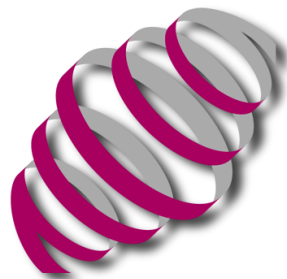




# ASTRON

Netherlands Institute for Radio Astronomy

ASTRON



LOFAR

---

# ASTRON's ATDB-based pipeline execution system

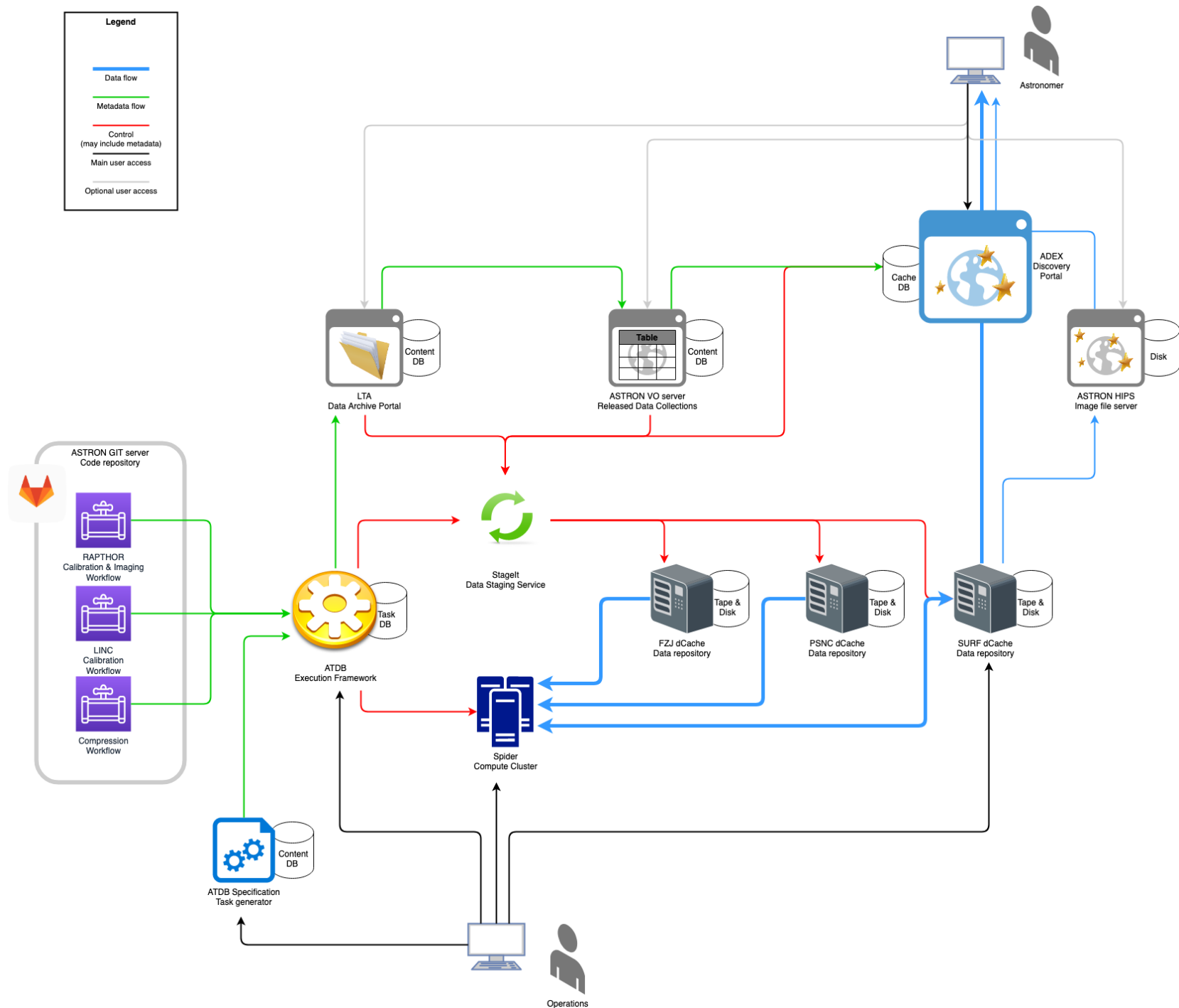
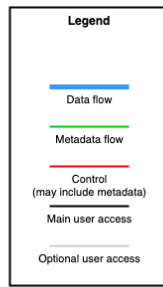
Hanno Holties

ASTRON

Netherlands Institute for Radio Astronomy



# ATDB-based Pipeline Execution System

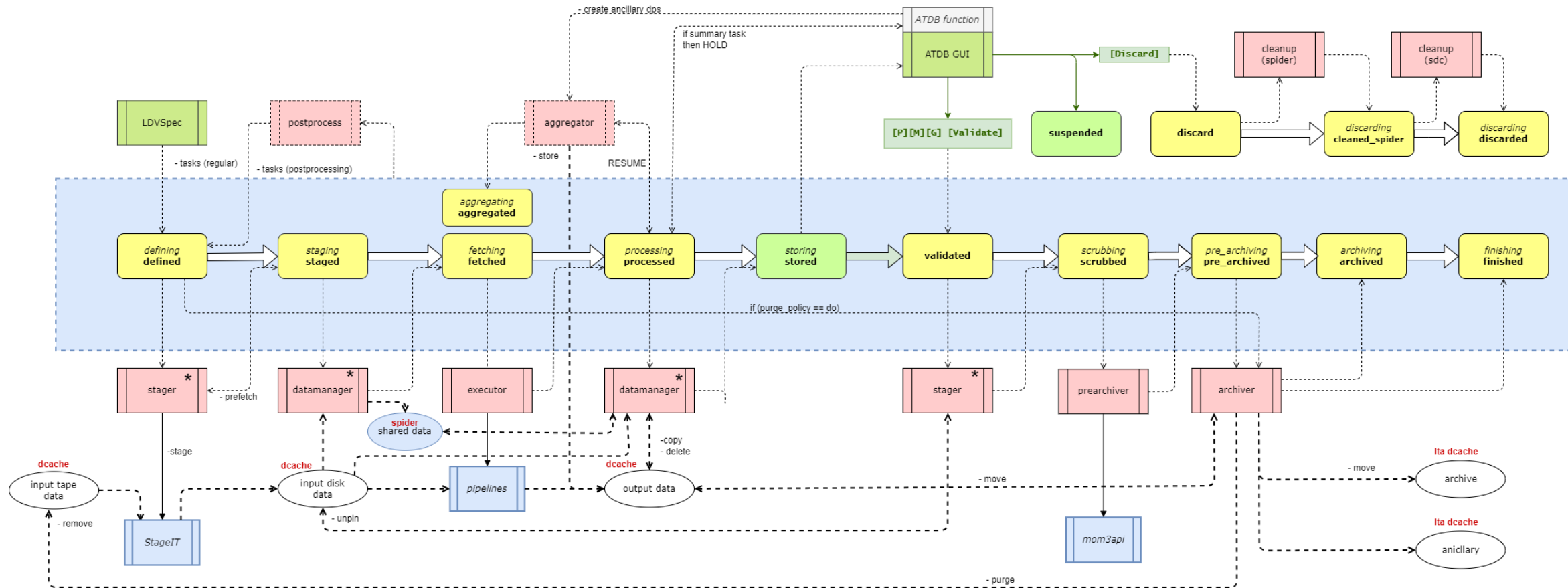


# ATDB - ASTRON Task DataBase

- Task vs Worklow vs Pipeline
  - A **Task** represents a concrete executable piece of work, usually consisting of
    - Input data to fetch
    - A workflow to execute
    - Output data to archive
    - A set of execution parameters
    - And potentially metadata and ancillary data products (e.g. inspection plots) to collect
  - A **Workflow** is a CWL-definition defining inputs, outputs, and steps for data processing
  - A **Pipeline** is a chained set of processing steps
    - May include branching & looping
    - For ATDB, expected to be implemented as a CWL workflow
    - Can also be executed outside of the ATDB framework
    - For ATDB, the 'core' pipeline is usually wrapped in a CWL workflow that handles framework-specific steps such as pre-fetching data, aggregating results, extracting metadata

# ATDB-based Pipeline Execution System

- State machine for microservice-based handling of tasks



id (api)	16663
slurm jobs (api)	jobs for task 16663
priority	100
status	stored
workflow	<a href="#">18 - linc_calibrator_v4_2</a>
<a href="#">Inputs</a> <a href="#">Outputs</a> <a href="#">Metrics</a> <a href="#">Quality</a>	
filter	prefactor_calibrator_test
project	lc10_010
sas_id (api)	667520
creationTime	2021-12-09 12:07:10
purge_policy	no
cleanup_policy	None
resume	True
is_summary	False
stage_request_id	<a href="#">29194</a>
size_to_process	83.9 GB
size_processed	662.8 MB
total_processing_time	98425
quality	None
calculated qualities	None
remarks	None

# Workflow Details View Api

GET /atdb/workflows/18/

HTTP 200 OK

Allow: GET, PUT, PATCH, DELETE, HEAD, OPTIONS

Content-Type: application/json

Vary: Accept

```
{
  "id": 18,
  "description": "Test",
  "tag": "Test",
  "workflow_uri": "linc_calibrator_v4_2",
  "repository": "https://git.astron.nl/RD/LINC.git",
  "commit_id": "ldv_v403",
  "path": "workflows/ldv_linc_calibrator.cwl",
  "oi_size_fraction": 1.0,
  "meta_scheduling": {
    "#SBATCH --cpus-per-task": 10
  },
  "default_parameters": null,
  "prefetch": false,
  "aggregation_strategy": "none",
  "aggregation_script": null,
  "quality_thresholds": null
}
```

master LINC / workflows / +



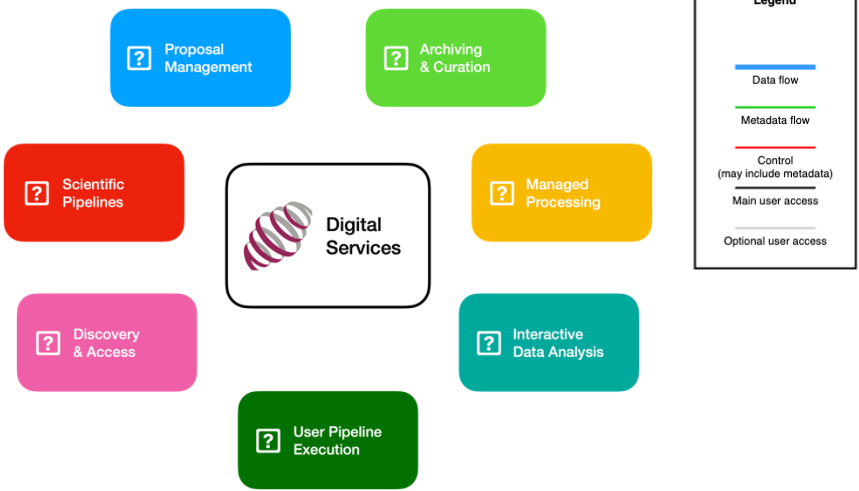
Add maxncpu\_flag option  
alex authored 1 week ago

Name	Last commit
..	
linc_calibrator	Add maxncpu_flag option
linc_target	[RAP-63]: Update calibrator p
HBA_calibrator.cwl	Add maxncpu_flag option
HBA_target.cwl	Use RMextract master branch
LBA_calibrator.cwl	[RAP-63]: Update calibrator p
LBA_target.cwl	Use RMextract master branch
ldv_linc_calibrator.cwl	Improve LDV workflow
ldv_linc_target.cwl	Improve LDV workflow
linc_calibrator.cwl	Add maxncpu_flag option
linc_target.cwl	Fix #56 - LINC selfcal crash i

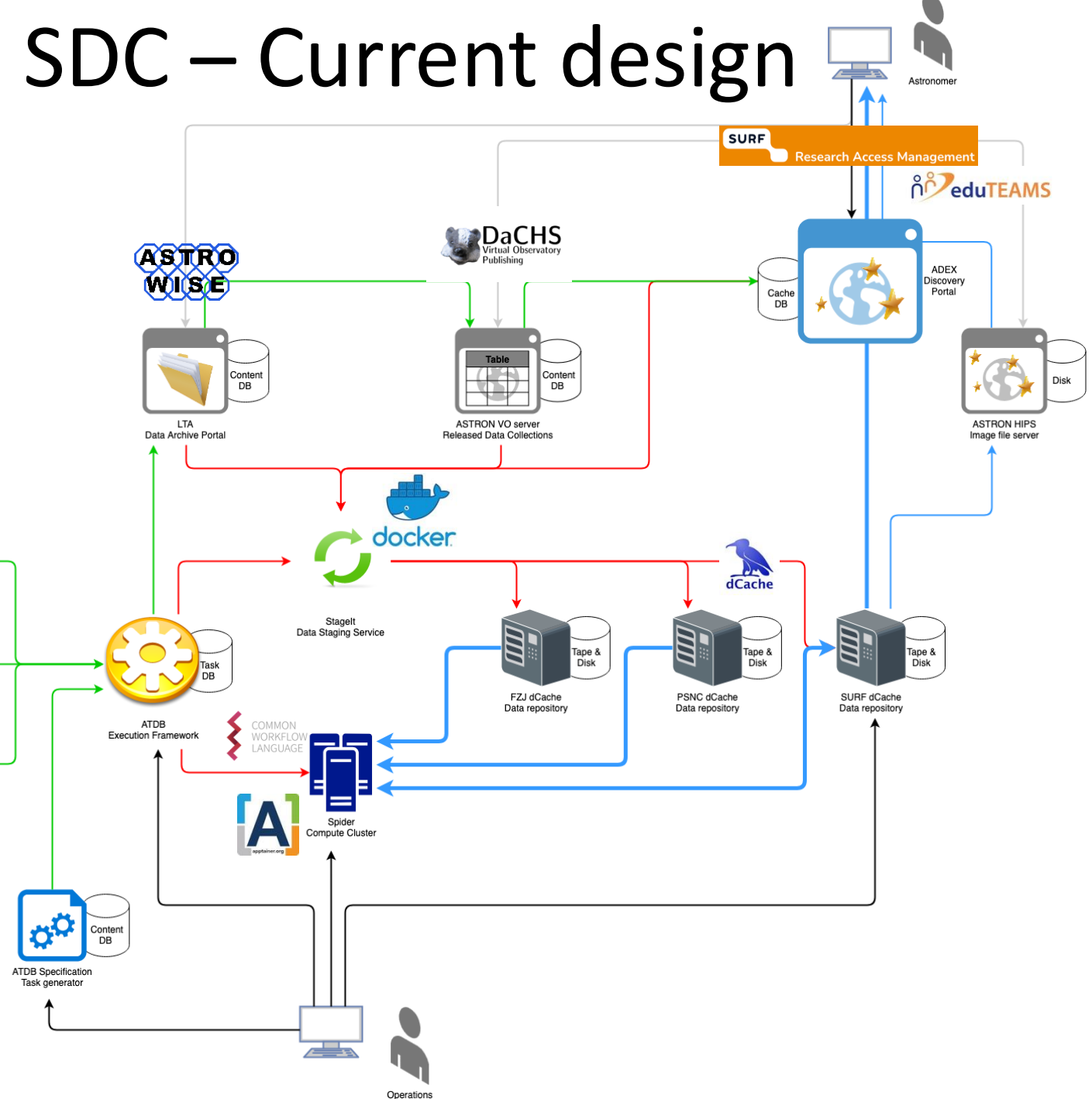
# LTA Current developments

## - Integrated/managed data processing

- Common Workflow Language (**CWL**) pipeline executors
- **Apptainer/Singularity** based software deployment
  - TBD: caching vs distribution
- Microservice based **framework**
  - **Centralized**
    - Central task database (process 'management')
    - Task specification
    - Stager (Prepare input data for access)
    - Archiver (Ingest output in LTA: catalogue update & move to permanent storage)
  - **Distributed** (running in data center – communicating with central task database)
    - Executor (job submission & monitoring)
    - Datamanager (move data between storage infrastructures)



# SDC – Current design



## • Core technology

- **Gitlab** code repository & CI/CD
- SURF **SRAM** / eduTEAMS FAAL provider
- **CWL** (Common Workflow Language) Pipeline definitions
- **Apptainer** (formerly Singularity) & **Docker** container technology
- **dCache** tiered storage data management
- **AstroWise** LTA Catalogue & Portal
- **GAVO DaCHS** VO standards service



# Block Diagram!

