

PFS Data Processing Workflow in Mitaka

Current DRP status, production workflow, and operational challenges

Speaker: Zhuoming Li

‘Ōnohi‘ula - Prime Focus Spectrograph

2026-06-17

Outline

How the Mitaka DRP workflow is organized, from raw data to final products

1. DRP status

Hardware, current pipeline versions, and environment layout

2. Data management

Raw-data sync, datastore, database, and product organization

3. 2D DRP workflow

From raw visits to coadded spectra

4. 1D DRP workflow

From coadded spectra inputs to lam1d outputs

5. Challenges

6. Summary

DRP Status

Current production environment and major software components

Production hardware

8 nodes in the LSC cluster

64 CPU cores and 450 GB RAM per node

512 cores and 3.6 TB RAM in total

Servers

Production server: LSC cluster

File server: mounted to LSC

Database server

Development server

**Test on development server before
applying changes to the production
environment**

Current pipeline versions

2D reduction pipeline: w.2026.24 (Princeton)

1D reduction pipeline: 1.18.0 (LAM)

The 2D and 1D DRP environments are kept
separate for easier management.

Data storage

File server: 48% of 500TB used

Data management

Raw data, processed products, and registry/database organization

Raw data

Synchronized from Hilo every day after the observing night ends

MD5 verification is performed after each observing run

Consumes most disk space but can be further pruned

Scripts and logs

Scripts used in the reduction process are archived on the internal Git server

Logs of each run are also archived for future review and reprocessing

2D products

Organized in Gen3 Butler datastores

Folders: store the physical files

Database: records configurations and relationships of data reduction products

1D products

Configurations and physical files are organized in folders

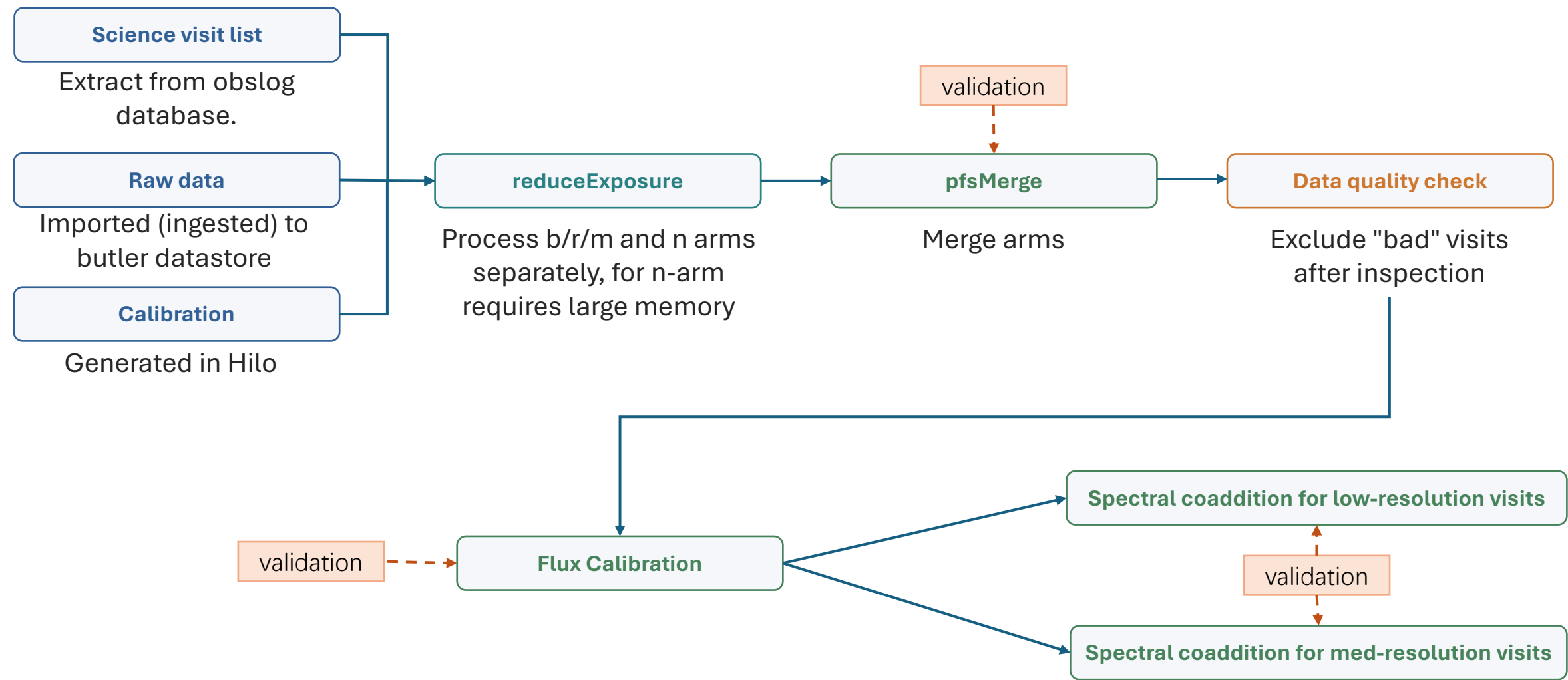
```
graph TD; RawData[Raw data] --> Ready[Ready to transit to Science Platform]; Scripts[Scripts and logs] --> Ready; 2D[2D products] --> Ready; 1D[1D products] --> Ready;
```

Ready to transit to
Science Platform

2D DRP workflow: raw visits to coadds

Main production steps for building science coadds in Mitaka

Validation: check whether the pipeline execution is done completely. Not a data quality check.



Key change from run26: low-resolution and medium-resolution visits are coadded separately.

2D data reduction workflow: calibration

Current strategy for keeping calibration products consistent

Goal

Generate reduction calibration products in Mitaka as part of the local production workflow

Keep calibration collections reproducible and easy to validate

Current practice

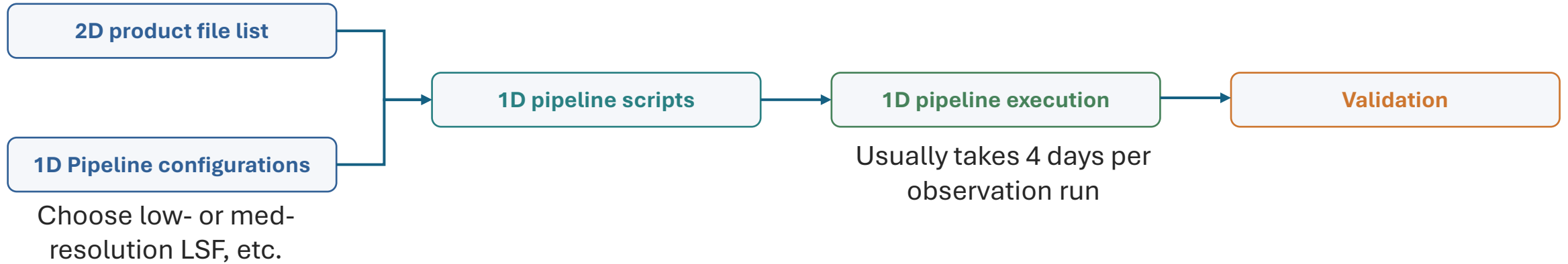
Reduction calibration is still under rapid development

To keep database contents consistent, calibration products are currently generated in Hilo and exported to the production environment

Next step: validate Mitaka-generated calibrations once the calibration workflow becomes stable.

1D DRP workflow

Processing 2D coadd products with the 1D pipeline



Key change from run26: low-resolution and medium-resolution spectras are processed separately.

Challenges

Resource scheduling

BPS deployment is being considered to improve scheduling and utilization

Parallelism must be balanced against memory and database limits

Documentation

Both 2D & 1D pipelines are evolving rapidly
DRP workflow is also changing to make the production process more robust

Documents should be updated accordingly to track changes, provide consistency, and leave clear operational footprints

Workflow evolution

Integrate more tasks into the local data reduction workflow, once those tasks are fully established:

Calibration, Quality assurance, etc.

Pipeline knowledge

Understandings of the pipeline source code

From an “user” to an active feedback provider and eventually a developer

Conclusion

Current status and discussion points

Current status

- Mitaka has a structured DRP workflow covering data management and 2D/1D data-reduction processes
- DEV/PROD environments and data management are now documented

Near-term priorities

- Stabilize the Mitaka calibration-generation workflow
- Evaluate BPS for better cluster utilization

Goal: make the Mitaka DRP workflow robust, reproducible, and easier to hand over.