EXTRACT

A distributed data-mining software platform for
extreme data across the compute continuum

TASKA

Transient Astrophysics
with a Square Kilometre Array pathfinder

Julien N. Girard, **Baptiste Cecconi** and the EXTRACT collaboration

See also: talk of E. Mauduit and the Poster of J. Girard

extract-project.eu

The EXTRACT Project has received funding from the European Union's Horizon Europe programme under grant agreement number 101093110.

EXTRACT

A distributed data-mining software platform for extreme data across the compute continuum

EU project

https://extract-project.eu/

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Université Paris Cité

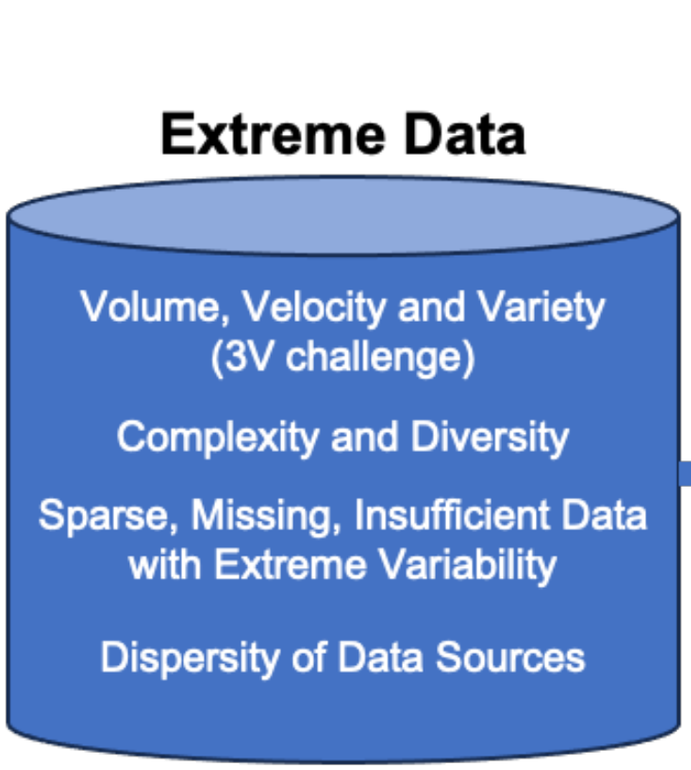
EXTRACT

A distributed data-mining software platform for extreme data across the compute continuum

EU project

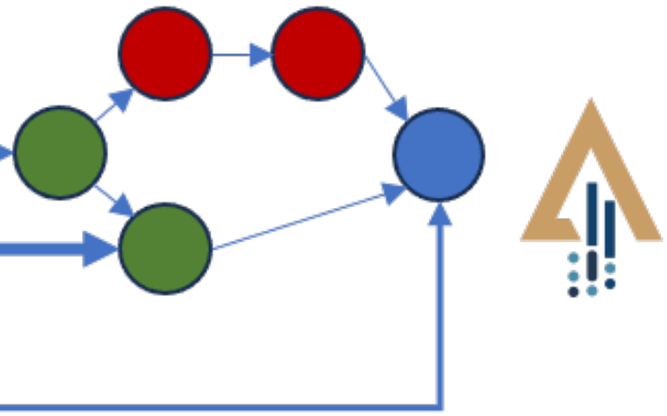
https://extract-project.eu/

EXTRACT aims to create a data-mining **software platform** for **extreme data** across the **compute continuum**



Data Mining Software Platform

- Data infrastructures and AI & Big-data frameworks
- Data-driven orchestration
- Interoperability



Compute Continuum
HPC
Edge
Cloud

EXTRACT

A distributed data-mining software platform for extreme data across the compute continuum

EU project

https://extract-project.eu/

ikerlan
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

UNIVERSITAT ROVIRA i VIRGILI

binaré

sixsq

cnrs

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

LOGOS RI
RESEARCH&INNOVATION

MATHEMA

IBM

Observatoire de Paris | PSL

Université Paris Cité

EXTRACT aims to create a data-mining **software platform** for **extreme data** across the **compute continuum**

Extreme Data

Volume, Velocity and Variety (3V challenge)

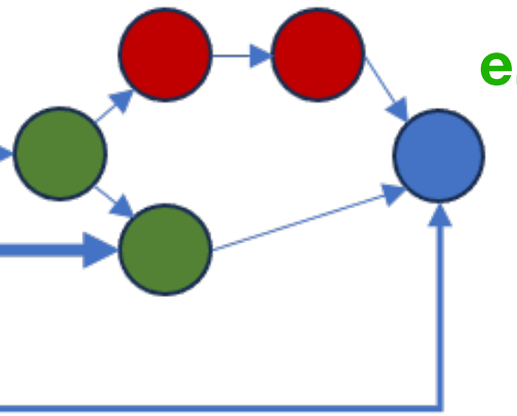
Complexity and Diversity

Sparse, Missing, Insufficient Data with Extreme Variability

Dispersity of Data Sources

Data Mining Software Platform

- Data infrastructures and AI & Big-data frameworks
- Data-driven orchestration
- Interoperability



e.g. NenuFAR

Compute Continuum
HPC
← Edge
Cloud

EXTRACT

A distributed data-mining software platform for extreme data across the compute continuum

EU project

https://extract-project.eu/

ikerlan
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

UNIVERSITAT ROVIRA i VIRGILI

binaré

sixsq

cnrs

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

LOGOS RI
RESEARCH&INNOVATION

MATHEMA

IBM

Observatoire de Paris | PSL

Université Paris Cité

EXTRACT aims to create a data-mining **software platform** for **extreme data** across the **compute continuum**

Extreme Data

Volume, Velocity and Variety (3V challenge)

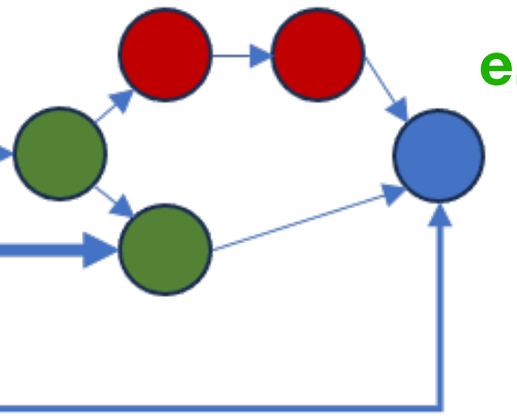
Complexity and Diversity

Sparse, Missing, Insufficient Data with Extreme Variability

Dispersity of Data Sources

Data Mining Software Platform

- Data infrastructures and AI & Big-data frameworks
- Data-driven orchestration
- Interoperability



e.g. NenuFAR

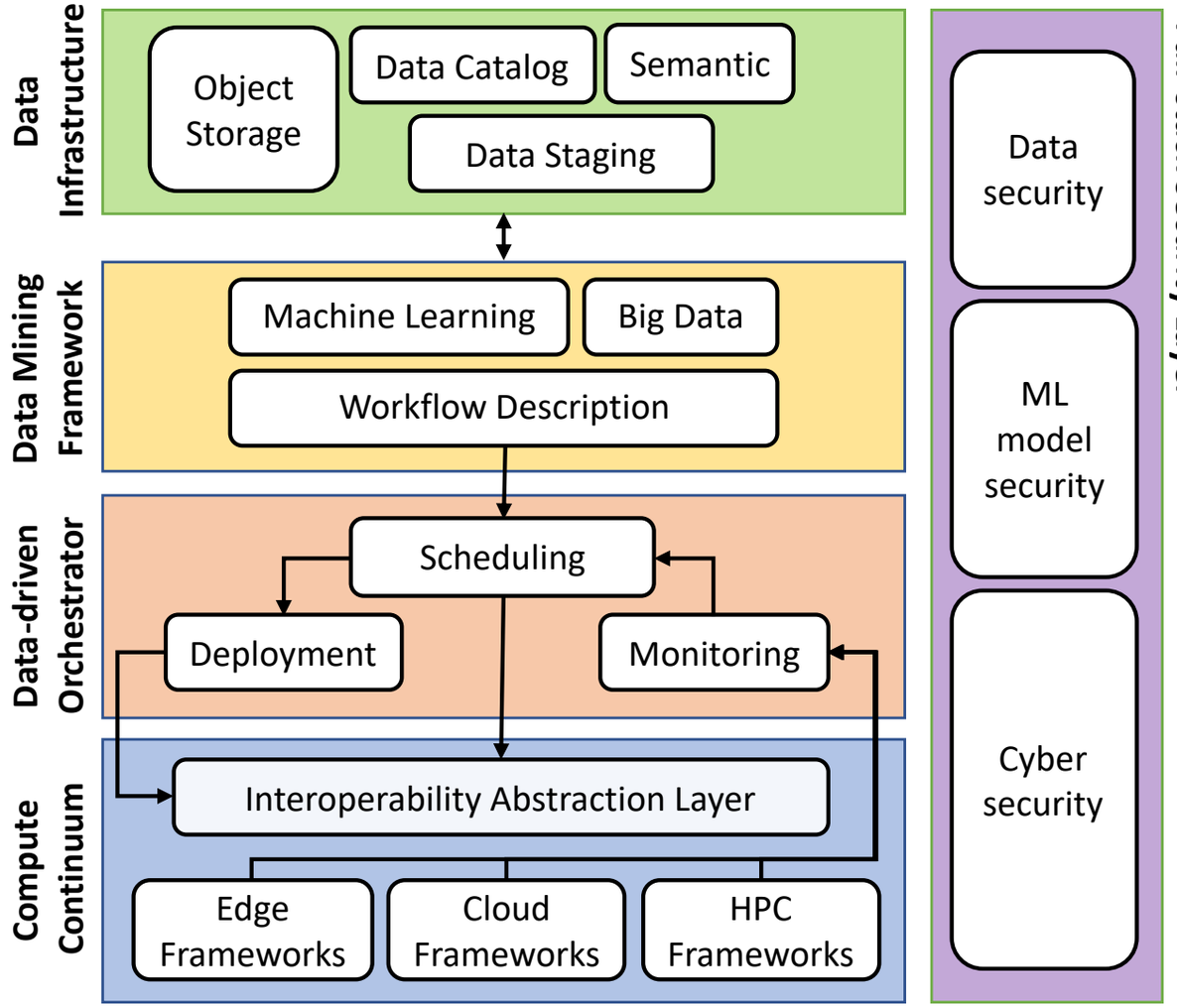
SKA ?

Compute Continuum
HPC
Edge
Cloud

EXTRACT

A distributed data-mining software platform for extreme data across the compute continuum

- Handle the complete lifecycle and value chain of extreme data
  - **Data collection** across highly distributed and heterogeneous sources
  - **Data mining** of meaningful, accurate, reliable and useful knowledge
  - **Secure and trustworthy used of knowledge** by applications and end users



- Everything looks local

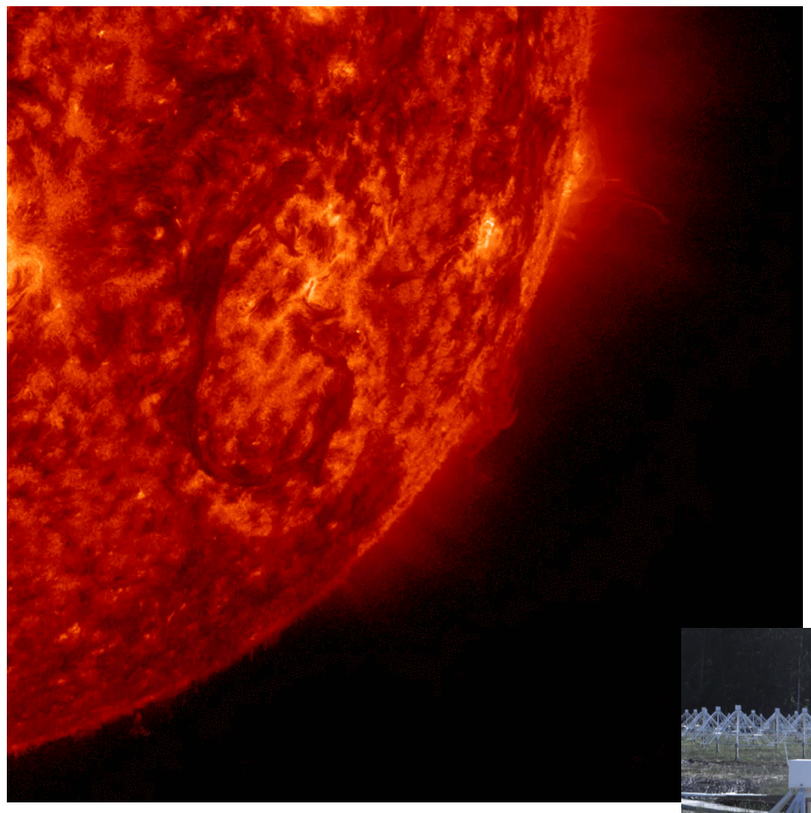
EXTRACT Use-Cases, sharing the same platform

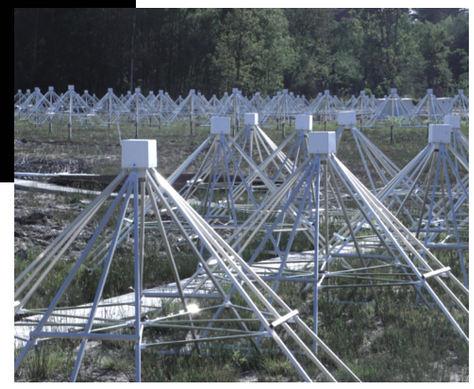
PER

Personalised Evacuation Route (PER)
in the City of Venice based on an
Urban Digital Twin and an AI engine

TASKA

Transient Astrophysics with the Square Kilometre Array pathfinder (TASKA)
NenuFAR generating high-volume and high-velocity data





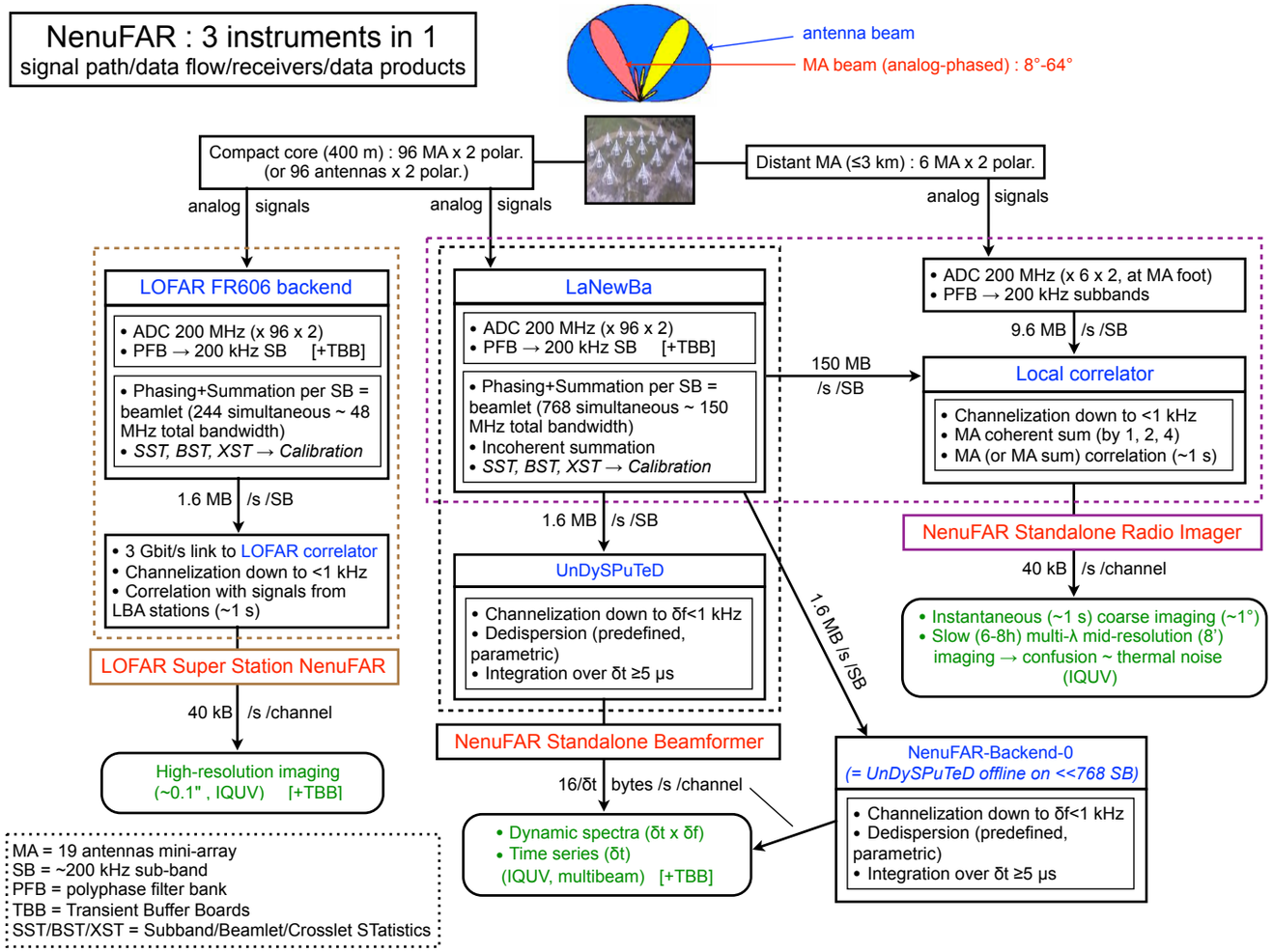
NenuFAR *New extension in Nançay Upgrading loFAR*

Pathfinder de SKA (LOW) , Infrastructure de recherche F = 20-80 MHz

N_A~2000 antennes Fonctionnement en mode réseau phasé et interféromètre

Extreme data with NenuFAR

- **~2000 antennas** (96 mini-arrays of 19 antennas + 6 remotes mini-arrays)
- **Beamform data:**
  raw data rate = 1.2 GB/s (~35 PB/yr)
- **Imaging data:**
  raw data rate = 8.6 TB/hr (~74 PB/yr)
- **Local data storage:** 3.5 PB
- **Science teams are reducing data**
  down to about 1 PB/yr
  (>1/100 of raw data rate)

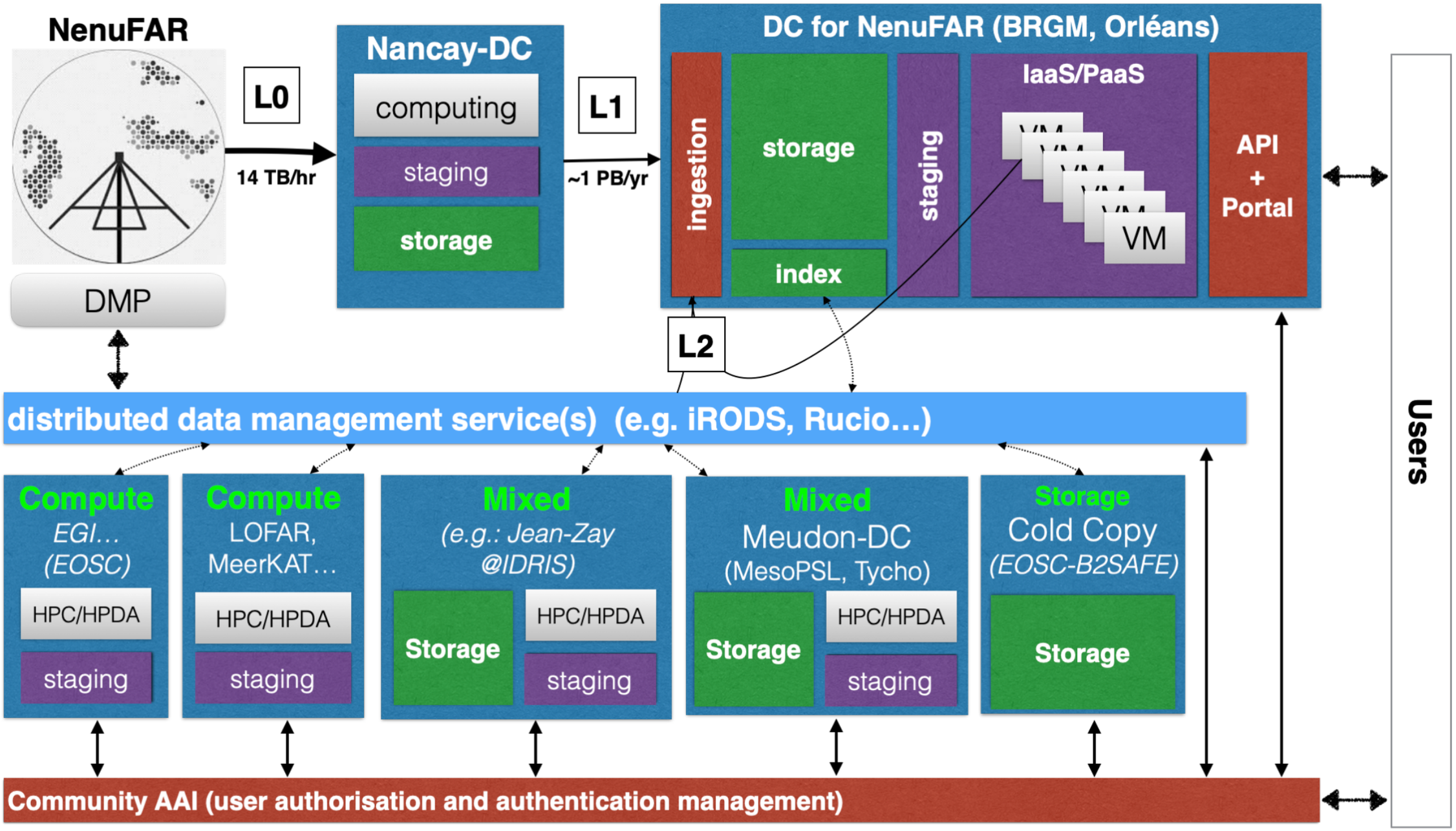


- **Reduced data transferred to distributed datacenter**
  (currently: 3 PB in Meudon and 5 PB in Orléans)
- **Processing of imaging data on distributed datacenter:**
  data can't be moved easily, need to process where the data is located

NenuFAR digital infrastructure

"Edge" = Nançay facility (NenuFAR backend + Nançay Data Centre)

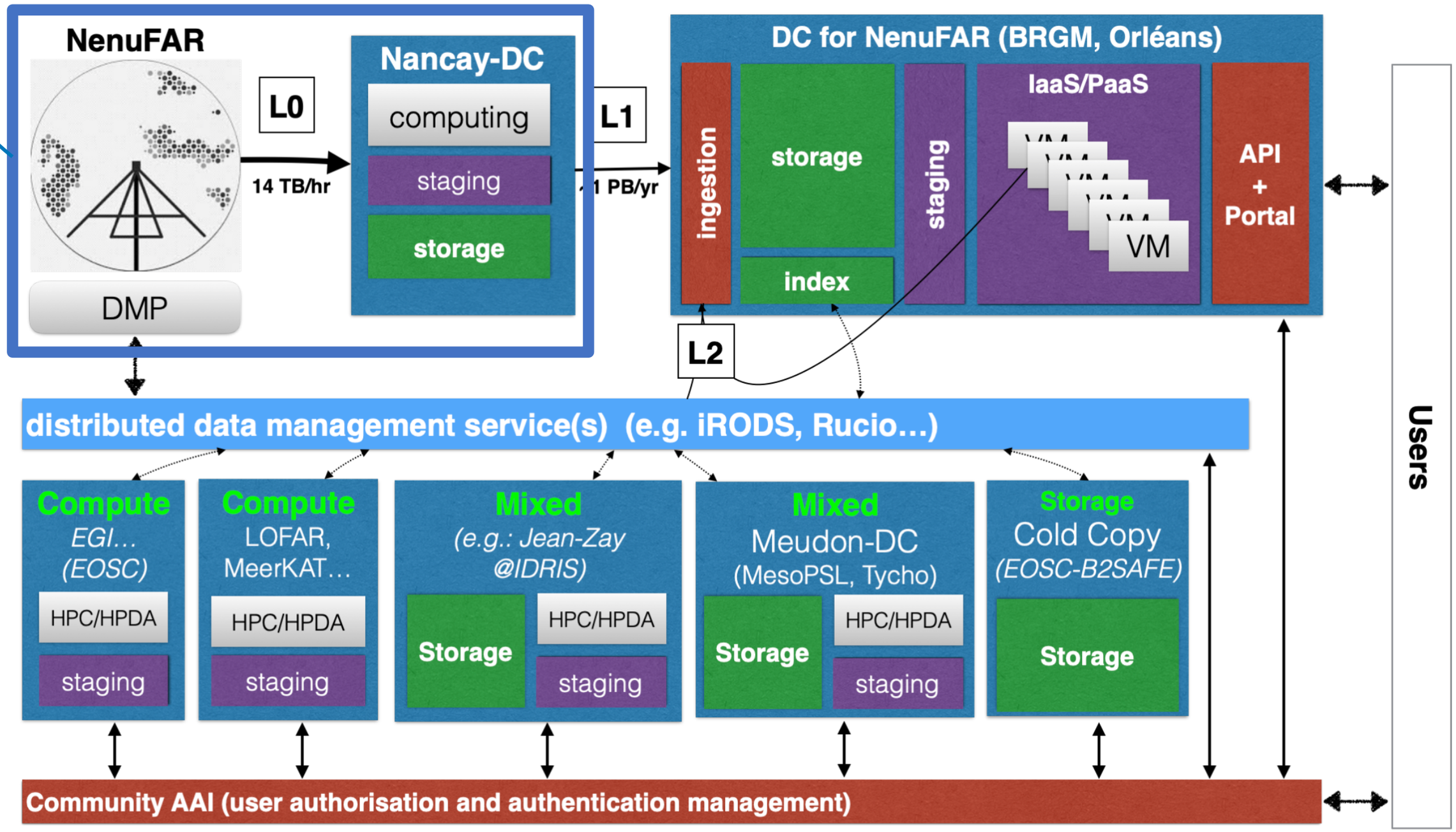
"Cloud" = "Datalake" (NenuFAR Data Centre + partners)



NenuFAR digital infrastructure

"Edge" = Nançay facility (NenuFAR backend + Nançay Data Centre)

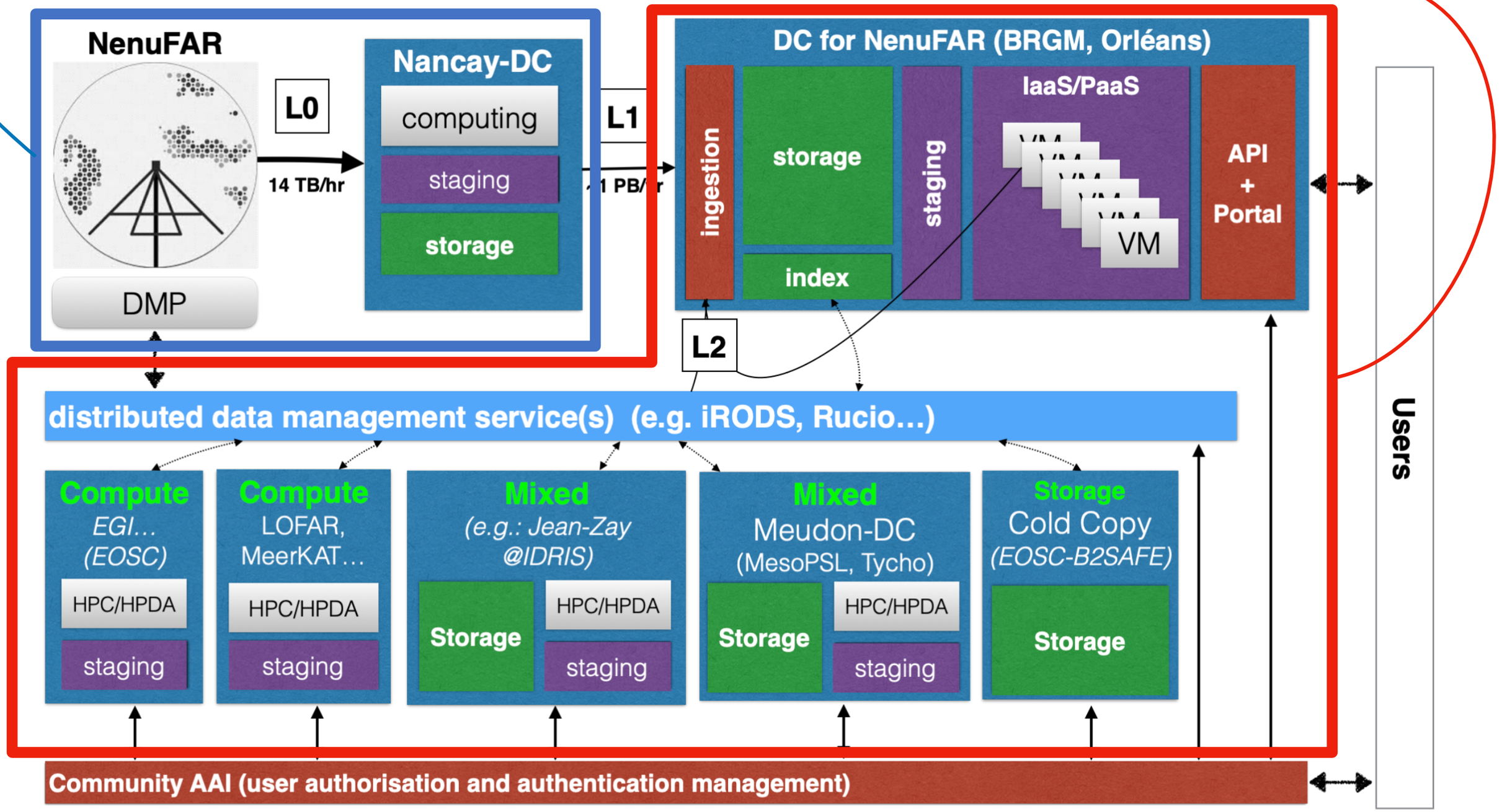
"Cloud" = "Datalake" (NenuFAR Data Centre + partners)



NenuFAR digital infrastructure

"Edge" = Nançay facility (NenuFAR backend + Nançay Data Centre)

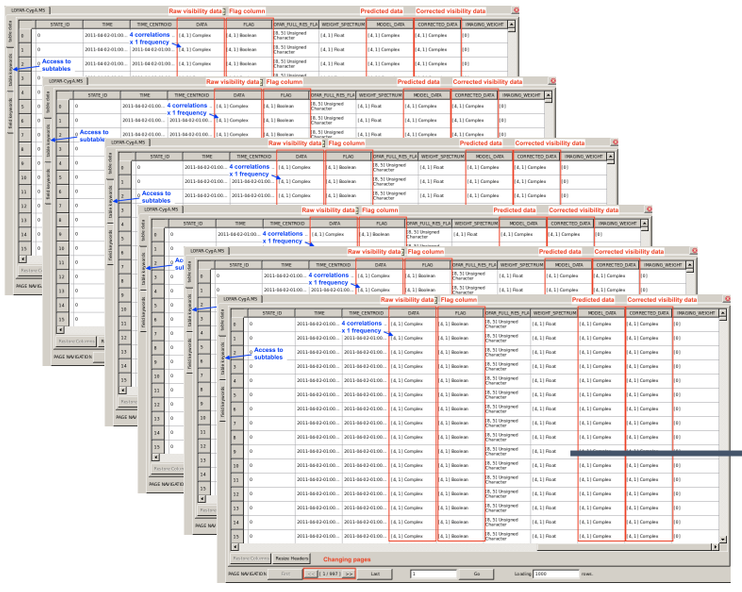
"Cloud" = "Datalake" (NenuFAR Data Centre + partners)



TASKA Use Case Overview

- **Use Case A:** Early detection and selective resolution data recording (space optimality)
- **Use Case C:** Workflow orchestration of interferometric data processing with a focus on improving the processing speed, accuracy and automation on large datasets
- **Use Case D:** Prototype development for “dynamic” imaging of the variable Universe (DL transient imaging)
- **Use Case E:** Advanced data reduction workflows for multi-dimensional real-time analysis and inference (joining A and C together)

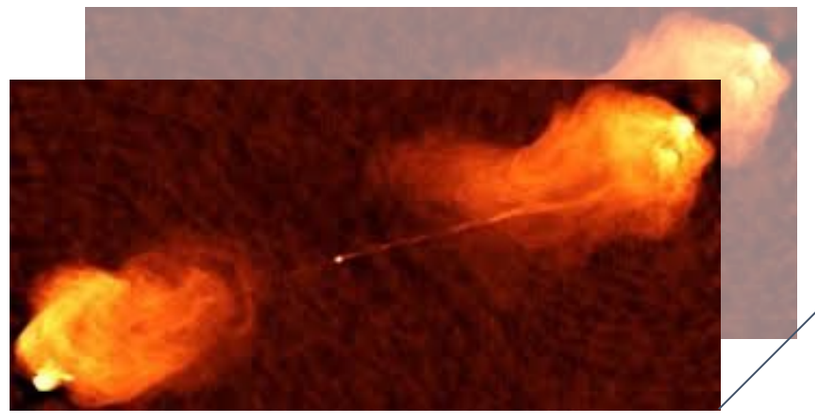
Use-Case C: Workflow orchestration for radiointerferometry



Starting dataset: Visibilities
(Measurement Sets (MS) Format)

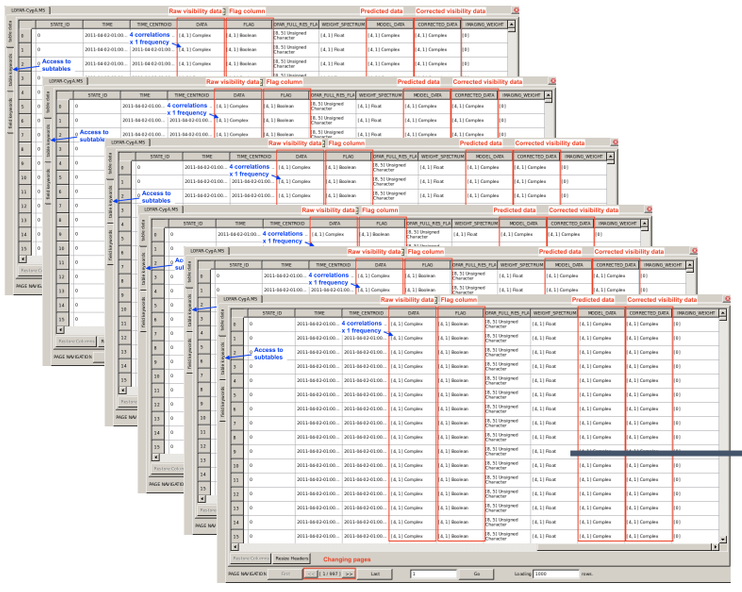
Data transfer
data processing

(Flagging, Rebining,
Calibration, Imaging)



Final product: time/freq
Image cubes

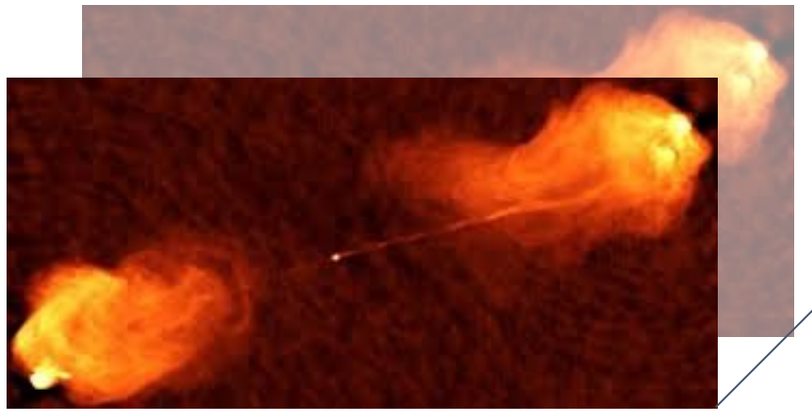
Use-Case C: Workflow orchestration for radiointerferometry



Starting dataset: Visibilities
(Measurement Sets (MS) Format)

Data transfer
data processing

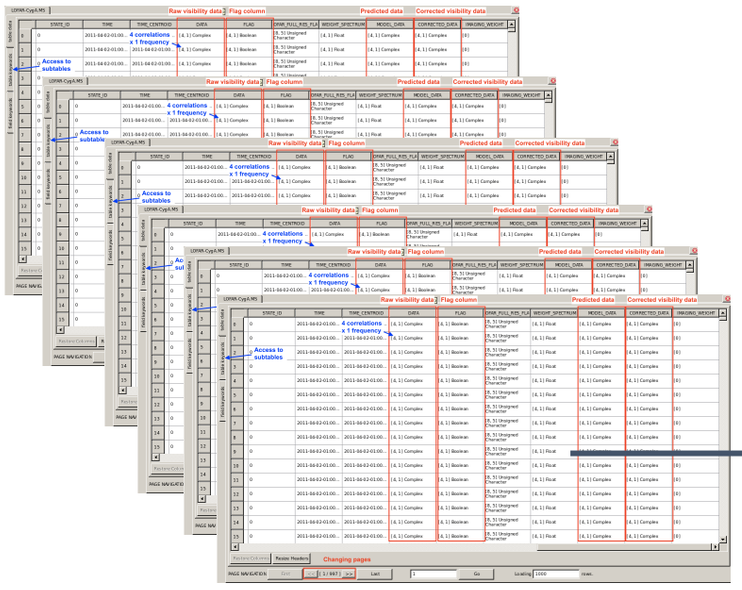
(Flagging, Rebining,
Calibration, Imaging)



Final product: time/freq
Image cubes

Optimal dataset distribution ?
(Multiple sites)

Use-Case C: Workflow orchestration for radiointerferometry



Starting dataset: Visibilities
(Measurement Sets (MS) Format)

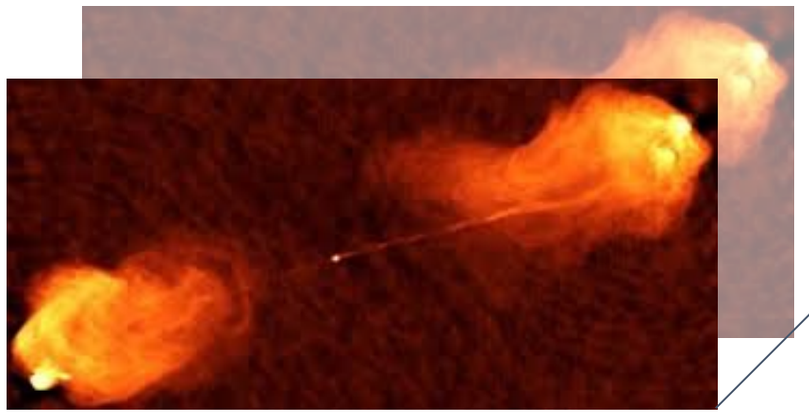
Optimal dataset distribution ?
(Multiple sites)

Data transfer
data processing

(Flagging, Rebining,
Calibration, Imaging)

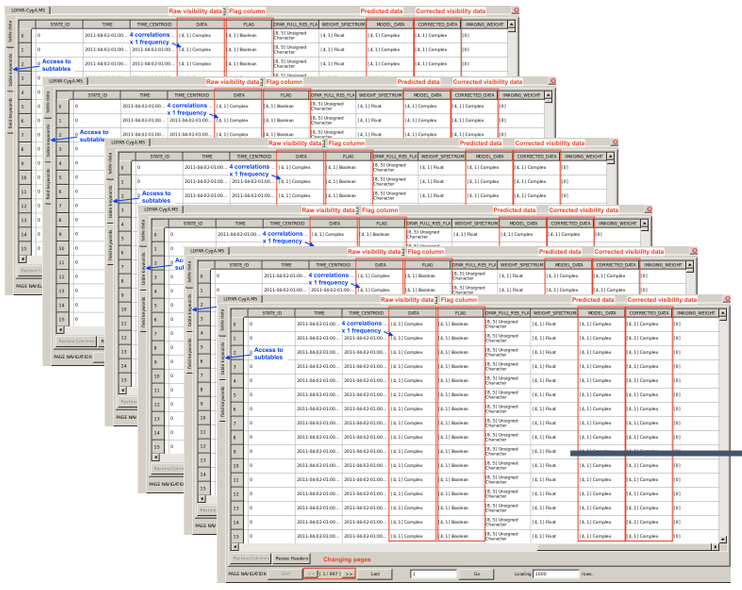
Flexible reduction?

(Multiple tools)
from combination of
known analytic “bricks”



Final product: time/freq
Image cubes

Use-Case C: Workflow orchestration for radiointerferometry



Starting dataset: Visibilities
(Measurement Sets (MS) Format)

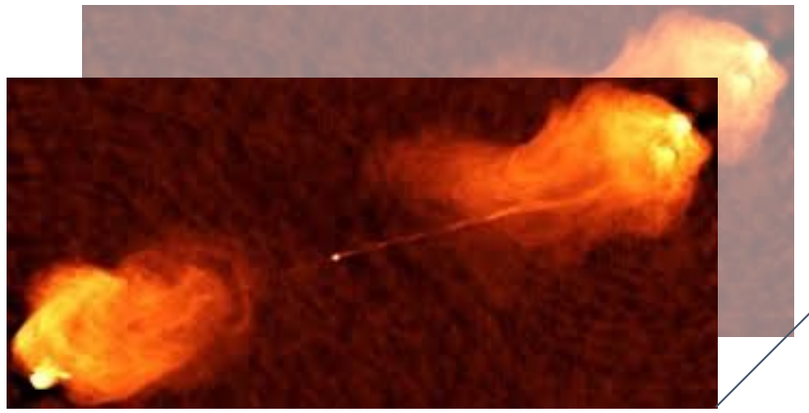
Optimal dataset distribution ?
(Multiple sites)

Data transfer
data processing

(Flagging, Rebining, Calibration, Imaging)

Flexible reduction?

(Multiple tools)
from combination of
known analytic "bricks"

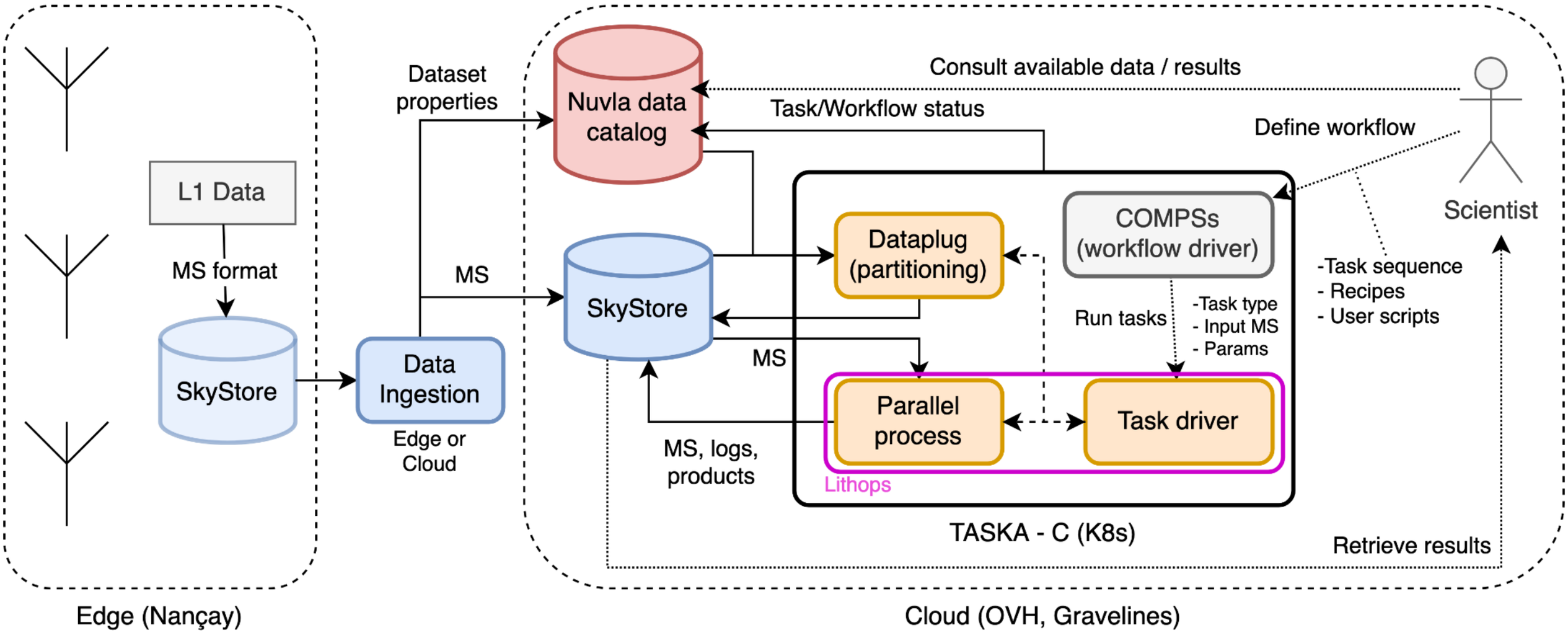


Final product: time/freq
Image cubes

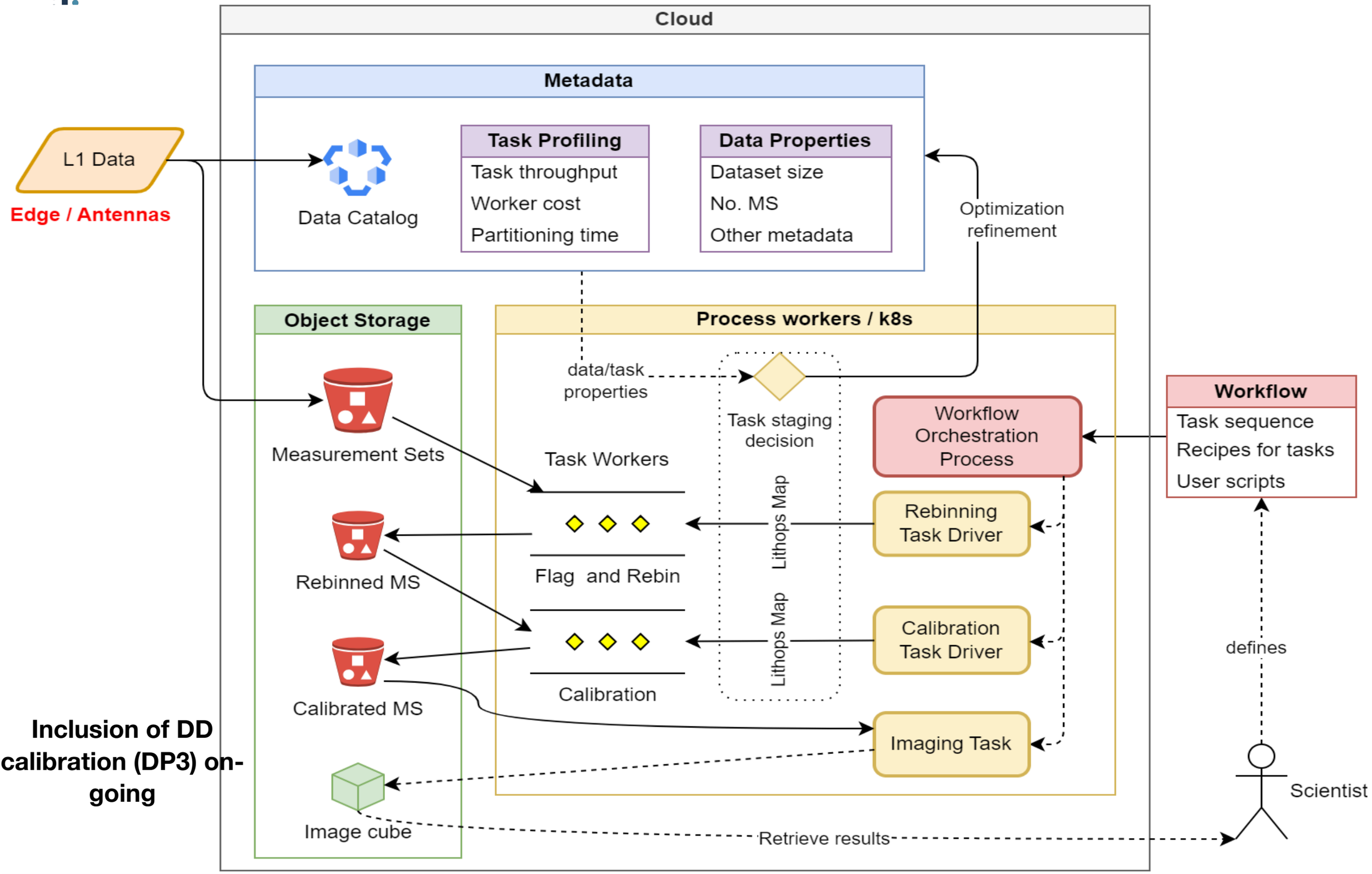
Analytics

Multiple tools
(scientific quality, fidelity)

Use-Case C: Architecture view



Use-Case C: User view

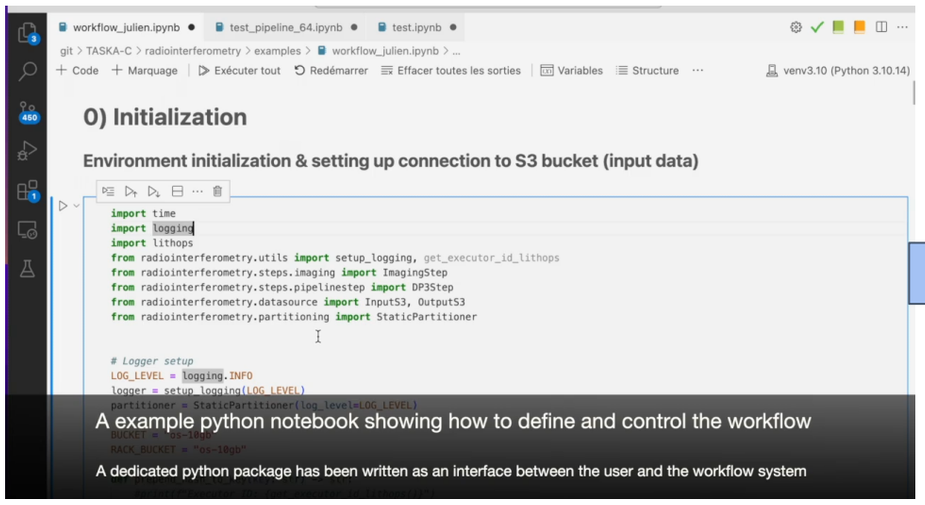


Inclusion of DD calibration (DP3) on-going

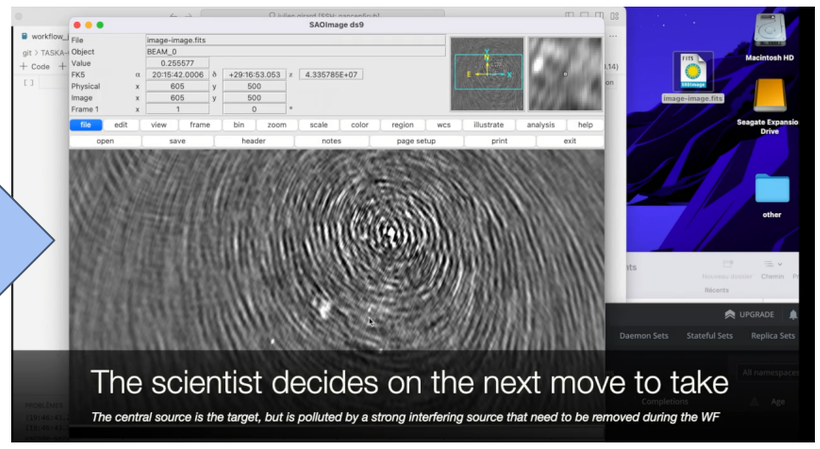
TASKA - MVP “Interactive” Workflow

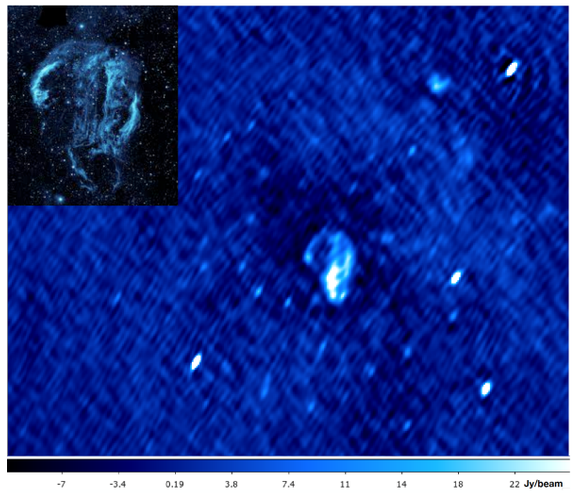
- Built as a “wrapper” that interacts with the astronomy community tools
  High potential impact because of the platform deployment in other communities (security, medical, resource management, etc.)
- Easy to invoke, easy to code, easy to customize, easy to “chain”: *natively made for workflows*
- Each task has a “definition” block and a “run” block: *separating the workflow building from its running*
- Run as a python script or in a python notebook (cf. DEMO video)

```
import time
import logging
import lithops
from radiointerferometry.utils import setup_logging, get_executor_id_lithops
from radiointerferometry.steps.imaging import ImagingStep
from radiointerferometry.steps.pipelinestep import DP3Step
from radiointerferometry.datasource import InputS3, OutputS3
from radiointerferometry.partitioning import StaticPartitioner
```



Controlled through a python notebook (S3, data partitioning, worker management, …)



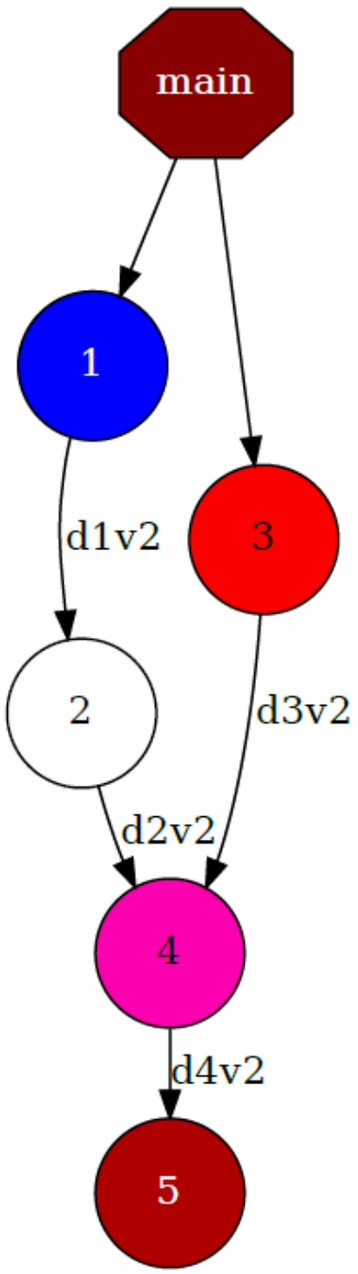


The final products are then retrieved on the scientist computer

…as if the process and data were **local**

TASKA - MVP “Automated” Workflow

Implementing the chaining of tasks and decision making & replay capabilities



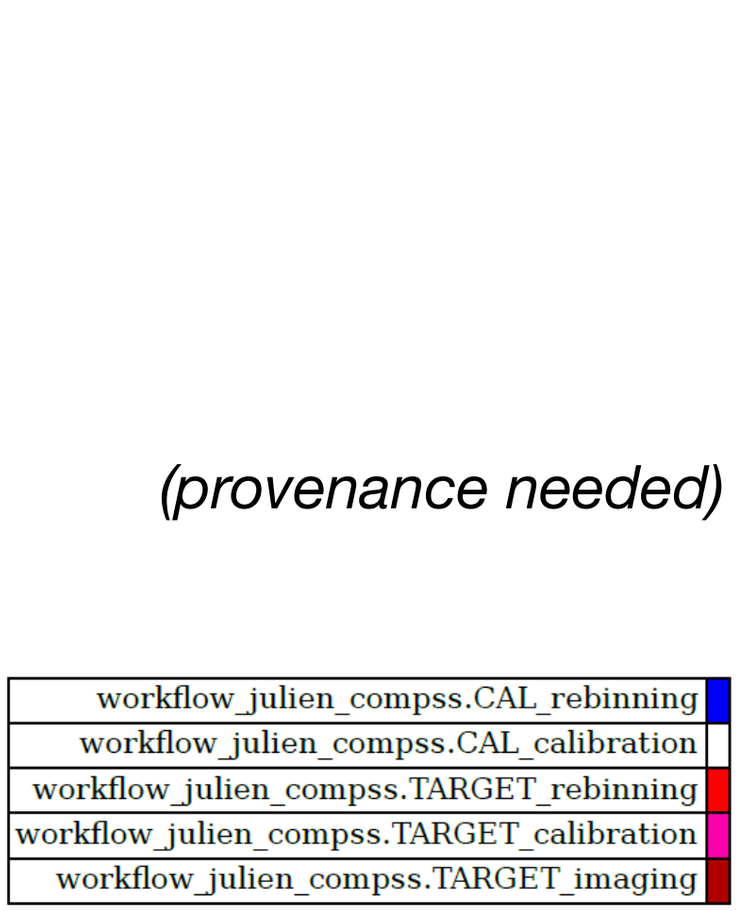
workflow_julien_compss.CAL_rebinning
workflow_julien_compss.CAL_calibration
workflow_julien_compss.TARGET_rebinning
workflow_julien_compss.TARGET_calibration
workflow_julien_compss.TARGET_imaging

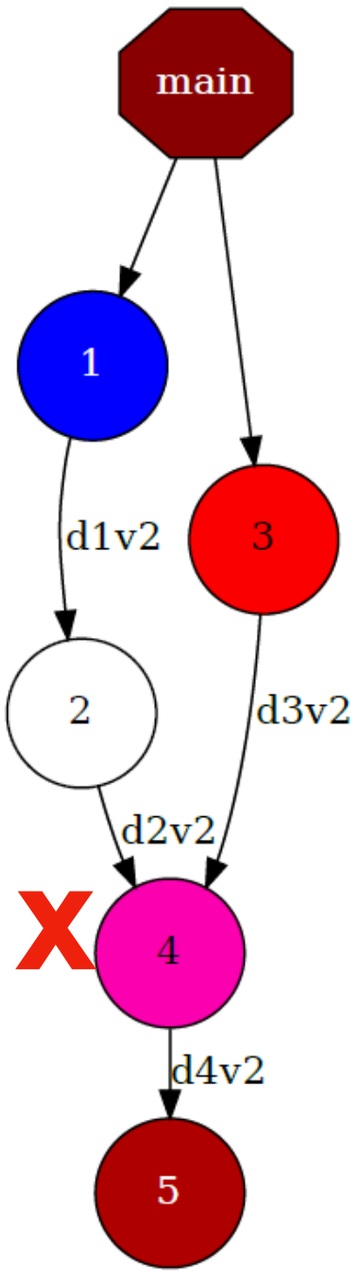
On-going development on EOSC, EGI/CESNET, OVH, (Soon Nançay/Obs)

TASKA - MVP “Automated” Workflow

Implementing the chaining of tasks and decision making & replay capabilities

(provenance needed)

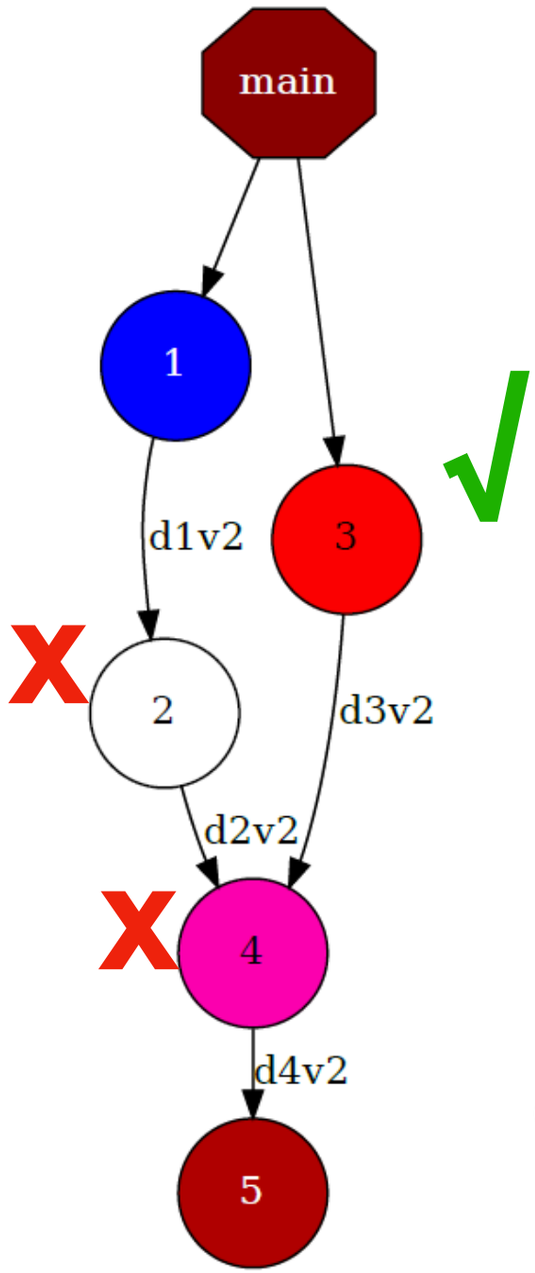




On-going development on EOSC, EGI/CESNET, OVH, (Soon Nançay/Obs)

TASKA - MVP “Automated” Workflow

Implementing the chaining of tasks and decision making & replay capabilities



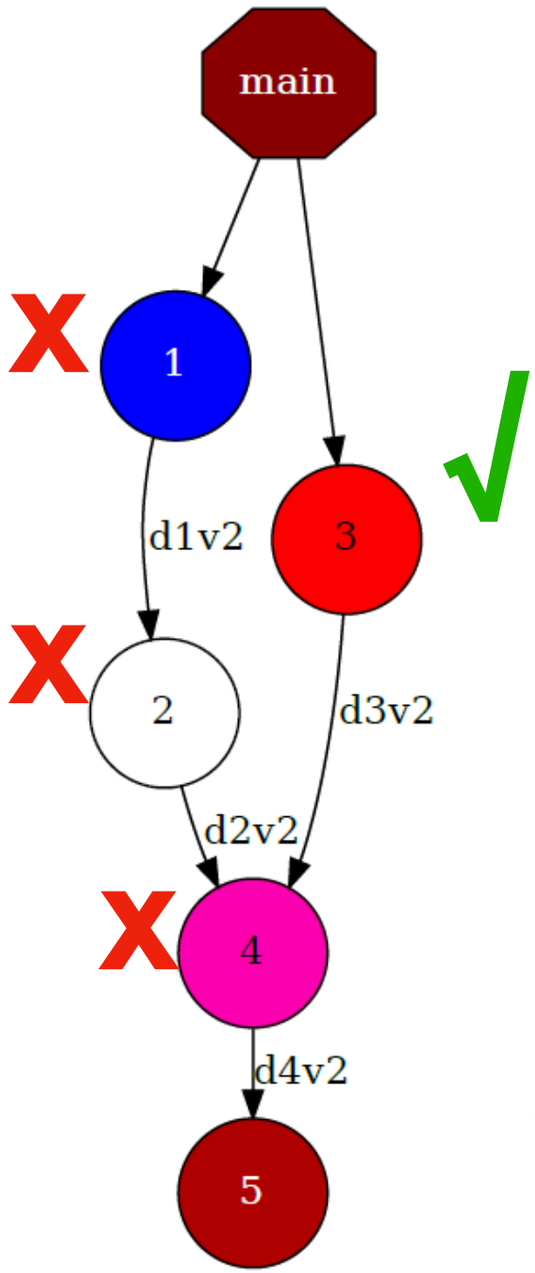
(provenance needed)

workflow_julien_compss.CAL_rebinning
workflow_julien_compss.CAL_calibration
workflow_julien_compss.TARGET_rebinning
workflow_julien_compss.TARGET_calibration
workflow_julien_compss.TARGET_imaging

On-going development on EOSC, EGI/CESNET, OVH, (Soon Nançay/Obs)

TASKA - MVP “Automated” Workflow

Implementing the chaining of tasks and decision making & replay capabilities



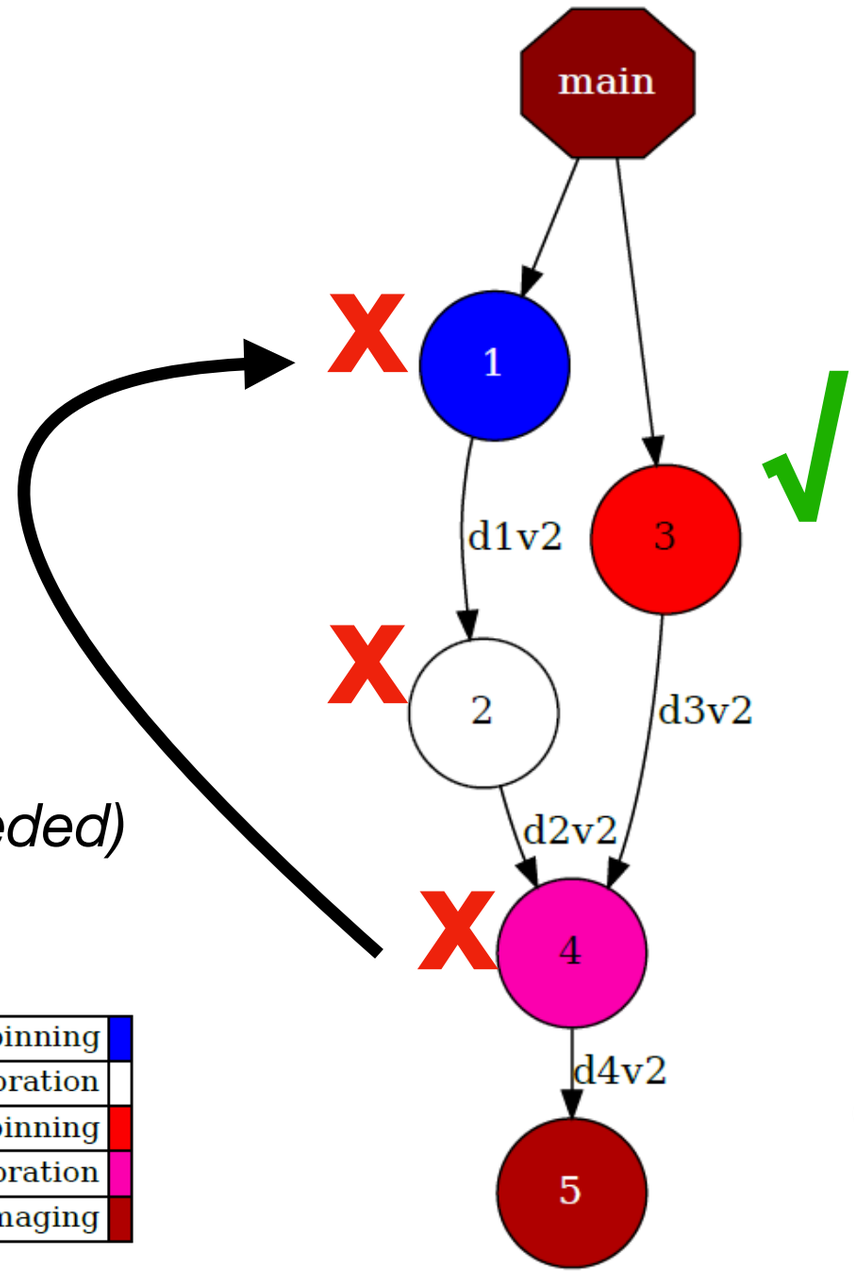
(provenance needed)

workflow_julien_compss.CAL_rebinning
workflow_julien_compss.CAL_calibration
workflow_julien_compss.TARGET_rebinning
workflow_julien_compss.TARGET_calibration
workflow_julien_compss.TARGET_imaging

On-going development on EOSC, EGI/CESNET, OVH, (Soon Nançay/Obs)

TASKA - MVP “Automated” Workflow

Implementing the chaining of tasks and decision making & replay capabilities



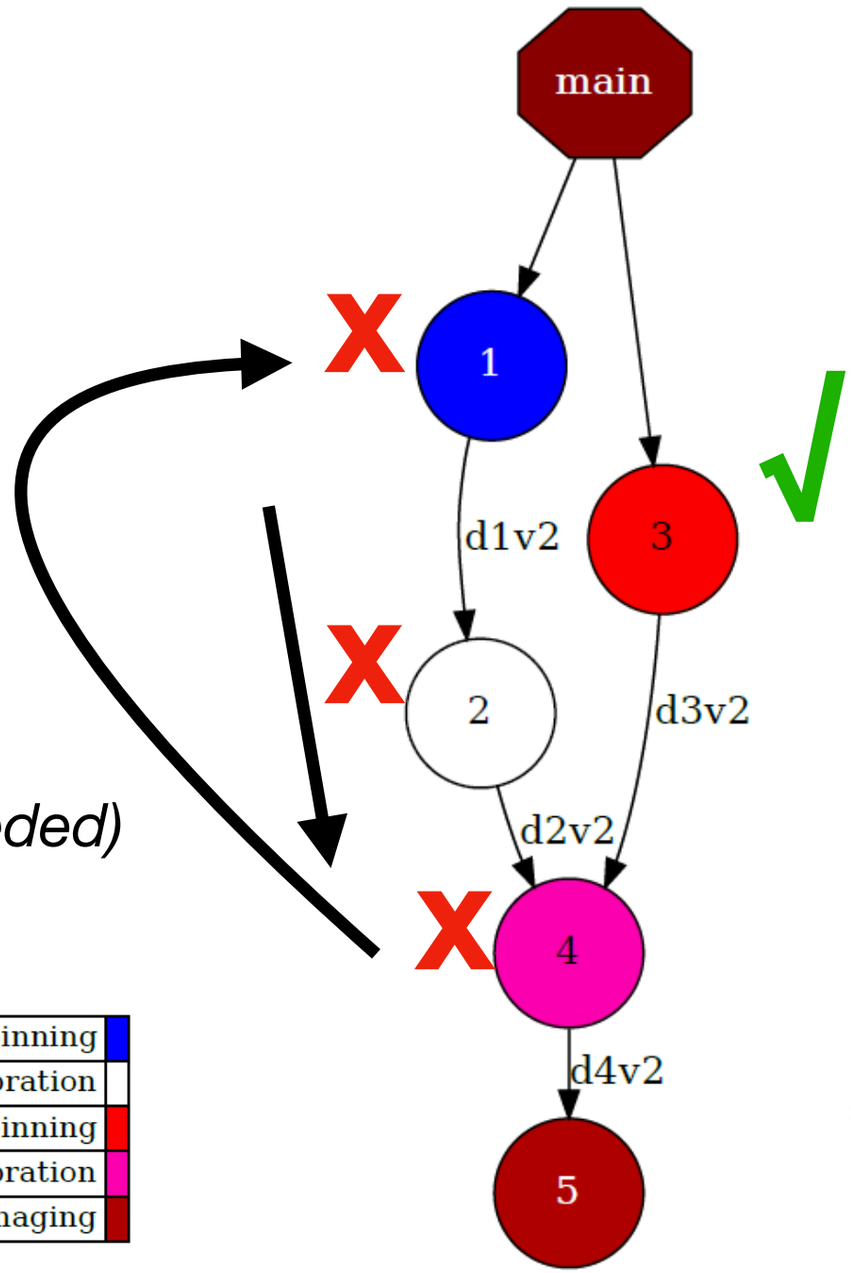
(provenance needed)

workflow_julien_compss.CAL_rebinning
workflow_julien_compss.CAL_calibration
workflow_julien_compss.TARGET_rebinning
workflow_julien_compss.TARGET_calibration
workflow_julien_compss.TARGET_imaging

On-going development on EOSC, EGI/CESNET, OVH, (Soon Nançay/Obs)

TASKA - MVP “Automated” Workflow

Implementing the chaining of tasks and decision making & replay capabilities



(provenance needed)

workflow_julien_compss.CAL_rebinning
workflow_julien_compss.CAL_calibration
workflow_julien_compss.TARGET_rebinning
workflow_julien_compss.TARGET_calibration
workflow_julien_compss.TARGET_imaging

On-going development on EOSC, EGI/CESNET, OVH, (Soon Nançay/Obs)

EXTRACT - TASKA - Summary

TASKA-C

- We have developed a **framework for distributed data computing** on cloud clusters
- Currently validating
  - unsupervised/automated workflow
  - running a step on an HPC resource
  - running on a multi-cluster scale (data distributed in several data centers)
- Application on NenuFAR (SKA pathfinder)
- Clear huge potential for SRCNet

Other work not addressed in this presentation:

TASKA-A

- Real time detection (possibly with AI) on high resolution data stream (dynamic spectra): implemented on NenuFAR beamformer backend

TASKA-D

- On going work on new imager for dynamical sources with AI-based video reconstruction

Components

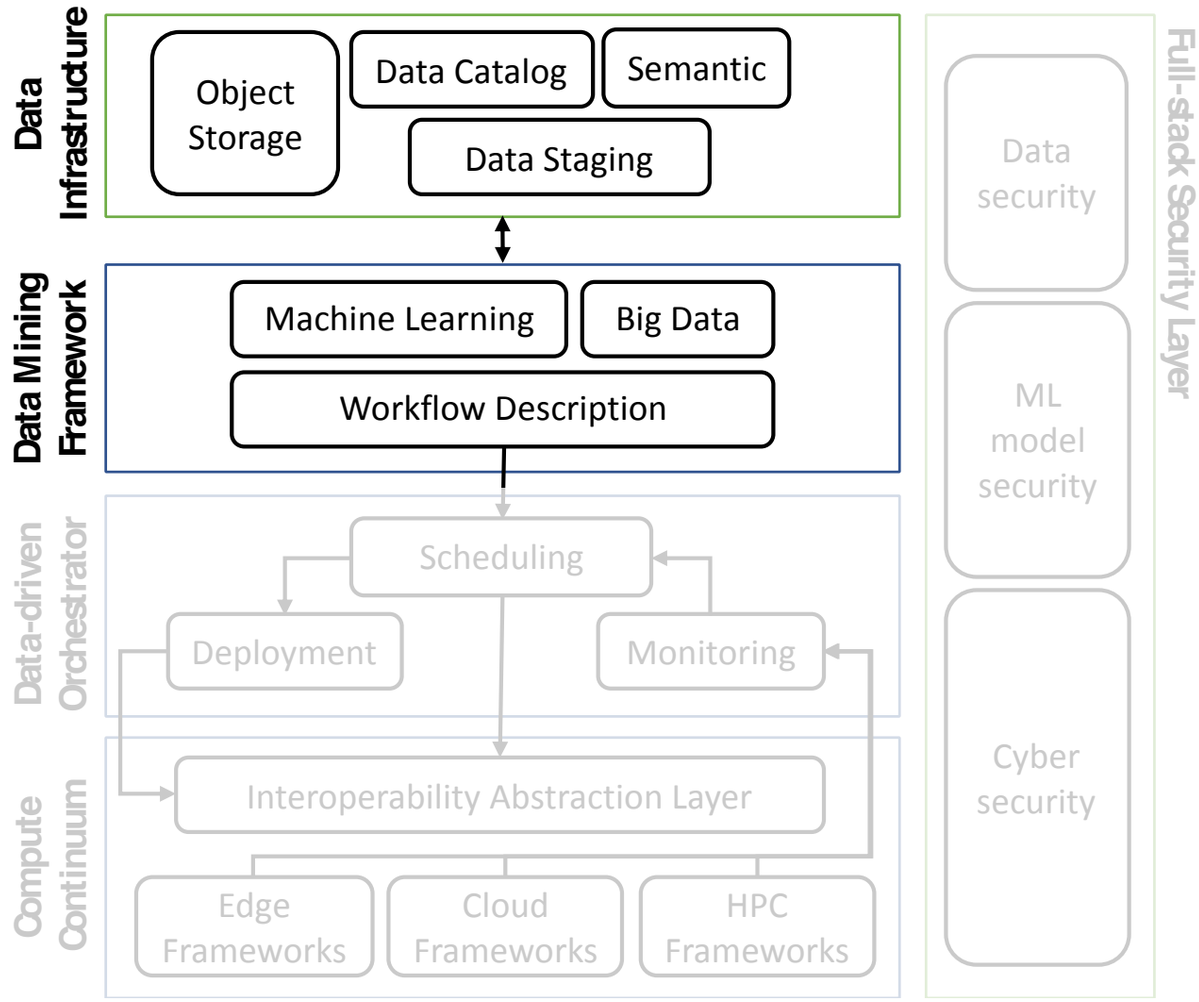
Data Infrastructure

- Object Storage
- Data Staging Engine
- Data Catalog
- Semantic engine

Data Mining Framework

- Complex workflows description
- Serverless approach
- Support to task- and data-based parallelism

Components
S3/InfluxDB
Dataplug
Nuvla
Virtuoso
COMPSs
Lithops
Pytorch
Kserve



Components

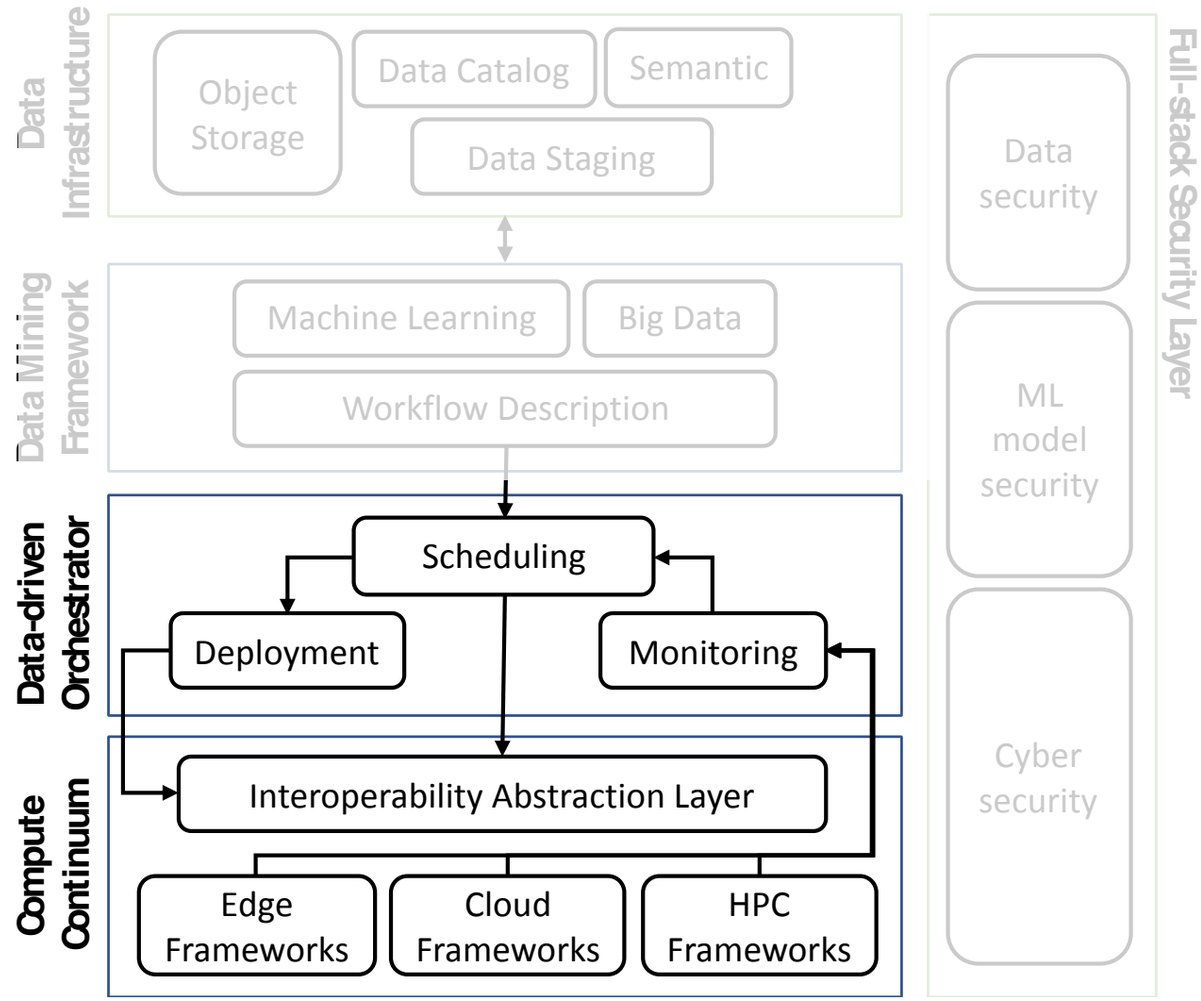
Data-driven Orchestrator

- Select computing resources for workflows based on monitoring

Compute Continuum

- Unified computing abstraction layer based on containers
- Programming paradigms optimized for edge, cloud and HPC

Components
COMPSs
Prometheus
Kubernetes
CUDA
SkyStore

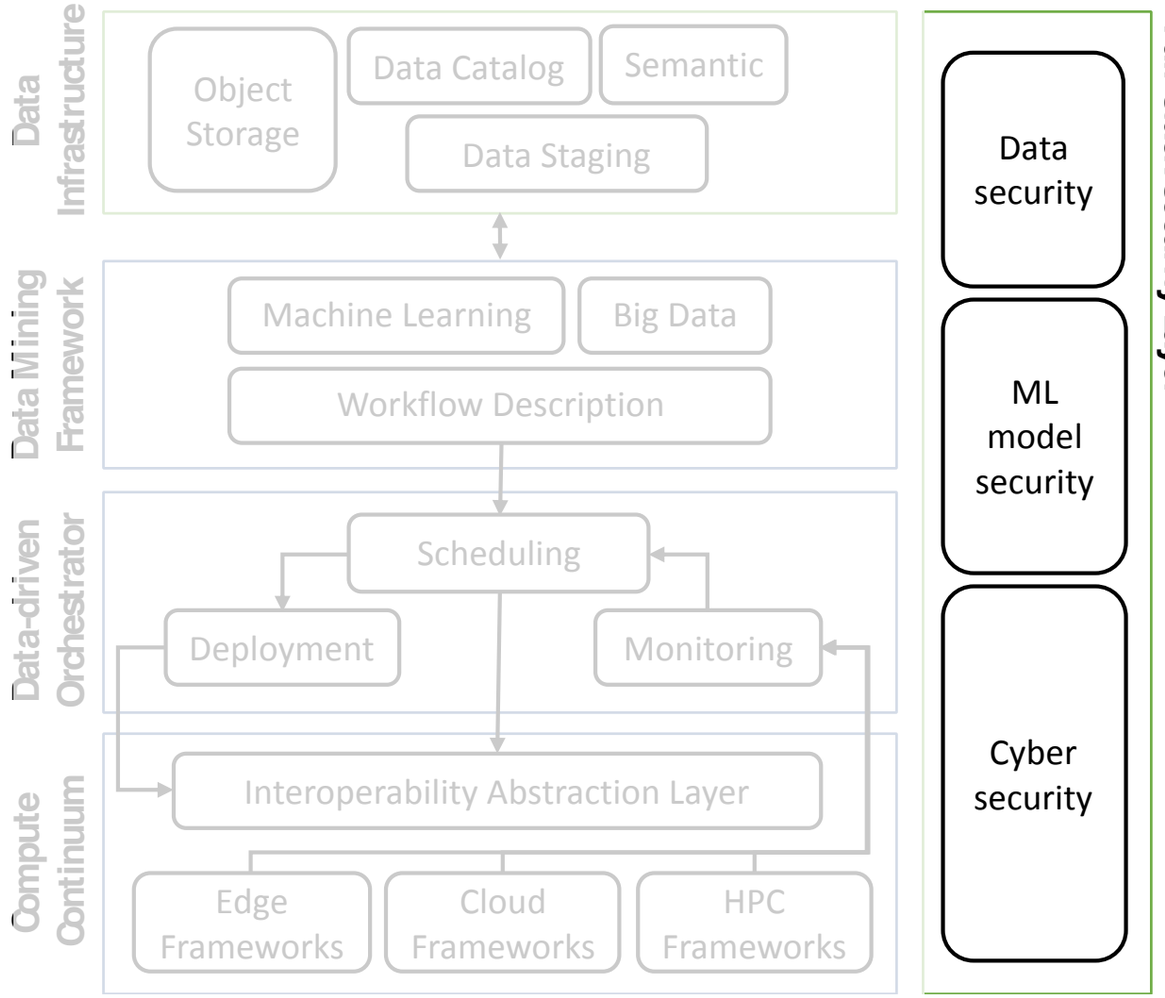


Components

Cybersecurity Capabilities

- Data protection, privacy and confidentiality
- AI models protection
- Authenticity and security for computing nodes
- Trustworthiness and verifiability of routines and libraries

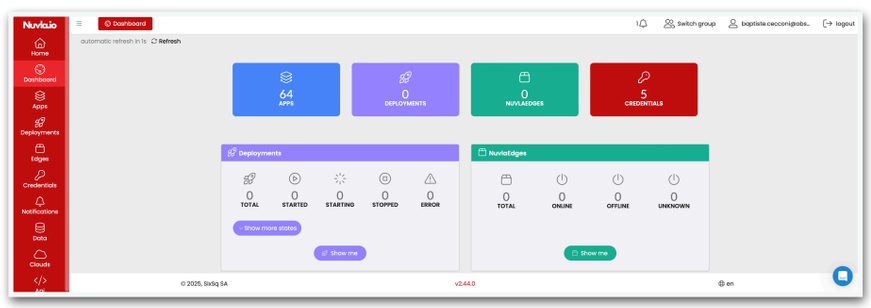
Components
Trivy
MultiParty Computing
Homomorphic Encryption



Data Catalog: Nuvla data management (DM)

Nuvla.io

https://nuvla.io/ui/



Elements:

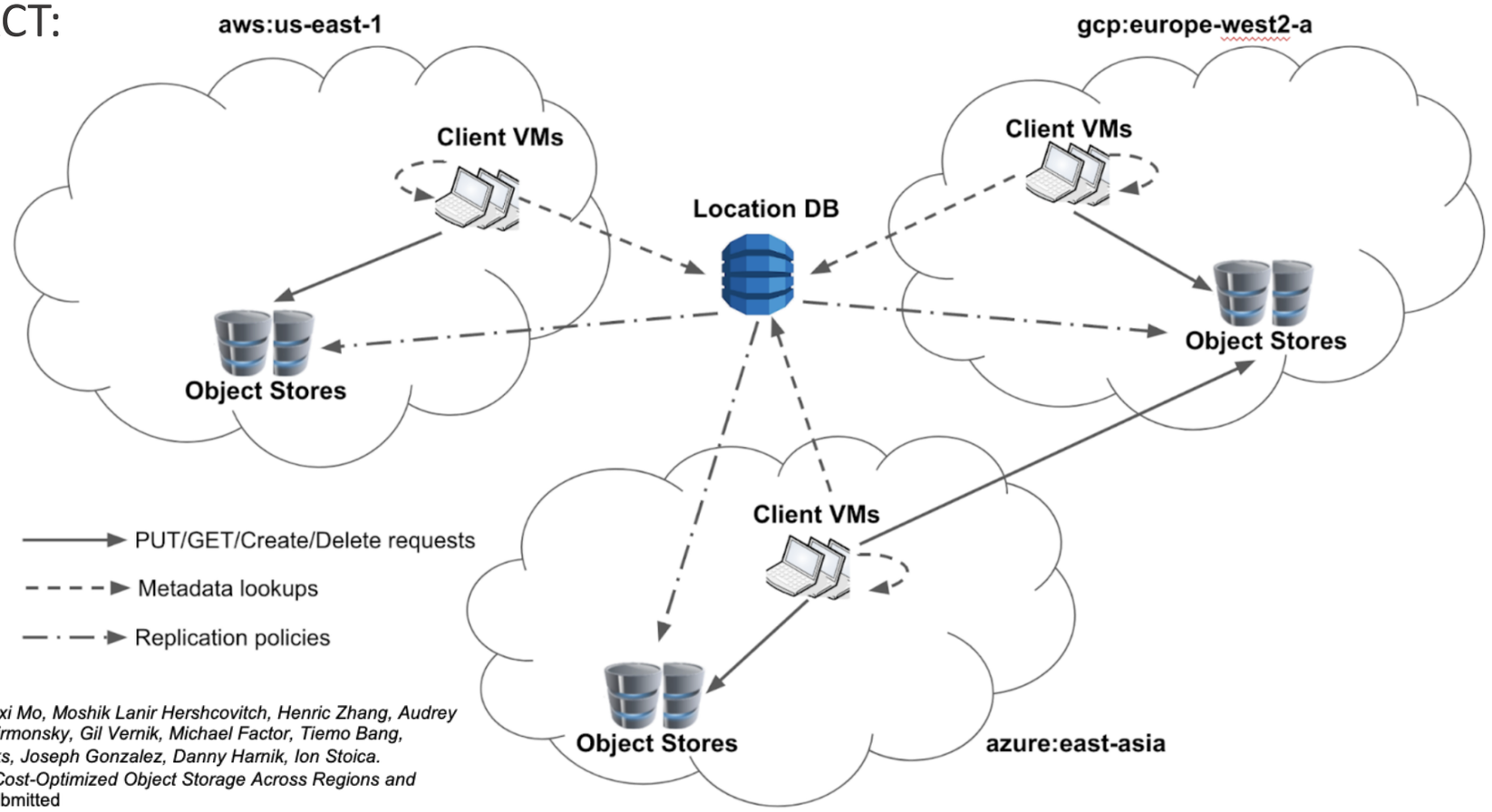
- *S3 infrastructure service*
  - endpoint and credentials
- *data object*
  - reference to S3 object on *S3 infrastructure service*
  - operations: create, obtain S3 pre-signed URL for upload, download, and delete
- *data record*
  - metadata about the *data object*
  - contains reference to the *data object* it describes
- *data set*
  - query against *data records*

Third-party app integration with DM workflow:

- Nuvla API (JSON over HTTPS) for data management
- Nuvla API python-library with data management examples
- Notifications to MQTT on data objects creation

SkyStore

- Joint project UC Berkeley - IBM
- Provides a **virtual global object store** namespace across clouds / premises
- Each client connects to local **S3-proxy**, which is connected to a nearby **S3 cluster** and to a central **location DB server**
- Data replication and consistency controlled via policies
- Specifically, remote objects can be automatically cached using closer/local object storage
- SkyStore work in Phase 2 of EXTRACT:
  - Matured base prototype
  - Joint paper (re-submission)

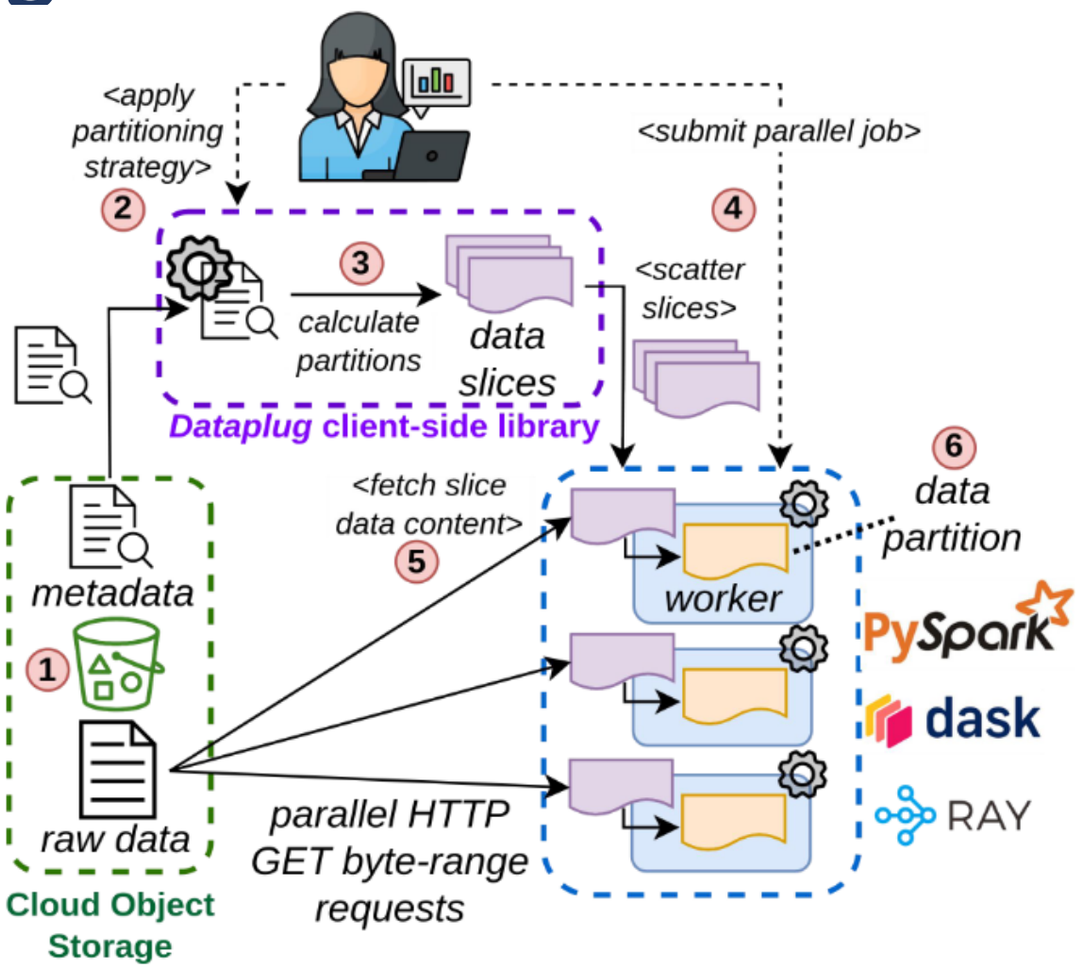


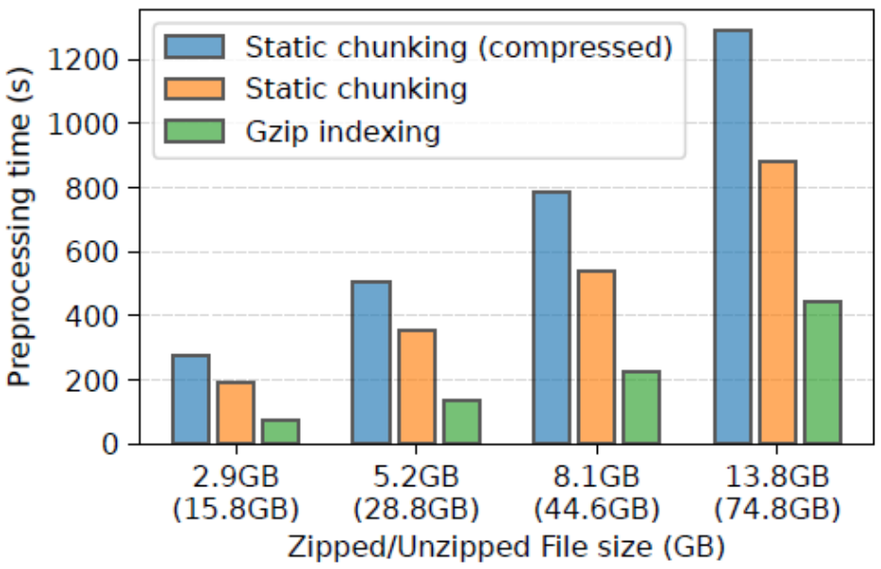
Shu Liu, Xiangxi Mo, Moshik Lanir Hershcovitch, Henric Zhang, Audrey Cheng, Guy Girmonsky, Gil Vernik, Michael Factor, Tiemo Bang, Natacha Crooks, Joseph Gonzalez, Danny Harnik, Ion Stoica. "GlobalStore: Cost-Optimized Object Storage Across Regions and Clouds". Re-submitted

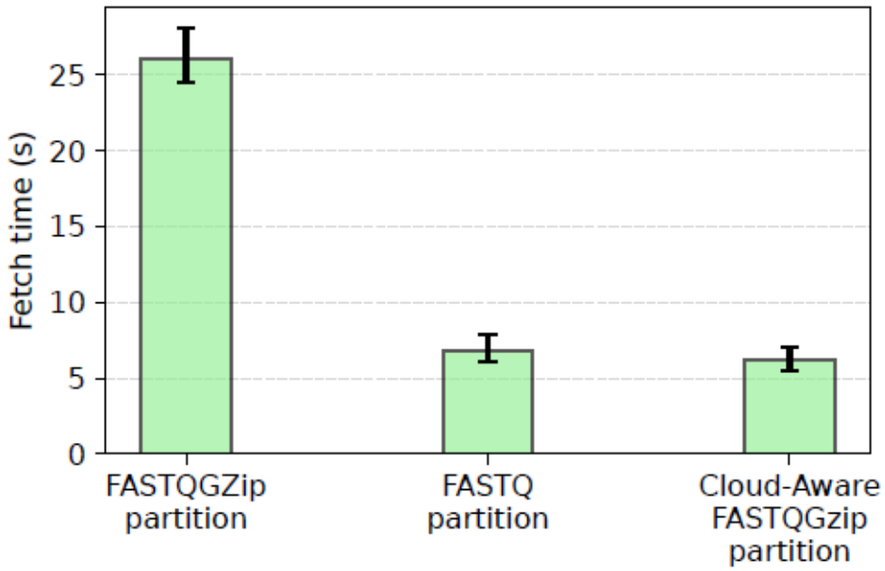
(Slide from Erez Hadad, IBM)

Dataplug dynamic data staging

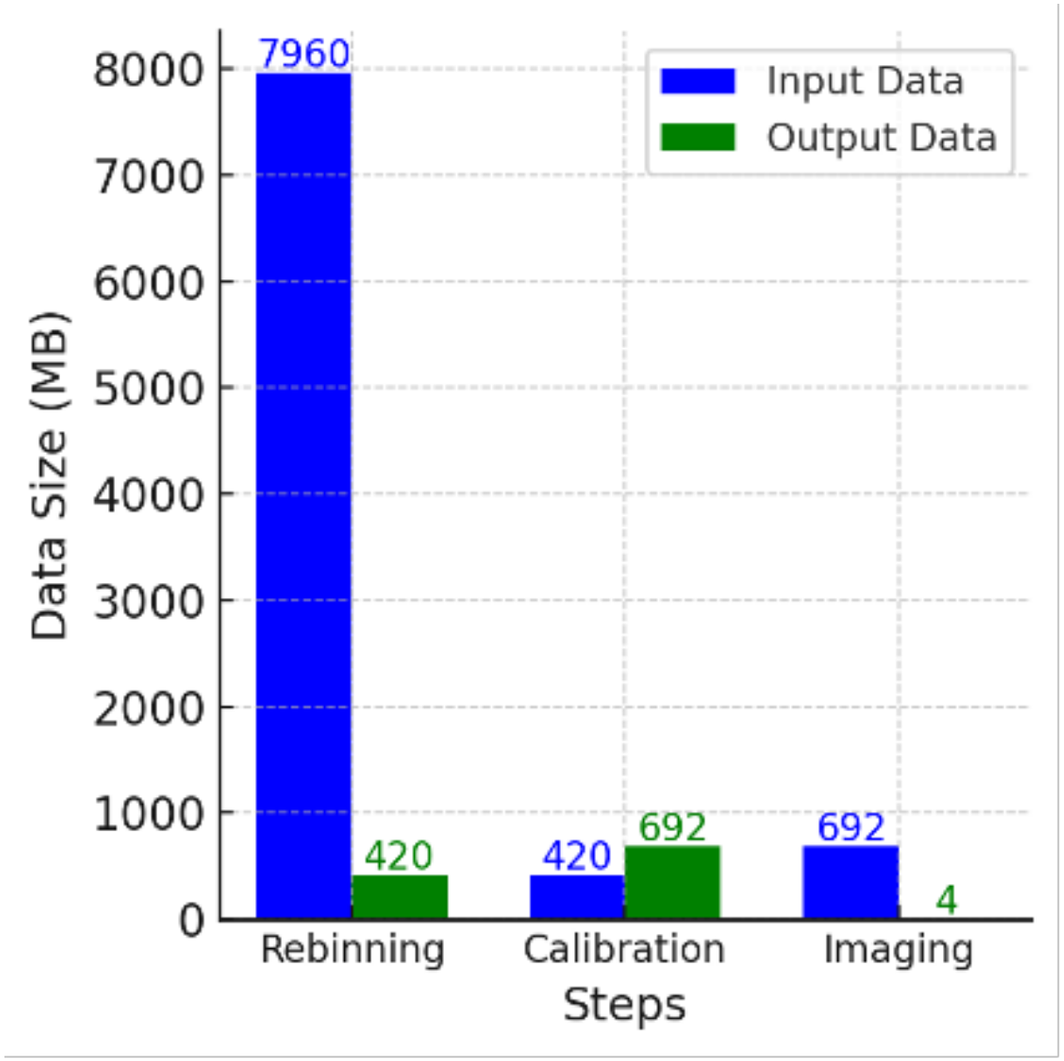
- **Dataplug:** extensible framework that implements on-the-fly data partitioning
- Hide complexities of pre-processing and partitioning unstructured scientific data
- Data-driven and dynamic, efficient parallel access to data
- Generate arbitrary data partitions without modifying existing data
- Extensible to multiple data formats
- **KPI 1.1**: faster partitioning (up to 65.6% less pre-processing time, and 3.7x in fetching partitions) and an important reduction of data transfers in staging



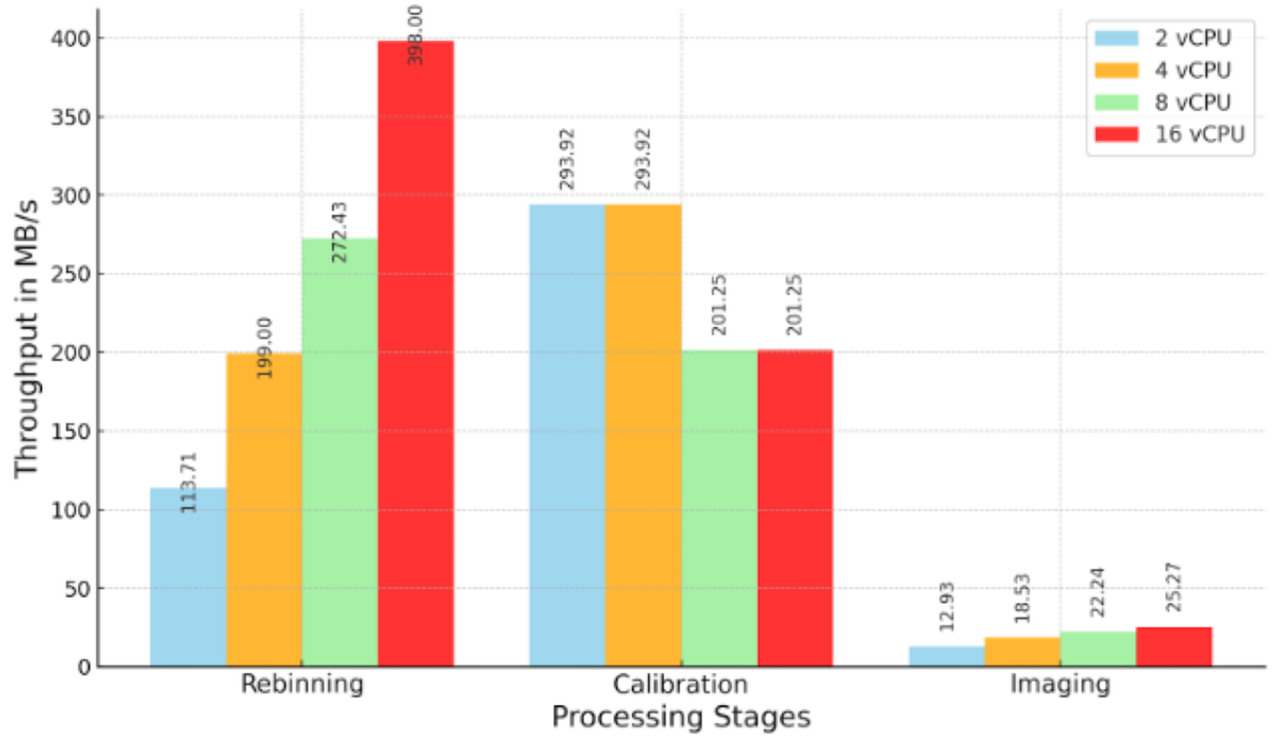




Data throughput analysis



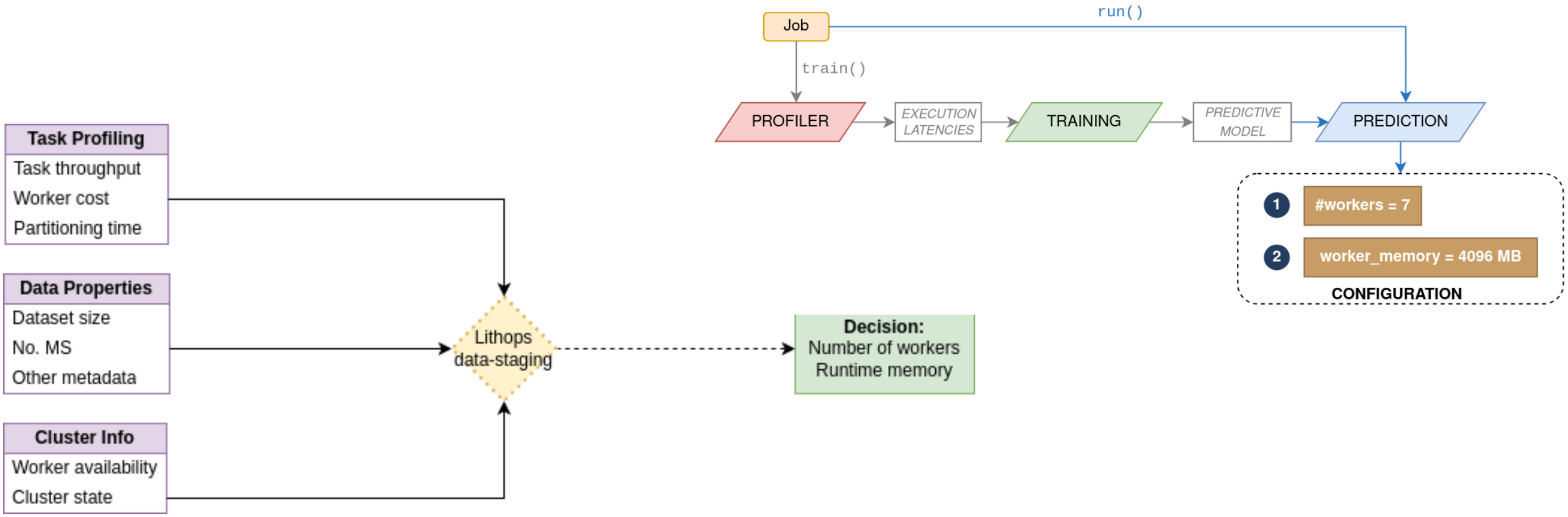
Data volumes per step (input-output



Throughput per step with different VCPU configurations.

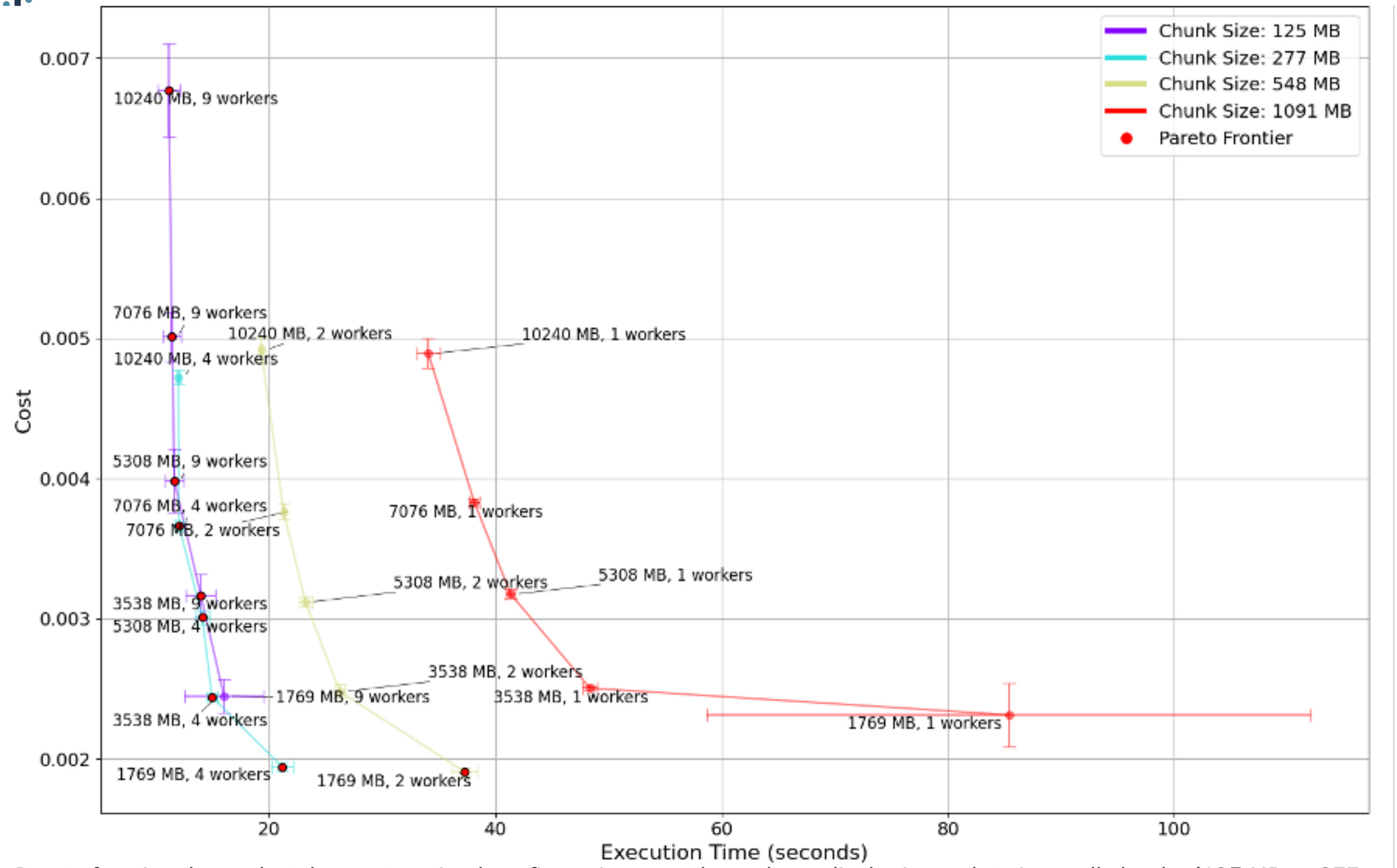
Lithops Smart Provisioning Tool

- **Smart provisioning:** new tool for Lithops to calculate the optimal number and size of workers for data staging tasks (WIP)
- Decisions to optimize job completion time and cost based on application performance, data size and transfer speed, and the cost of partitioning
- Working directly with TASKA use case
- **KPI 1.1 and 1.2:** enhancing user experience in developing extreme data processing workflows by abstracting compute resource configuration while optimising performance



Optimal resource allocation studies

1 (small) MS = 1090 MB



The Pareto frontier shows that the most optimal configurations are those that split the input data in small chunks (125 MB or 277 MB) and use multiple (small) workers to run the process