo-zs



Future

Spine-leaf architecture
GPUs, agentic AI
New catalogs, advanced
interactive image analysis.

Credits and acknowledgements

Attending SA3CC in person



Julia Gschwend
Software Lead



Carlos Adean
Infrastructure Lead



Luiz da Costa
General Director



Rosane Melo
Secretary



Mariana Fernandes
EPO Specialist



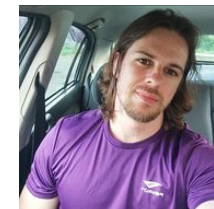
Rodrigo Bouffleur
Data Scientist



Luigi Silva
Data Scientist



Nubia Garcia
System Analyst



Eloir Troyack
System Analyst



Cristiano Singulani
Full-stack developer



Glauber Vila-Verde
Full-stack developer

Sponsors / funding support



Questions?

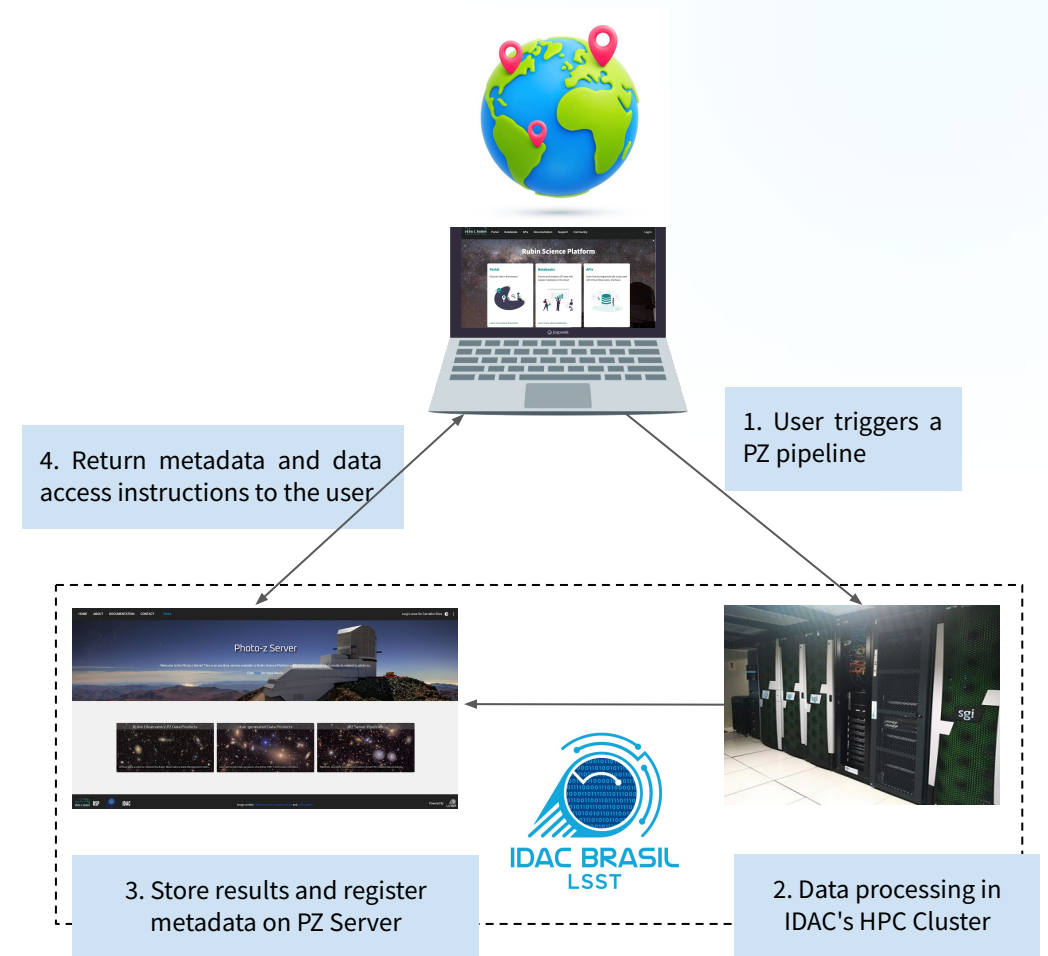
Thank you for your attention.

julia@linea.org.br

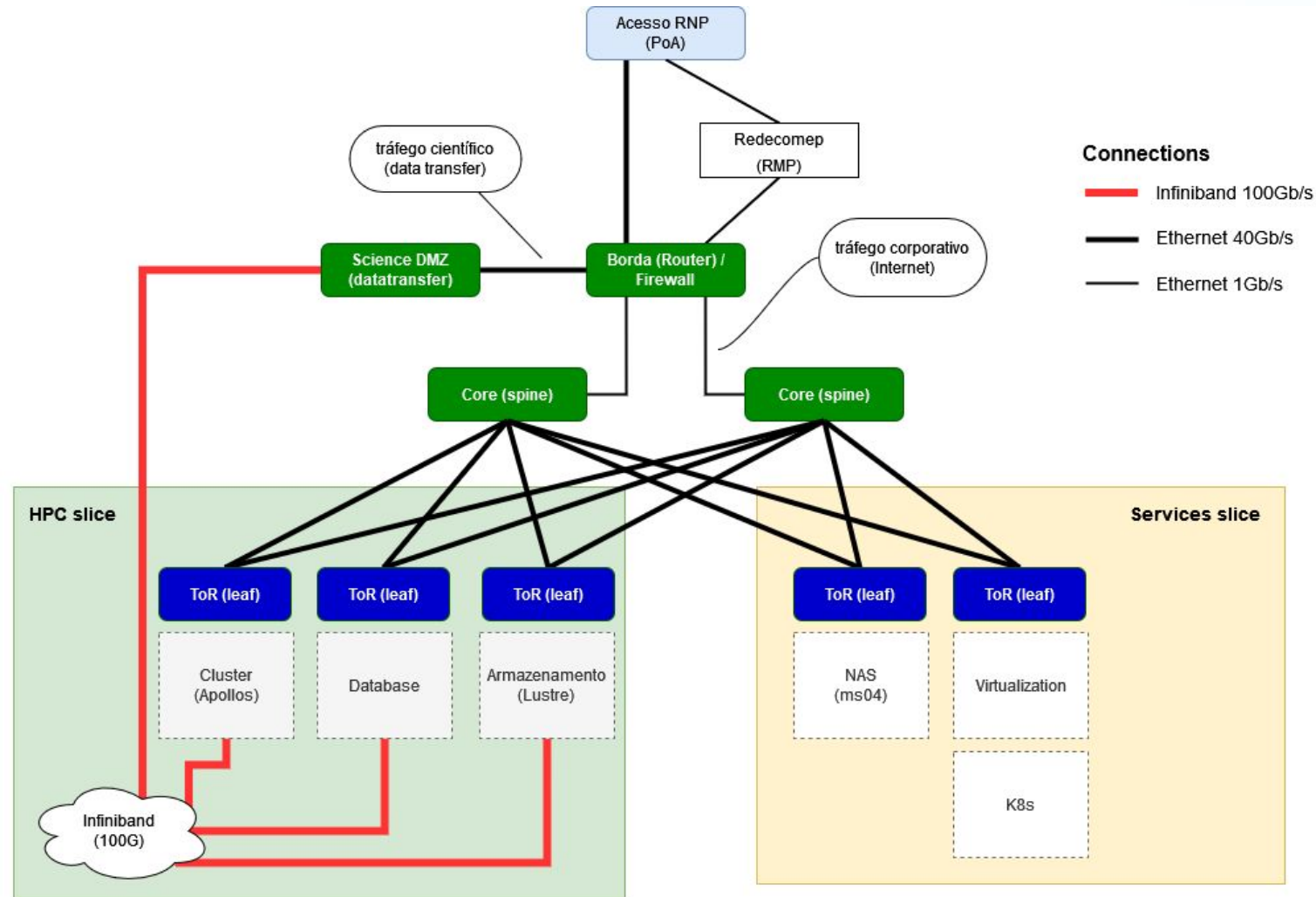
Extra slides

PZ Server API and Remote Triggering of HPC Pipelines

- **PZ Server:** web-based service to create and host lightweight data products related to photo-z
- **PZ Server Data Products:** reference redshift catalogs, training sets, training results, validation results, photo-z estimates, others
- **PZ Server Pipelines:** crossmatching with LSDB using the IDAC-BR infrastructure



Spine-leaf Architecture



Hardware Infrastructure

