Karlsruher Institut für Technologie

Latest Developments of the PUNCH4NFDI Compute and Storage Infrastructures

CHEP 2024 - Cracow, Poland

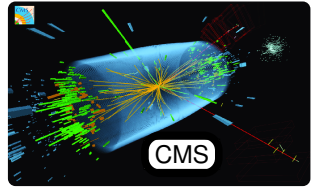
Benoit Roland (KIT), Harry Enke (AIP), Oliver Freyermuth (Uni Bonn), Manuel Giffels (KIT), Matthias Hoeft (TLS), Michael Huebner (Uni Bonn), Arman Khalatyan (AIP), Christoph Wissing (DESY)

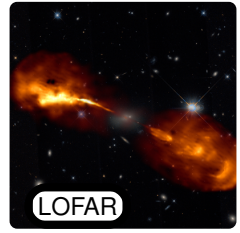
What is PUNCH4NFDI?

Consortium within the **NFDI** - National Research Data Infrastructure in Germany

Particles, **U**niverse, **N**u**C**lei and **H**adrons for the **NFDI**

- From elementary particles to large scale structures
- **Similar challenges** with large data volume
- **Different expertise** in dealing with it





Setup a **federated** and **FAIR** science data platform

- Provide **infrastructures** to process and store **data**

  → Latest news about storage, compute and AAI

- Provide **data portal** to build and re(use) **research products**

  → Integration into analysis platform REANA

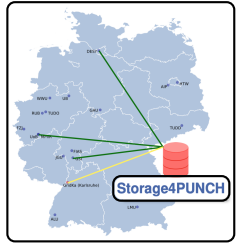
PUNCH4NFDI communities

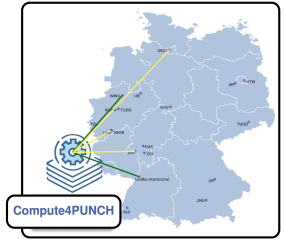
Institutions committed to providing compute and storage resources

Universities and research centers throughout Germany

Infrastructure pillars

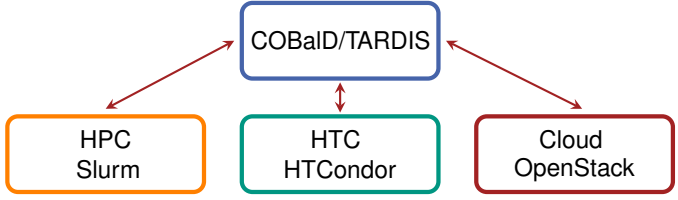
- PUNCH4NFDI has to cope with increasing amount of data and compute need
- Integrate available **storage** resources into a federated storage platform
  → **Storage4PUNCH** (Bonn, DESY, GSI, KIT)
- Federate available **compute** resources
  → **Compute4PUNCH** (Göttingen, LMU, LRZ Cloud, KIT, WWU)
- Provide **transparent access** to these resources
  → **Single entry point**
- **Secure** access to these resources
  → **Token-based Authentication**





Compute4PUNCH key ingredients

- Heterogenous compute resources with different schedulers, architectures, operating systems
- Aggregate resources in a single **Overlay Batch System** based on **HTCondor**
- Provision resources to the OBS using **COBalD/TARDIS meta-scheduler**
  → Single federated pool with dynamic extension on a user-demand basis
- **TARDIS** integrates available resources into OBS
- **COBalD** does balancing, job to resource matching, ensures effective usage

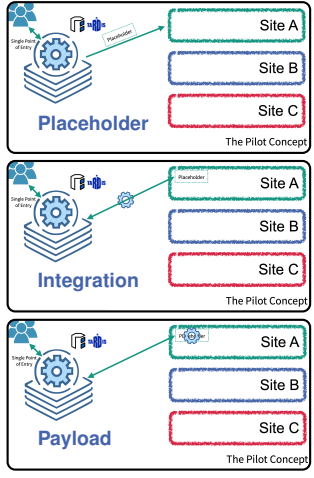


Resource provisioning and payload assignment

- Placeholder job - a drone - is submitted to the site
- Drone allocates and integrates resource into the OBS and provides a specific environment via a container
- Payload pulled from the OBS and assigned to the resource

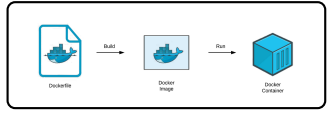
User interacts with the OBS through a single entry point:

- Login node - provided
- JupyterHub - expected soon
- OIDC token-based authentication

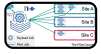


Typical analysis workflow

- Containerise analysis software
- Make it available in PUNCH4NFDI container registry
- Distribution via the CVMFS file system



- Job submission from login node 

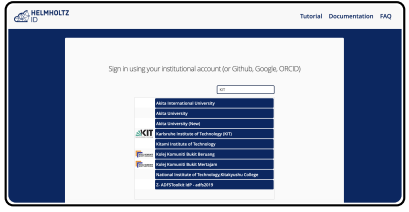
- HTCondor worker node started 

- Job executed with container retrieved from CVMFS

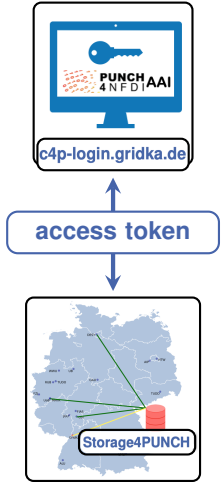
```
[root@c4p-login ~]# ls /cvmfs/unpacked.cern.ch/gitlab-p4n.aip.de\:5005/compute4punch/container-stacks/
caracal_container:latest          htcondor-wn:latest          simulation-wn:latest        wltg-wn:latest
caracal_container_new_stimela:latest  lincwn:latest          snakemake_arcane:latest
h4leptons:latest                  lqcd_analysis:latest        snakemake-wn:latest
```

Storage4PUNCH

- Endpoint at DESY and KIT using dCache
- Endpoints at Bonn and GSI using XRootD
- WebDav/XRootD as transfer protocol
- **Token-based access using Helmholtz AAI**



- Register once to Helmholtz AAI provider
- Once registered - access token generated with oidc-agent
- Storage4PUNCH accessed from login node



Access token

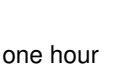
- Helmholtz AAI access tokens are mandatory to access Storage4PUNCH from Compute4PUNCH worker nodes

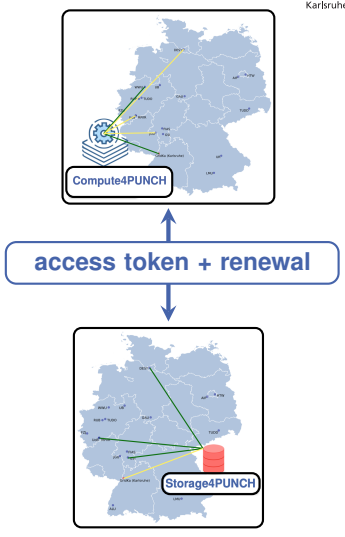
→ Have to be **inserted** in the HTCondor job

- Access tokens have a limited life time of about one hour

→ Have to be **renewed** for jobs exceeding their life time

→ **Need a mechanism to renew access tokens on-the-fly on the worker nodes**





Access token management

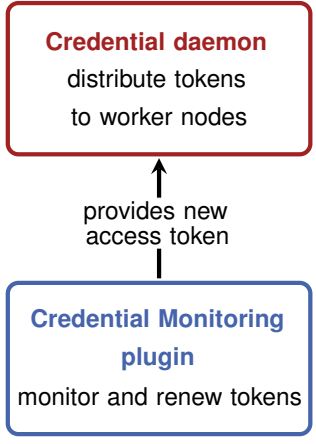
HTCondor provides a **Credential daemon (Credd)**

- **Agnostic** to type of tokens
- Responsible for their **distribution**

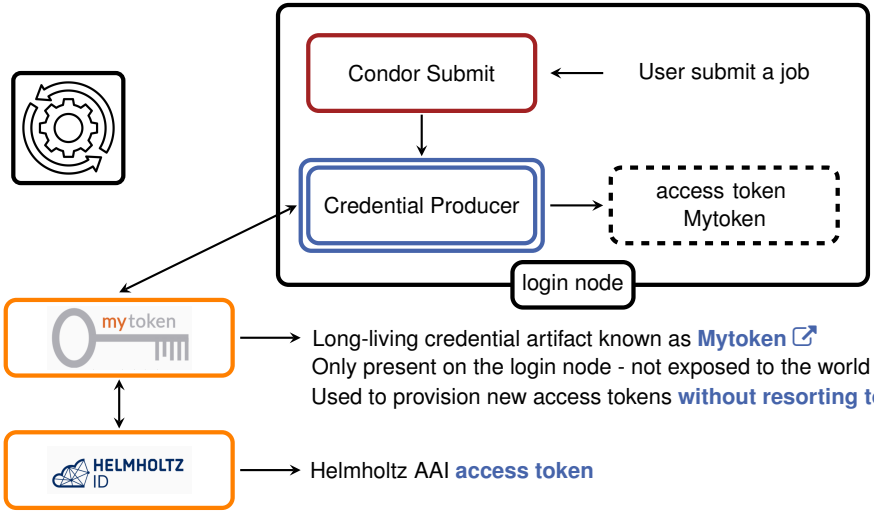
This daemon relies on **Credential Monitoring plugins (CredMon)**

- **Specific** to type of tokens
- Responsible for their **monitoring** and **renewal**

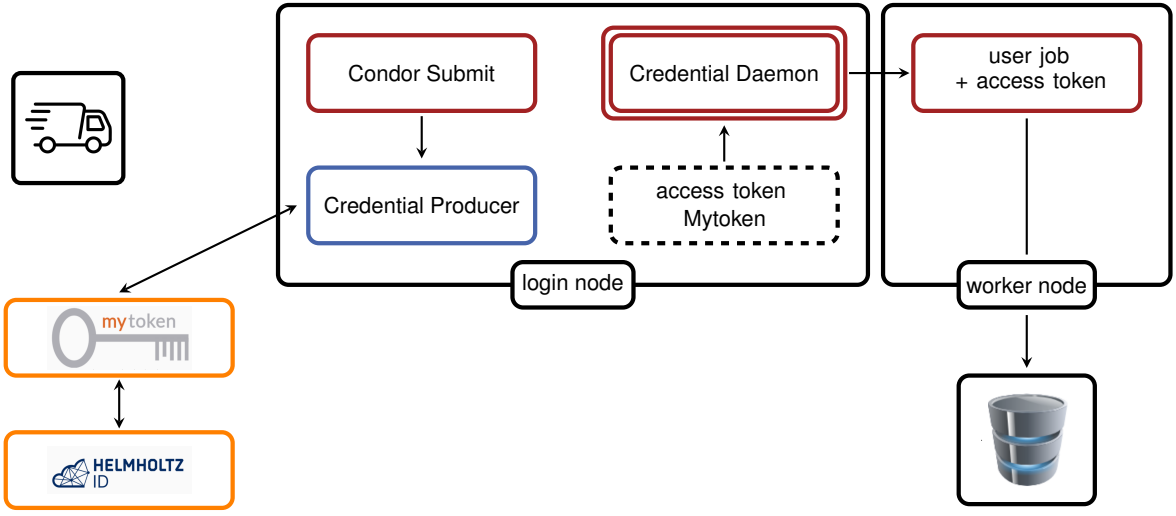
→ **Specific CredMon plugin to monitor and renew Helmholtz AAI access tokens**



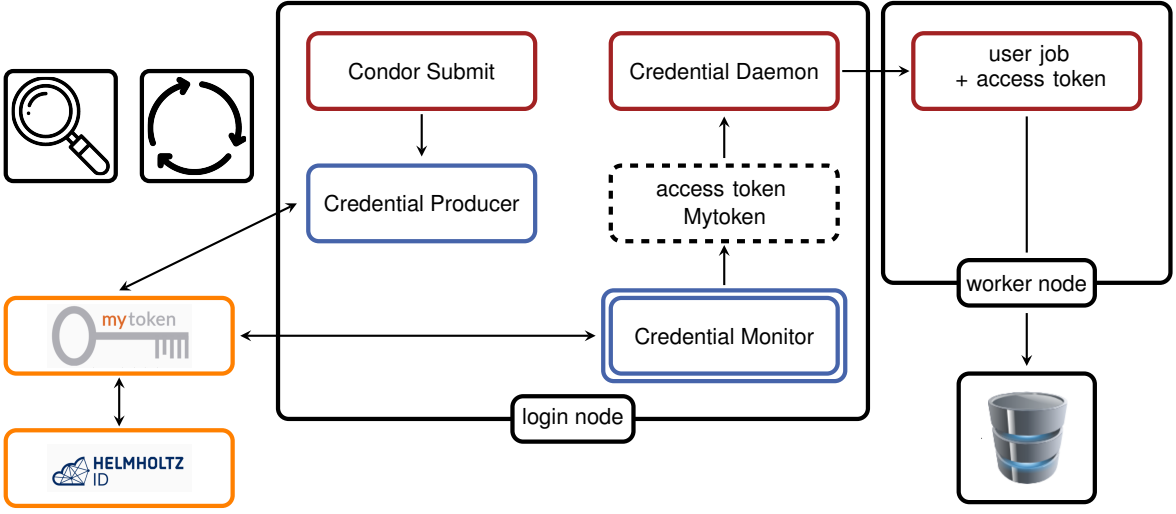
Production


Condor Submit
User submit a job
Credential Producer
access token
Mytoken
login node
mytoken
Long-living credential artifact known as Mytoken
Only present on the login node - not exposed to the world
Used to provision new access tokens without resorting to the user
HELMHOLTZ ID
Helmholtz AAI access token

Distribution



Monitoring and Renewal



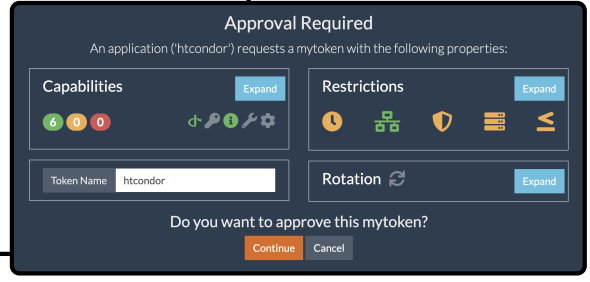
Job configuration

```
executable = analysis.sh
use_oauth_services = helmholtz  → Only need to specify AAI prodiver
should_transfer_files = YES
output_destination =
helmholtz+https://dcache-desy-webdav.desy.de:2880//pnfs/desy.de/punch/analysis
output = logs/$(machine).$(cluster).$(process).out
error = logs/$(machine).$(cluster).$(process).err
log = logs/$(machine).cluster.log
request_cpus = 1
request_memory = 256
+SINGULARITY_JOB_CONTAINER = "analysis:latest"
queue 1
```

Job submission

```
condor_submit analysis.jdl
Submitting job(s)
Hello benoit_roland!  You are going to submit your HTCondor jobs.
No credential has been found!
Please visit the following url in order to generate your credential:
https://mytoken.data.kit.edu/c/XWSdsgNC
Starting polling and waiting for your approval...
Your credential has been successfully created!
Its remaining life time is 6 d 23 h 58 m 46 s.
1 job(s) submitted to cluster 1142
```

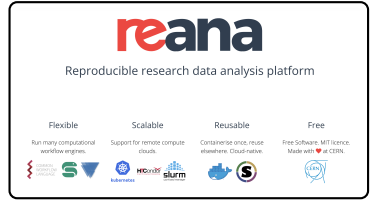
→ **Visit specified URL and approve Mytoken**



REANA overview

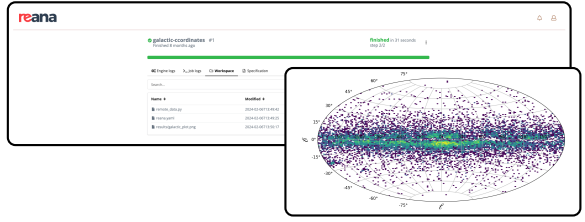
PUNCH4NFDI Data Portal → Single entry point to Research Products

- Data, metadata, software
- Workflow description and history
- Results and publication



Reproducible research data analysis platform REANA → Build Reproducible Research Products

- Supports various computational workflow engines
- Container-based workflow
- User-friendly CLI client and GUI



PUNCH4NFDI REANA server hosted at AIP in Potsdam

Integration in REANA

Scalable compute model and several compute backends known as Job Controllers

- Responsible for job execution and management
- Available for HTCondor, Slurm and Kubernetes
- HTCondor compute backend based on Kerberos tickets

→ **Developed HTCondor Job Controller based on OIDC tokens**

- Produce and upload Mytoken credential to REANA server
- Upload from server to login node for job submission
- Execute and manage HTCondor job on Compute4PUNCH

Job Controller in production and in REANA stable release soon

Summary

- Overview of Compute4PUNCH and Storage4PUNCH latest developments
- Implementation of access token management in HTCondor
- Used by PUNCH4NFDI consortium and DARWIN collaboration
- Integration of Compute4PUNCH into REANA analysis platform

Thanks a lot for your attention!

References

- **FLATICON** - Free Icons and Stickers for your projects.
- Max Fischer, Eileen Kuehn, Manuel Giffels, Matthias Schnepf, Stefan Kroboth, Thorsten M., Oliver Freyermuth. **MatterMiners/cobald (0.14.0) 2023.**
- Manuel Giffels, Max Fischer, Alexander Haas, Stefan Kroboth, Matthias Schnepf, Eileen Kuehn, P. Schuhmacher, Rene Caspart, Florian von Cube, Peter Wienemann. **MatterMiners/tardis (0.8.2) 2024.**

- **CVMFS** - CERN Virtual Machine File System.
- **Mytoken** - Web service to provide OIDC access tokens to long-running compute jobs.
- **C4P-HTCondor** - HTCondor with OIDC access tokens monitoring.
- **REANA** - Reproducible research data analysis platform.
- **REANA Compute4PUNCH Job Controller** - HTCondor Job Controller with OIDC access tokens.

Job submission, production and monitoring

- Next job submission

```
condor_submit analysis.jdl
Submitting job(s)
Hello benoit_roland!  You are going to submit your HTCondor jobs.
A valid credential has been found with a remaining life time of 5 d 4 h 3 m 12 s.
1 job(s) submitted to cluster 1155.
```

- Mytoken standalone production also possible - allow automated job submission afterwards
- Monitoring also handles access token revocation
- User informed at submission time when Mytoken life time below 1 day