

PUNCH4NFDI Infrastructure: The Platform for the analysis of Astrophysical Data

September 8, 2026

Presented by

Prateek Gupta

Thüringer Landessternwarte Tautenburg



PUNCH4NFDI

A Consortium in the NFDI

Nationale ForschungsDaten
Infrastruktur

P articles
U niverse
N uClei
C
H adrons

Particles, Universe,
NuClei and Hadrons
for the NFDI

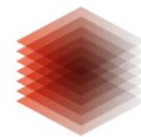
A consortium in the NFDI.



- FAIRification: data & workflows
- "big data" and "open data"
- heterogeneous data & infrastructures
- transfer of knowledge

Efficient usage of DATA and resources
with minimum technical barriers

Meet the team



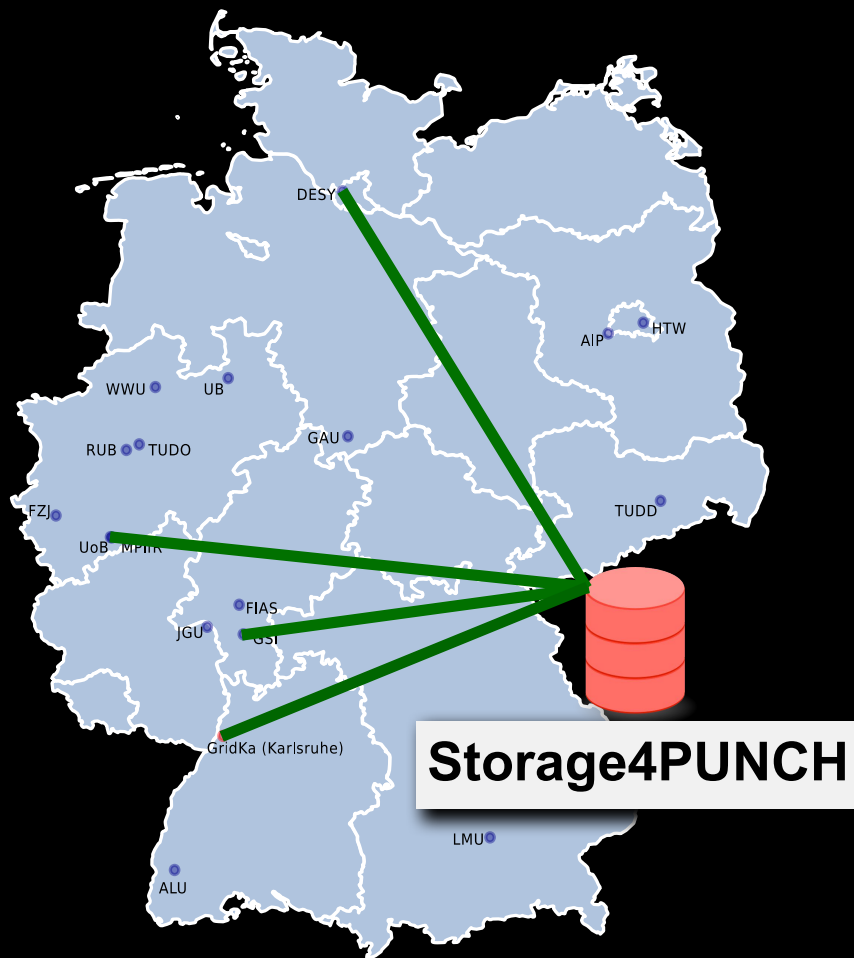
Storage4PUNCH

Integrated 4 Storage sites

Based on **two storage technologies**:

- **dCache** (@ DESY and KIT)
- **XrootD** (@ U Bonn and GSI)

Storage in PUNCH 1.0: ~ **4 PB**



Compute4PUNCH

Dynamic integration of various heterogeneous compute resources

Acquire resources :

- dynamically by CoBalD/TARDIS

Integrated →

HTC, GPU, Kubernetes cluster (ReAna)

Container technology :

- software, modules, libraries encapsulated in a container; lightweight isolated environment



Sufficient Storage & Compute resources !

Way more than the resources in personal laptops/desktops

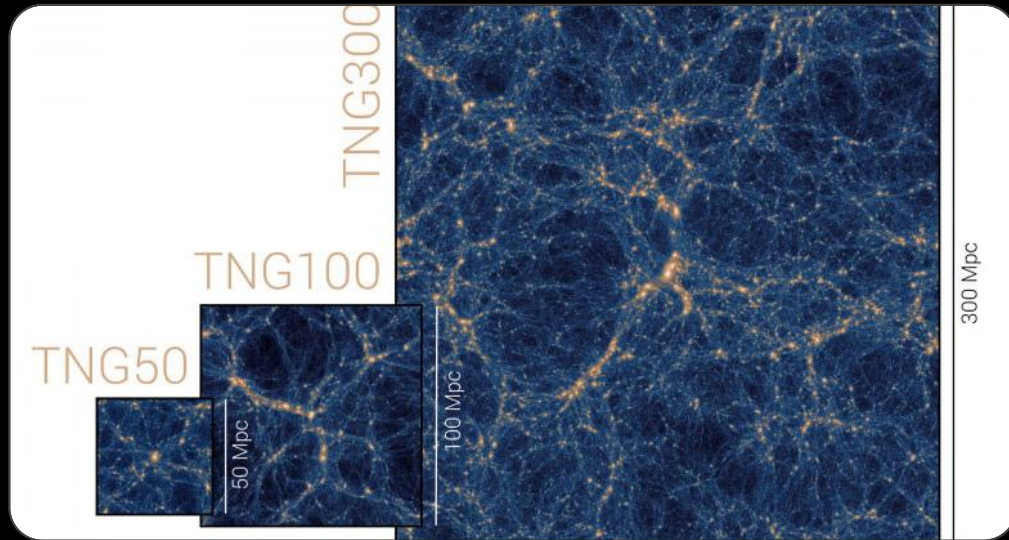
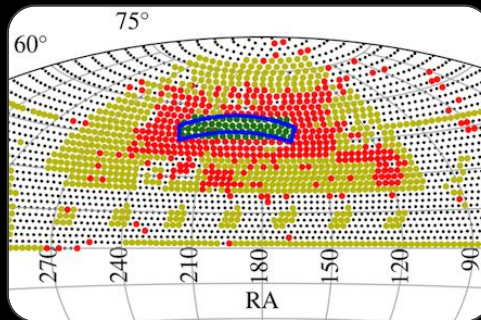
Can Large astrophysical data ($>$ a few TB) be processed on PUNCH resources?

Let's exploring with different kind of data sets

Large Astronomical Data

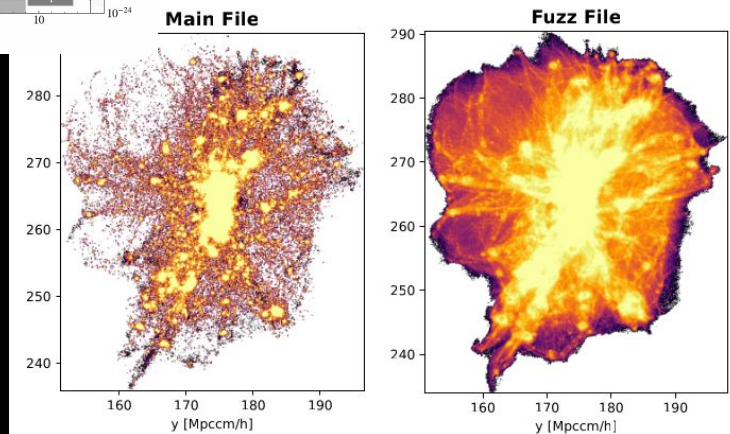
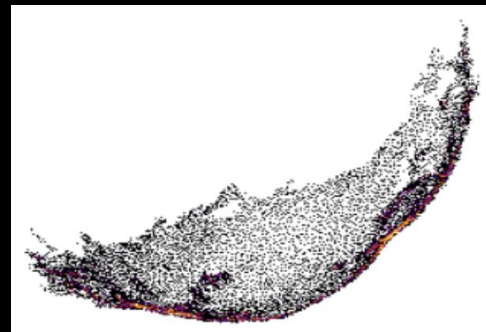
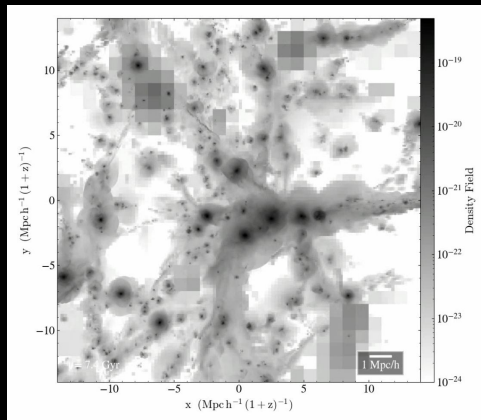
Various **Radio Sky Surveys** and **Large Cosmological simulations** are conducted in order to **understand** the **formation history** of the **Universe**

- This leads to significant amount of data ~ a few 10's of PB
 - ~ 500 GB to 1.5 TB / Snapshot
 - ~ 1.5 TB / observation hours (LOFAR)
 - ~ 2.5 TB / observation hours (MeerKAT)



1. Post-processing Cosmological Simulations

- Public Simulation data:
TNG, TNG-Cluster
- Private Simulations
- Simple slice or projection plots
- Volume rendering or time series
- Running halo finder algorithms
- Implementing semi-analytical models to derive new fields from simulated fields



2. Processing LoTSS data

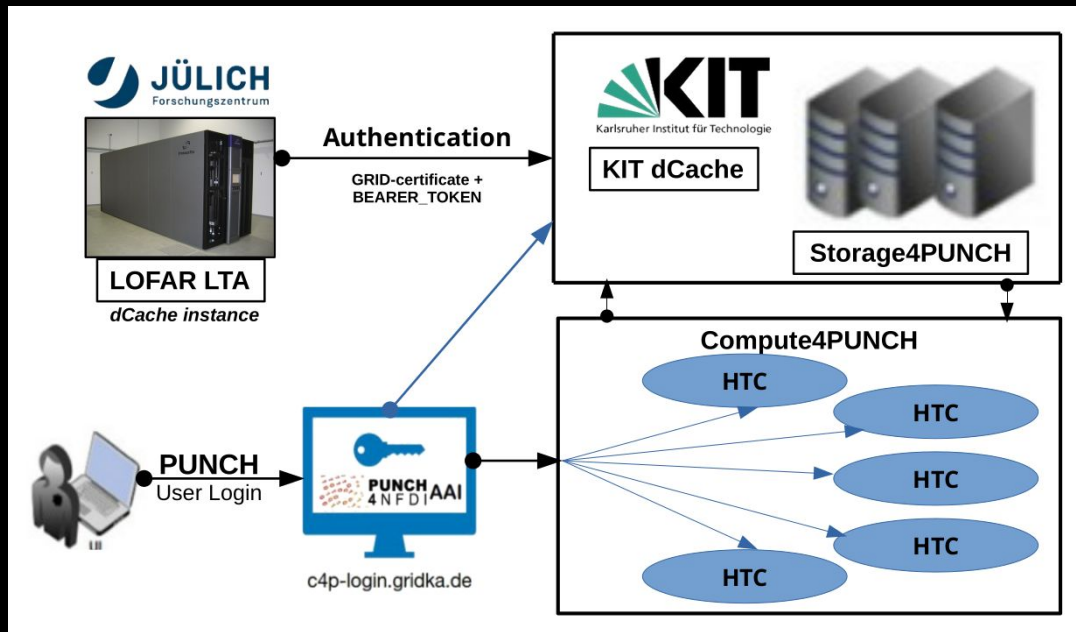
The Largest radio sky survey ~ 46% of the sky

→ Significant amount of data:
~ 1.5 TB / observation hours (LOFAR)

Run Initial Calibration pipeline (LINC):

- To correct for various instrumental and ionospheric effects
- To prepare data for further scientific processing/analysis.

Aims to produce high-resolution images ~ 5" and sensitivity ~ 100 μ Jy/beam



LOFAR data processing Pipeline on PUNCH resources

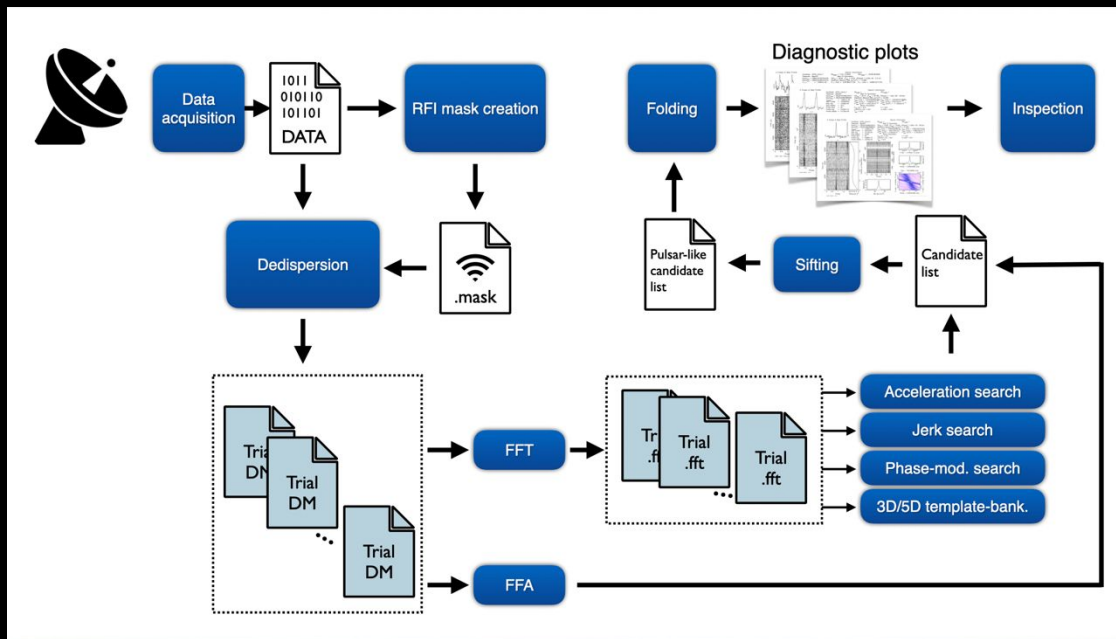
3. Searching Pulsars in MeerKAT data field

One of the key science project of MeerKAT

→ Significant amount of data:
~ 2.5 TB / observation hours
(MeerKAT)

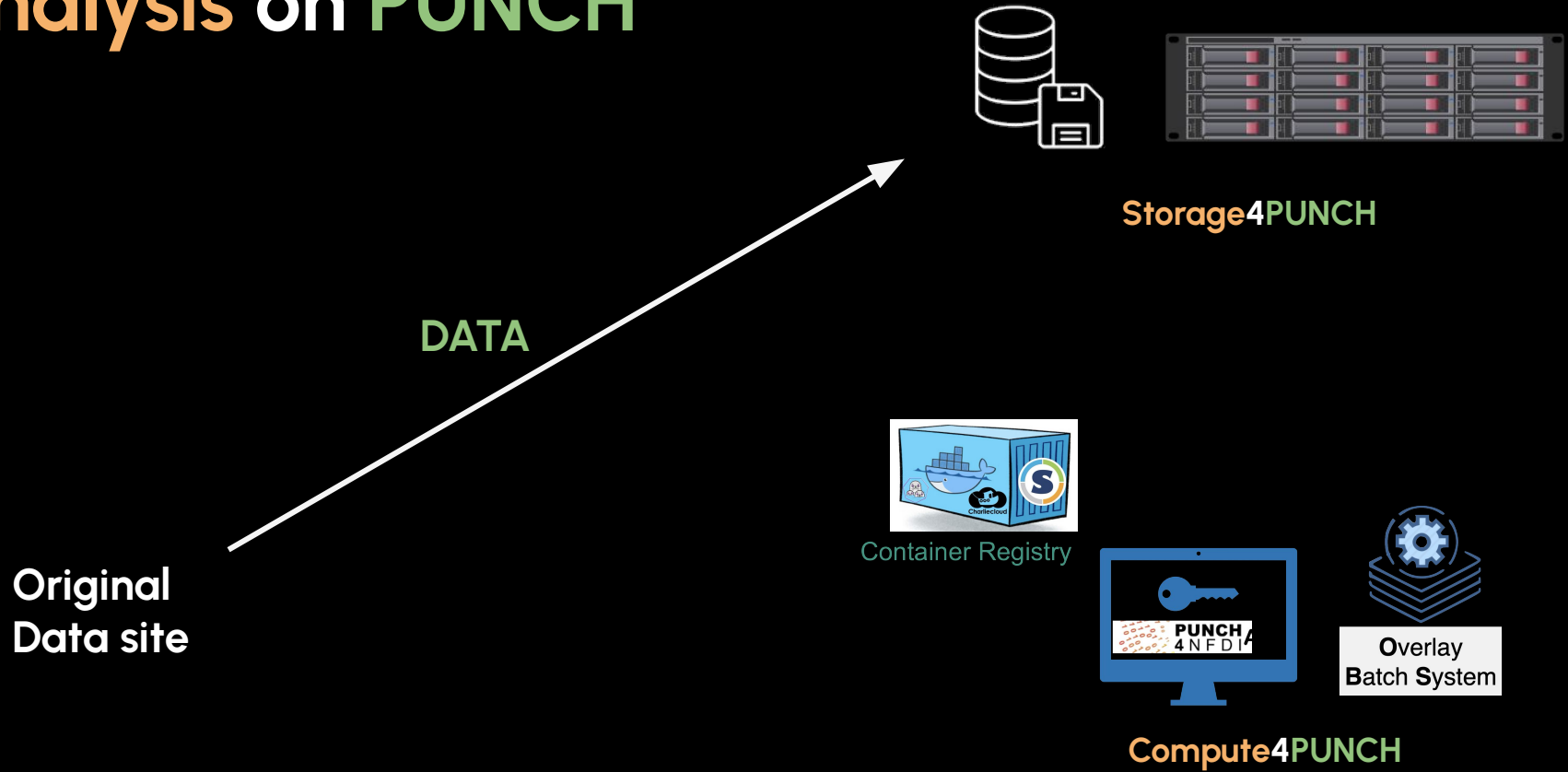
Pipeline processing requires HTC
and GPU infrastructure:

Heavily accelerated Fast-Fourier
Transformation



Pulsar Search Pipeline

Analysis on PUNCH



Analysis on PUNCH



Storage4PUNCH

Original
Data site

DATA



Container Registry



Compute4PUNCH



Overlay
Batch System

Analysis on PUNCH

Preferred:

No breakdown in the
retrieval of data at C4P

DATA

Original
Data site



Storage4PUNCH



Container Registry



Compute4PUNCH



Overlay
Batch System

Analysis on PUNCH

Preferred:

No breakdown in the retrieval of data at C4P

Original
Data site

DATA



Storage4PUNCH



Container Registry



Compute4PUNCH



Overlay
Batch System

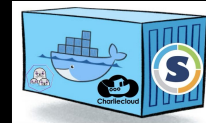
DATA

Analysis on PUNCH

Compute environment ?



Storage4PUNCH



Container Registry



Overlay
Batch System

Compute4PUNCH

Analysis on PUNCH

Compute environment ?

PUNCH GitLab
Container Registry,
hosted at AIP

The unpacked image of containers made
available to the **Compute4PUNCH** via
CVMFS filesystem



Storage4PUNCH



Container Registry



Compute4PUNCH



Overlay
Batch System

PUNCH GitLab Container Stack Repository

Compute4PUNCH / Container Stacks

C Container Stacks

main container-stacks + Find file Code

Merge branch 'alex-main-patch-65b8' into 'main'
 Manuel Giffels authored 3 months ago

Name	Last commit	Last update
astro-ml	Merge branch 'arm2arm-main-patch-65b8' into 'main'	10 months ago
caraca_container	Reverted change by commenting o...	3 years ago
caraca_container_new_stimela	Update file Dockerfile	3 years ago
h4leptons	Latest version of H4Leptons analysis	2 years ago
htcondor-wn	Fix issue with trailing slash on toke...	1 year ago
linc-wn	Update Dockerfile for LINC	3 months ago
lqcd_analysis	Merge branch 'ding-ze.hu-main-patch-65b8' into 'main'	2 years ago
presto	Prateek main patch 38033	5 months ago
simulation-wn	Prateek main patch 38033	5 months ago
snakemake-wn	Add apptainer to snakemake-wn	3 years ago

Users can create their own containers:

- Provide a Dockerfile via the Container Stack Repository
- Within a day or so, the container will be available on Compute4PUNCH

An upcoming update to Container Registry:

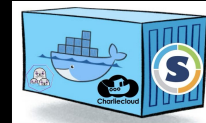
User will be able to version their containers.

<https://gitlab-p4n.aip.de/compute4punch/container-stacks>

Analysis on PUNCH



Storage4PUNCH



Container Registry



Overlay
Batch System

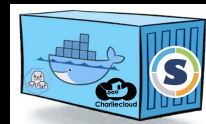
Compute4PUNCH

Analysis on PUNCH



Storage4PUNCH

How to access
these distributed
resources?



Container Registry



Overlay
Batch System

Compute4PUNCH

Analysis on PUNCH



Storage4PUNCH

How to access
these distributed
resources?

PUNCH uses TOKEN-based
authentication:

Obtain a TOKEN and access these
resources



Container Registry



Overlay
Batch System

Compute4PUNCH

PUNCH-AAI

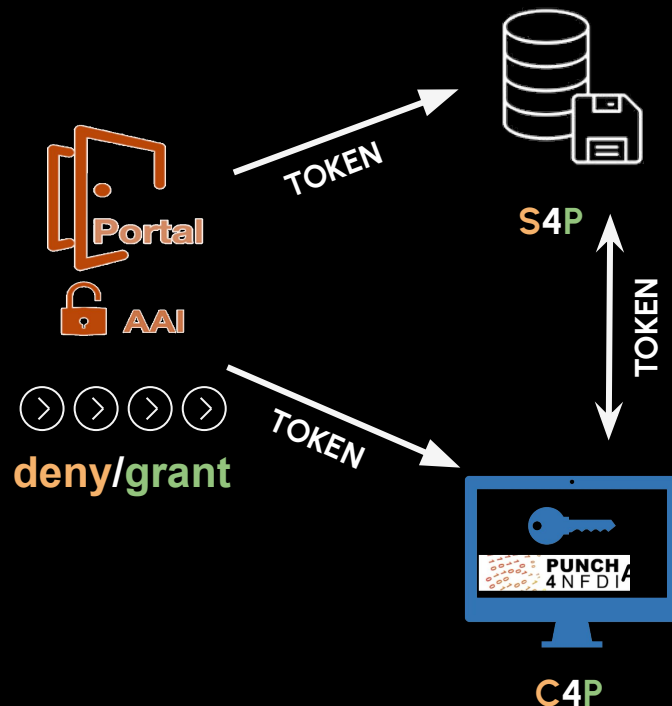
Authentication and Authorization Infrastructure

Authentication:

- ❖ to securely access distributed resources
- ❖ using existing institutional credentials
- ❖ a Single Sign-On mechanism

Based on Helmholtz AAI:

- implemented using the UNITY AAI toolkit
- provided by HIFIS





PUNCH ecosystem



Storage4PUNCH



Container Registry



Overlay
Batch System

Compute4PUNCH



PUNCH ecosystem



Storage4PUNCH



USER



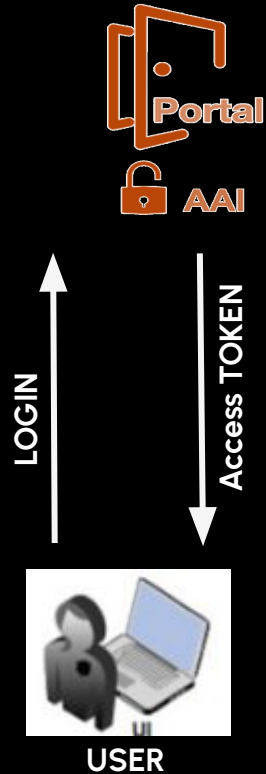
Container Registry



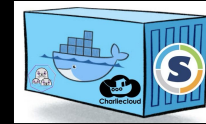
Overlay
Batch System

Compute4PUNCH

PUNCH ecosystem



Storage**4**PUNCH



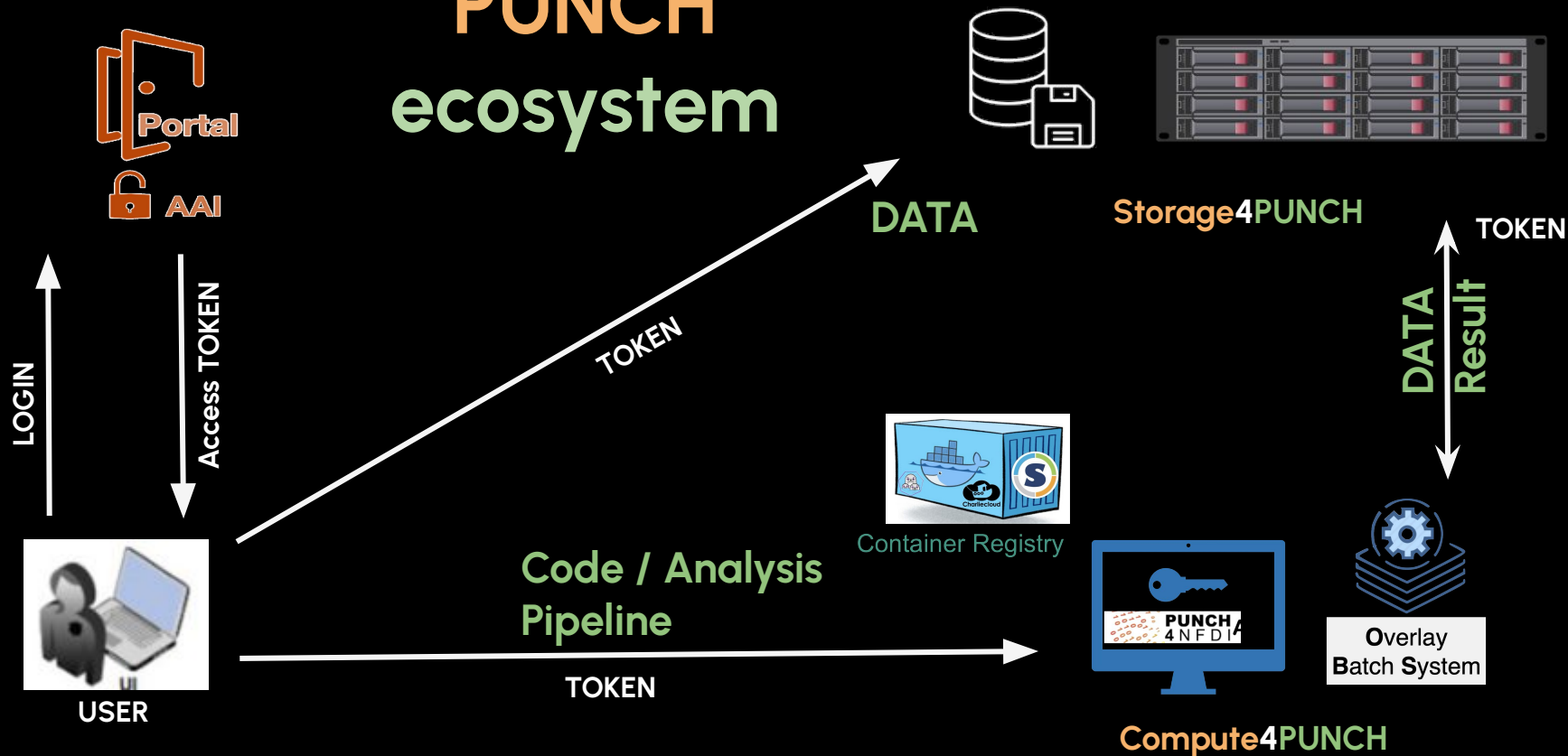
Container Registry



Overlay
Batch System

Compute**4**PUNCH

PUNCH ecosystem



DRP-Hub

Discover and manage research tools and workflows

Key Features:

- 1). Job submission on C4P or REANA via DRP-Hub
- 2). Each submission will create a card
- 3). One can keep this card private or public
- 4). Even, can be shared with the collaborators to reproduce the results or for doing same analysis with different data or parameters.

The screenshot displays the DRP-Hub web interface. On the left is a sidebar with sections: DRP-Hub (Welcome back, prateek), NAVIGATION (Dashboard, Browse, Bookmarks), TOOLS (Open console), and SECRETS & CONNECTIONS (SSH Keys, Git Config, S3 Storage, Reana, C4P). The main content area is titled 'DRP-Hub' with the subtitle 'Discover and manage research tools and workflows'. It includes a search bar, filters for 'Categories' and 'Tags', and a 'Sort: Newest' option. Below this, it states '24 DRPs available (12 public, 12 internal, 0 private)'. A grid of eight cards is shown, each representing a research tool or workflow. Each card includes a logo, title, author, category, description, and tags. The cards are: 'CERN Open Data with C4P' (hep), 'FIAS 2026 Frankfurt' (general), 'Key4hep tutorial' (hep), 'Cosmology data analysis ex...' (astro), 'Halo Mass Distribution' (astro), 'StarHorse interactive' (astro), 'REANA Environments' (general), and 'REANA Tutorials' (general). Each card also shows a heart icon for likes and a share icon.

Summary and Conclusion

PUNCH infrastructure can serve as an ideal platform for large astrophysics data processing;

- making it a valuable resource for researchers in the field

Easy access:

- No new credentials are required
- Use existing institutional ones

PUNCH ecosystem encourage reproducibility in the research environment

Thank you



© 2023–2026, Prateek Gupta & PUNCH4NFDI, No Right Reserved.

Acknowledgment:

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure (NFDI) funding programme, using project number 460248186 (PUNCH4NFDI)

Contact:

Prateek Gupta

Thüringer Landessternwarte, Tautenburg

Email: prateek@tls-tautenburg.de

