

Processing Cosmological data on ReAna4**PUNCH** Analysis Platform

Prateek Gupta

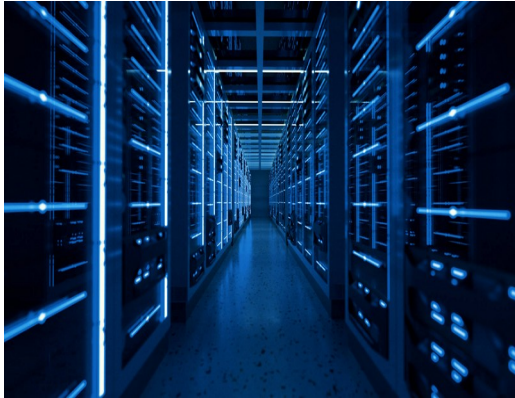
Thüringer Landessternwarte Tautenburg

4th TA 2 Mini Workshop in Göttingen
January 31, 2025



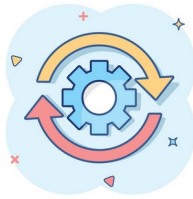
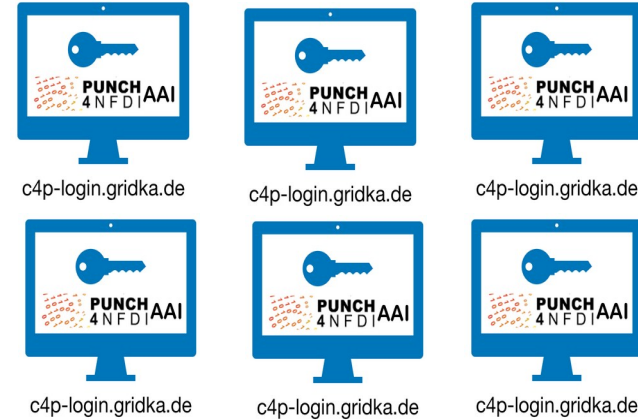
ReAna4PUNCH

Access to
Storage4PUNCH



Large Storage

Access to
Compute4PUNCH



Running
Analysis Pipelines/
Workflows
on container registries

ReAna4PUNCH Workflows

1) Generate source catalog from FITS image (using PyDBSF)

Workflow type: Serial

Data & Results: Storage4PUNCH

Computation@ AIP

2) Processing LoFAR data [tested only flux calibrator]

Workflow type: Serial & CWL

Data & Results: Storage4PUNCH

Computation@ AIP & Compute4PUNCH

3) Post-processing of simulated data [Single timestamp/snapshot]

Workflow type: Serial

Data & Results: Storage4PUNCH

Computation@ AIP & Compute4PUNCH

4) Galaxy Group/Cluster finder in galaxy redshift surveys [mock sample]

Workflow type: Serial

Data & Results: Local machine

Computation@ AIP

Experience as a ReAna4PUNCH User

- Workflows has pre-defined format (reana.yaml)
Education required ---> working with CWL-workflows
- Data Transfer to the ReAna workspace is required !!
- Use of multiple-containers in a single (serial) workflow
- To access Storage4PUNCH “davix” / “curl” / “gfal”
More detailed documentation required on their usage
- Update the docker container registry, if new library is required
- What about performing cosmological simulations?

May I ask for Additional Features ?

Missing:

- 1) Submitting CWL workflows on compute4PUNCH backend
- 2) Running jobs on multiple-nodes !!

Suggestion:

Self-test of the workflow is required:

Workflows with some libraries that are missing in the mentioned container, should not be in queue at the very first place

Idea: Option to listout libraries somewhere in the reana.yaml file, even in case of using a multiple containers



Thank you