PUNCH4NFDI TA2
Compute4PUNCH Storage4PUNCH

Kilian Schwarz
DESY IT/Scientific Computing

TA 2
Work packages

Storage and Compute 4PUNCH

WP1 *Access to Data*

Harmonize access methods

- Standard protocols & interfaces

Advanced storage technology

- Federated & interoperable

- **Employ** recent developments from ESCAPE & WLCG

Community overarching **AAI**
Interfaces to existing meta data

Integration and extension of existing tools

WP2 *Federated computing*

set up common federated compute infrastructure (demonstrator)

integrate heterogeneous resources

realisation of **entry points** (JupyterHub, batch)

Data-locality aware scheduling

integrate opportunistic cache systems

container registry

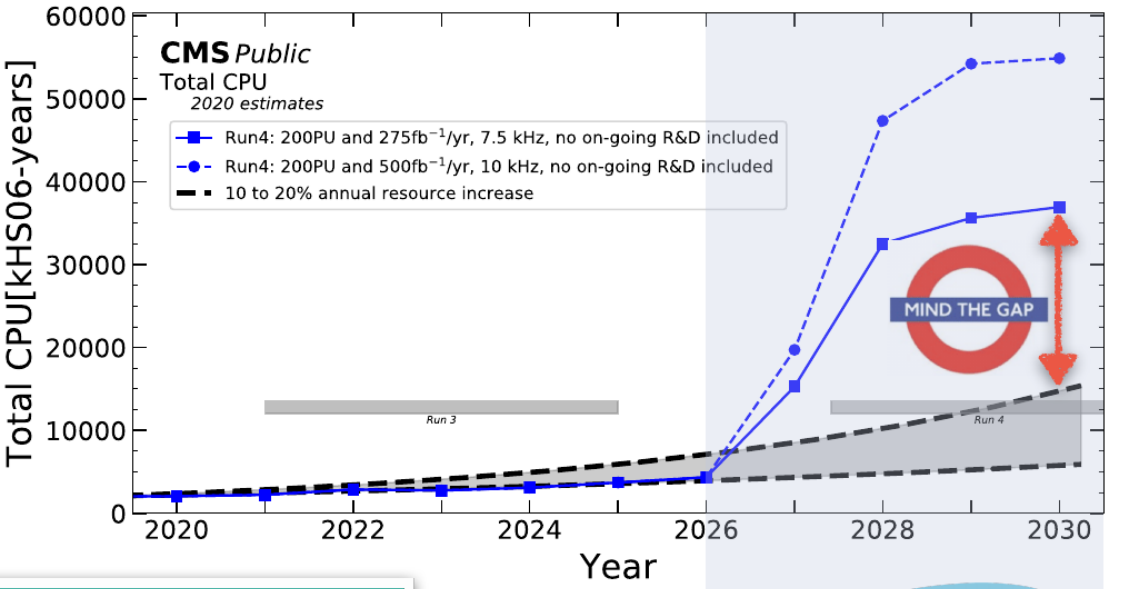
WP3 *Automation*

Storage Workflows
- real-time monitoring
- dynamic data placement and replication

Compute Workflows
- multi Cloud scenarios
- JupyterHub workflow templates

Upcoming Computing Challenges in HEP

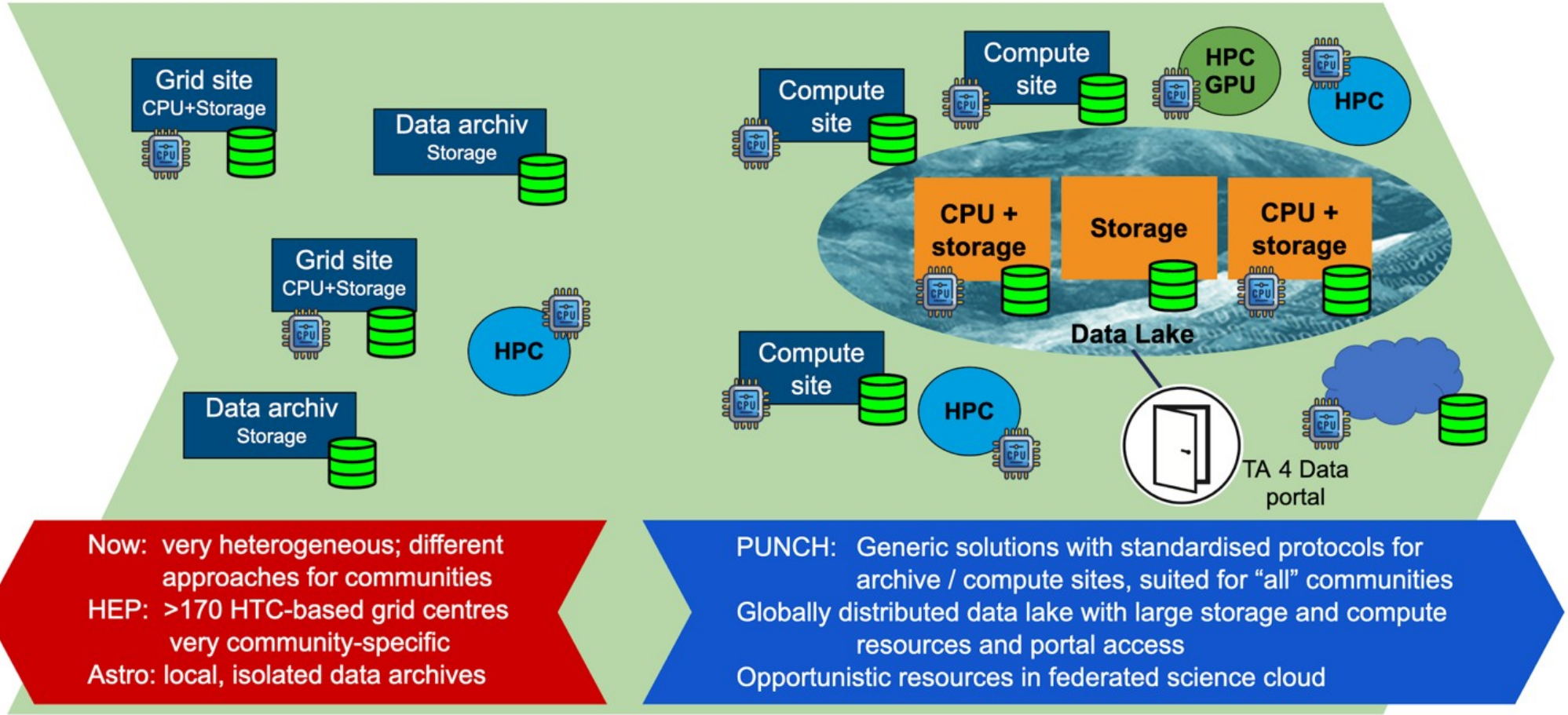
- HL-LHC poses unprecedented challenges to HEP computing
- Assuming flat budget and 10-20% technology advance per year
- CPU shortcoming of factor 2 - 3 estimated in 2027
- ➡ Critical conditions for HL-LHC (Run 4)



Potential Measures:

- Software & Computing Evolution (R&D)
- Add additional resources (opportunistic resources)

TA 2: Data Management



Ongoing Prototyping – Federated Storage & Computing

Common PUNCH-AAI: Based on Helmholtz-AAI

Storage4PUNCH

- Two test systems in place:
- DESY (dCache based)
- Bonn (Xrootd based)
- Access for all PUNCH4NFDI members
- Read and write access though PUNCH-AAI web tokens

Compute4PUNCH

- Implemented via an Overlay Batch System
- First login node at KIT
- Access for all PUNCH4NFDI members
- OIDC token based (mccli)
- Initially included resources at

First (still less demanding) workflows from the community are being ported

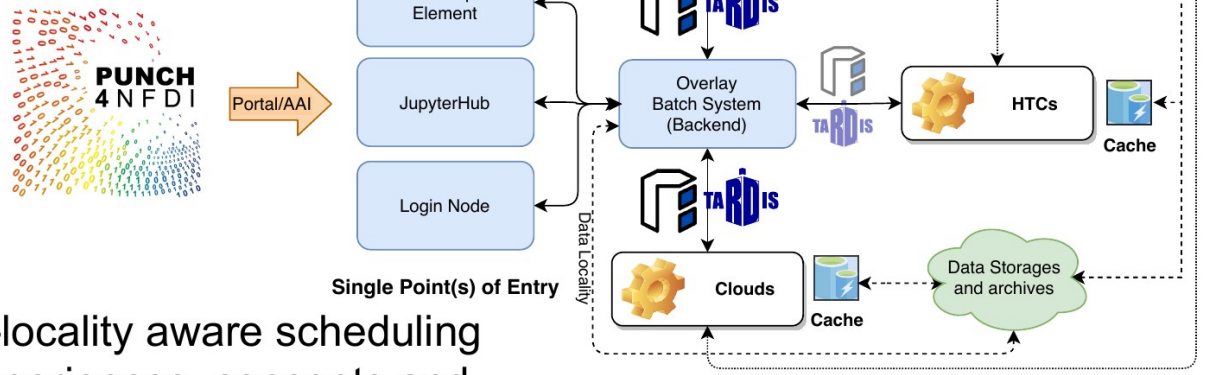
WebDAV, Xrootd

- Container registry

Federated Compute Infrastructure Compute4PUNCH

Outlook: Towards the Compute4PUNCH Infrastructure

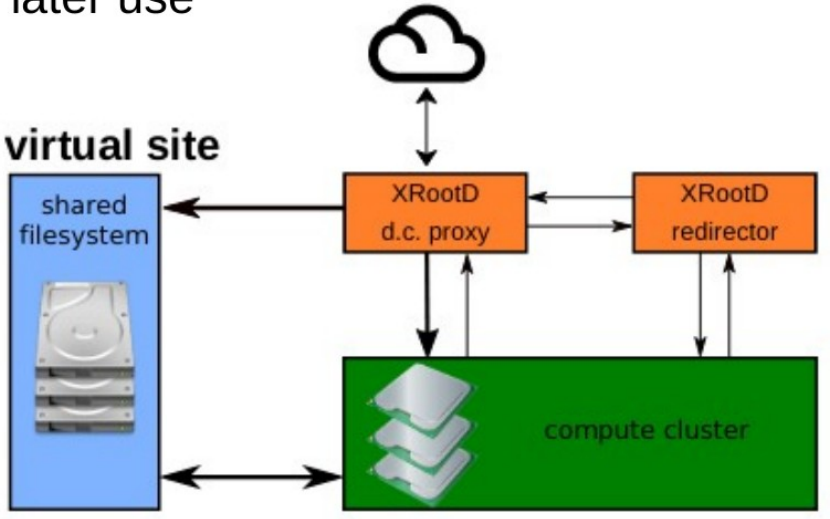
- Establish a federated heterogeneous compute infrastructure for PUNCH
- Integrate data storages, archives and opportunistic caches



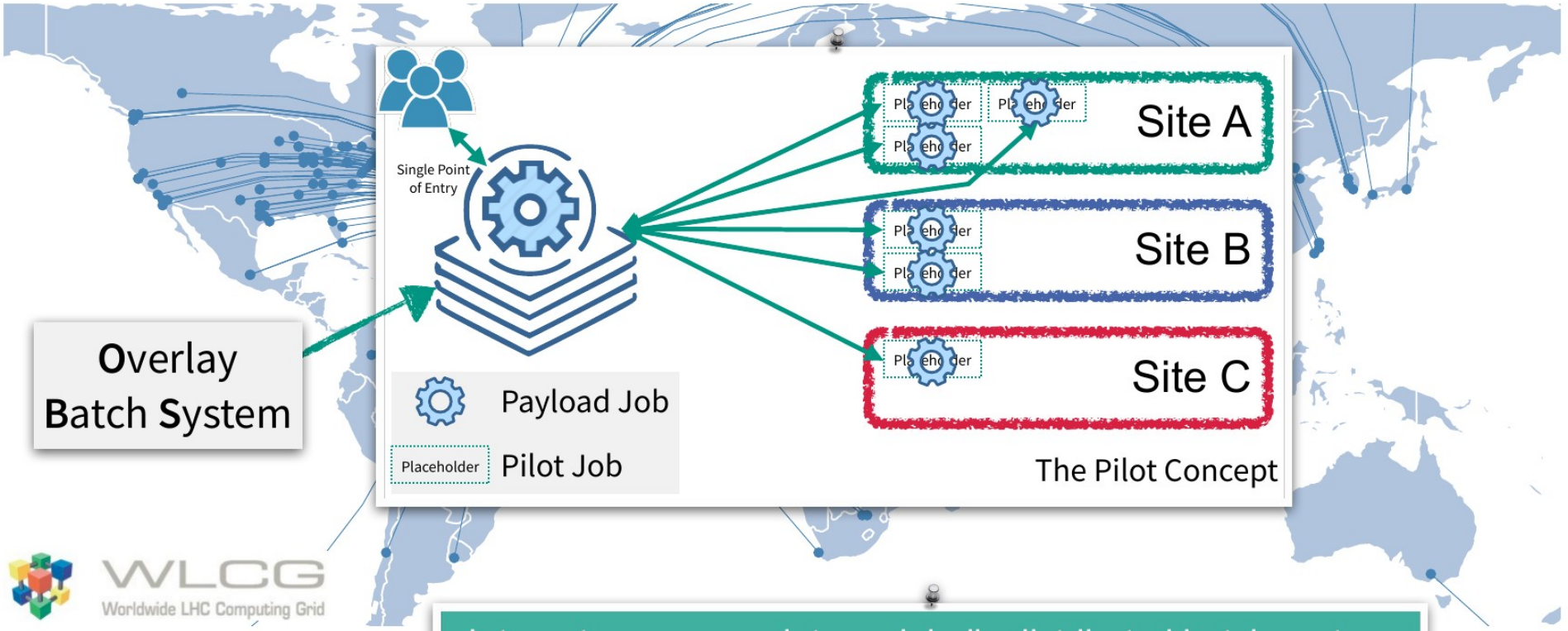
- Introduce data-locality aware scheduling
- Benefit from experiences, concepts and tools available in HEP community

Disk Caching Proxy Setup (Disk Cache on the fly)

- Clients access data through the redirector
- If data exists redirector redirects clients to local file system
- Otherwise redirector redirects to disk caching forward proxy
- Disk caching proxy forwards request to external site and retrieves the data
- Data are being cached on local file system for later use



From Grid towards Global Distributed Computing



The COBalD/TARDIS - Resource Scheduler

[COBalD - the **O**pportunistic **Bal**ancing **D**aemon] [**T**ransparent **A**daptive **R**esource **D**ynamic **I**ntegration **S**ystem]

Karlsruhe Institute of Technology

COBalD

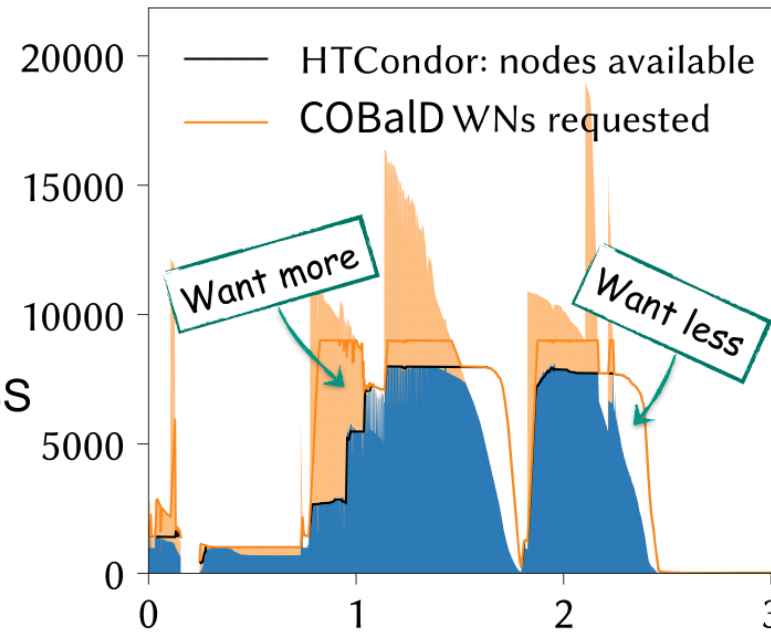
TARDIS

COBalD:

- Look at what is used, not what is requested
  - Simple logic: more used, less unused resources
  - COBalD acquires/releases resources
  - Batch system scheduler handles jobs

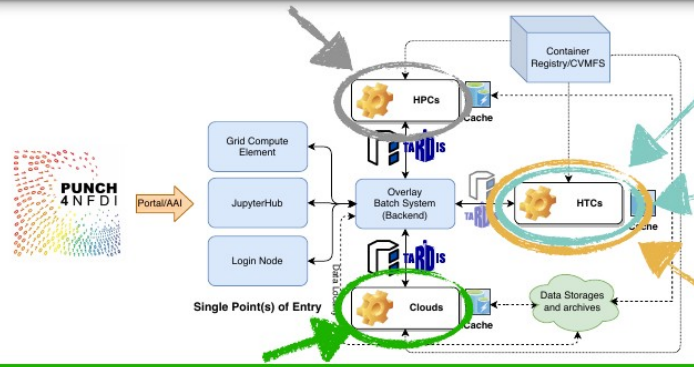
TARDIS (a COBalD plugin):

- Defines VM/Container/Job(Script) as resource
- Provides access to resource provider APIs
  - OpenStack, CloudStack, HTCondor, Slurm, Moab and K8S
- Integrates resources into Overlay Batch System
  - HTCondor and Slurm are supported
- Manages resource life cycle



Status: Compute4PUNCH Infrastructure

MOGON II/Himster II (JGU Mainz)
- Shared access to HPC clusters (~1000 cores for tests)
- Dynamically managed by COBaLD/TARDIS
- To be integrated into Compute4PUNCH
- Status: Work ongoing



OpenStack Cluster (WWU Münster)
- Contingent: 96 cores, 256 GB Memory, 1TB disk space
- Dynamically managed by COBaLD/TARDIS
- Fully integrated into Compute4PUNCH
- Status: Ready to be tested!

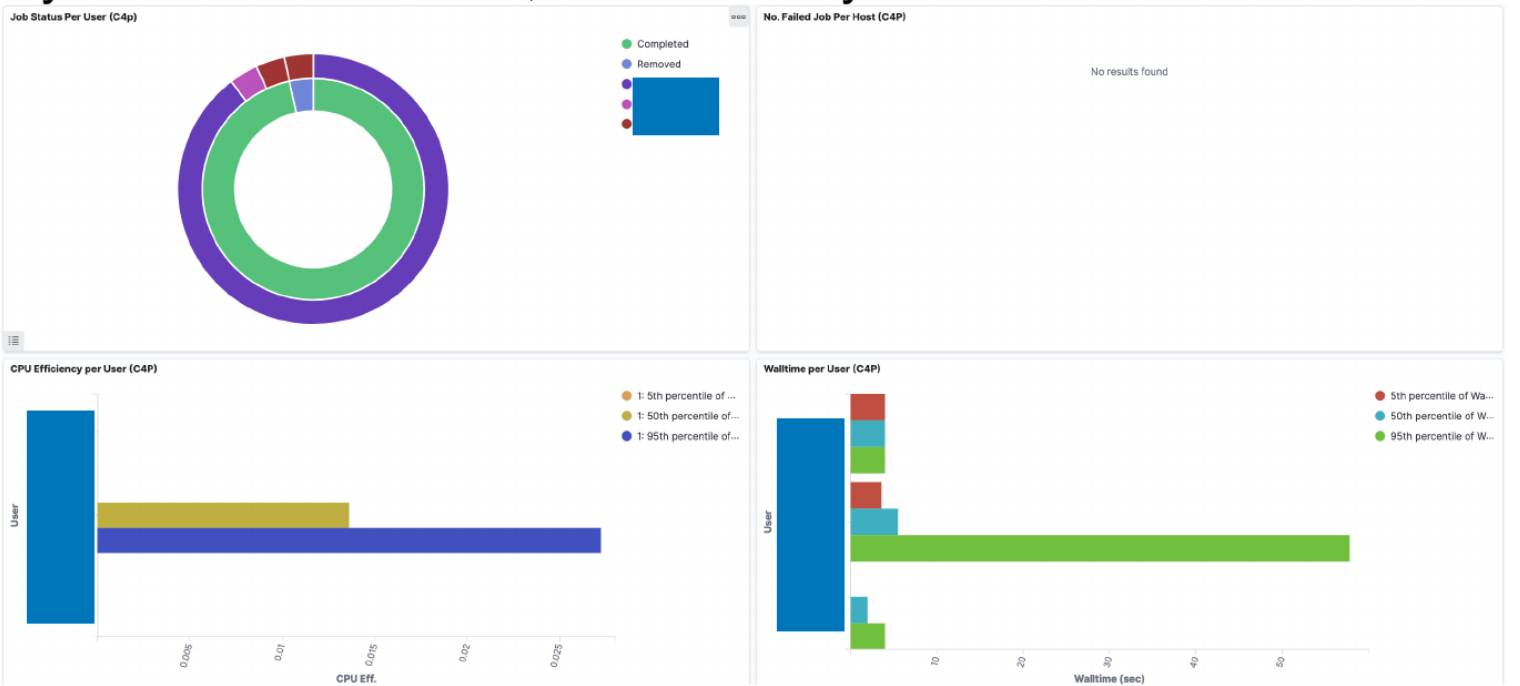
High Throughput Cluster (TOPAS@KIT)
- 8 NVidia V100, 64 cores, 160 GB RAM (shared)
- More cores or RAM available on request
- Dynamically managed by COBaLD/TARDIS
- Fully integrated into Compute4PUNCH
- Status: Ready to be tested!

GridKa Cluster (KIT)
- Up to 2000 cores, 2-3 GB RAM per core (shared)
- Dynamically managed by COBaLD/TARDIS
- To be integrated into Compute4PUNCH
- Status: Work ongoing

Bonn Analysis Facility BAF (Bonn)
- 10 M core-hours on local compute cluster (3.5kCores)
- Dynamically managed by COBaLD/TARDIS
- To be integrated into Compute4PUNCH
- Status: Work ongoing

News: ElasticSearch Compute4PUNCH Monitoring

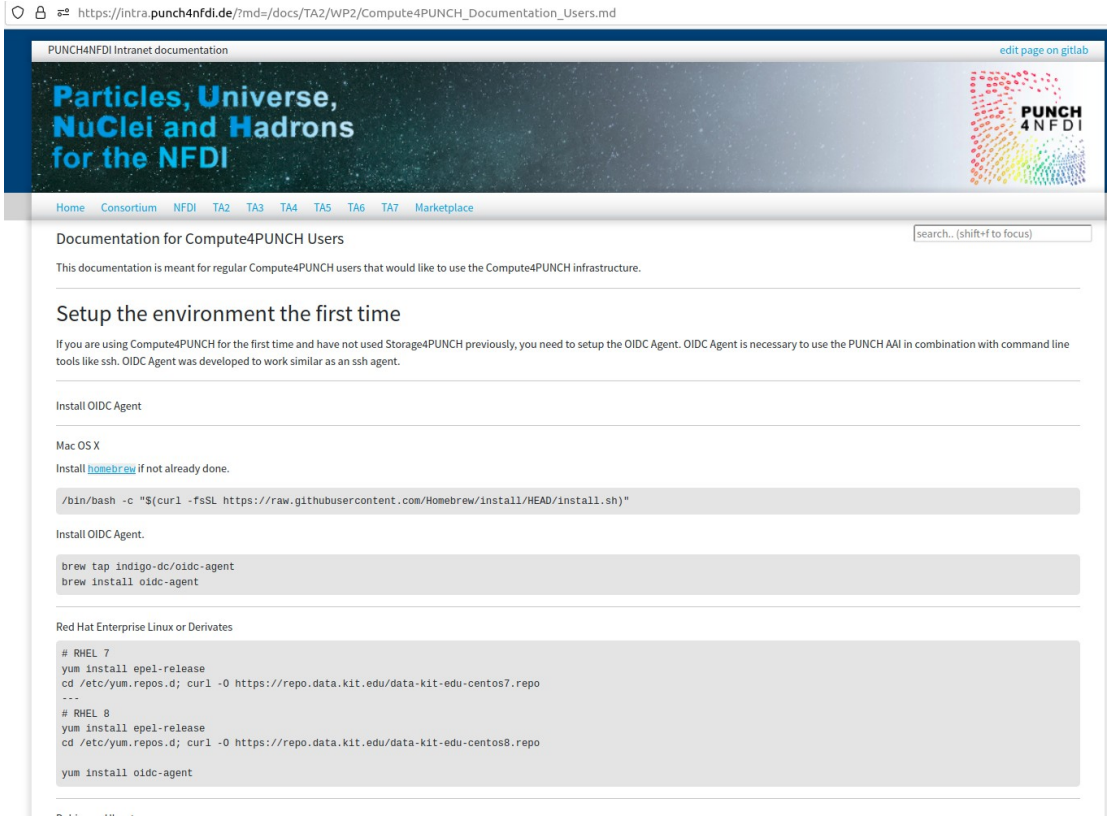
- Uses `condor_adstash` running on login node(s) (HTCondor Sched)
- Currently hosted at KIT-SCC, unfortunately no access via PUNCH AAI



Thanks to M. Schnepf (KIT SCC) for making this possible!

Compute4PUNCH documentation

- Detailed documentation for users and admins available



News: Container Registry / CVMFS

- Utilize the docker registry at AIP (gitlab-p4n.aip.de)
- Dockerfiles managed in directories on GitLab Project (compute4punch/container-stacks)
- Prototype CI to build and upload docker containers is ready and working
- Automated conversion into singularity sandbox format and distribution via CVMFS (unpacked.cern.ch)

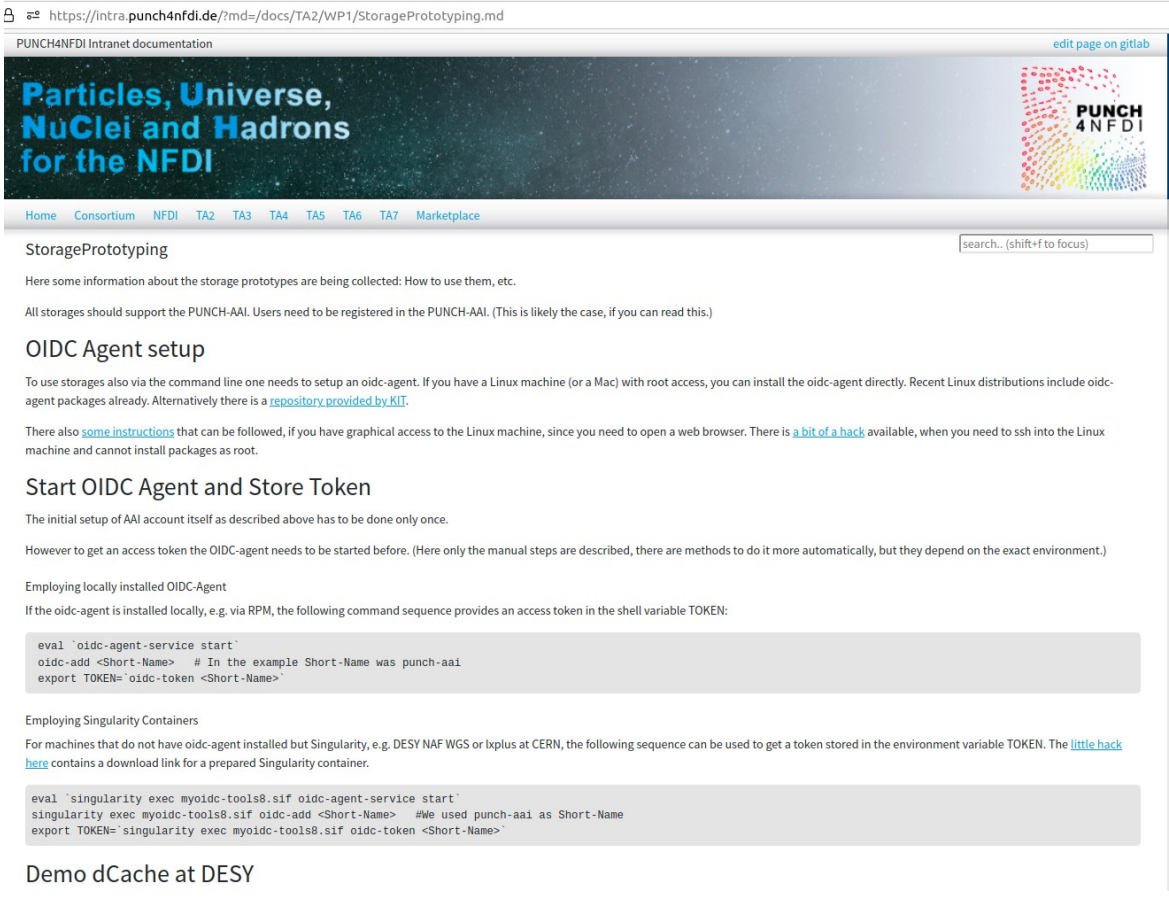
Ongoing/ToDo:

- Need a discussion on container versioning
- Documentation on how to use setup

Ready to give it a try!

Federated Storage Infrastructure Storage4PUNCH

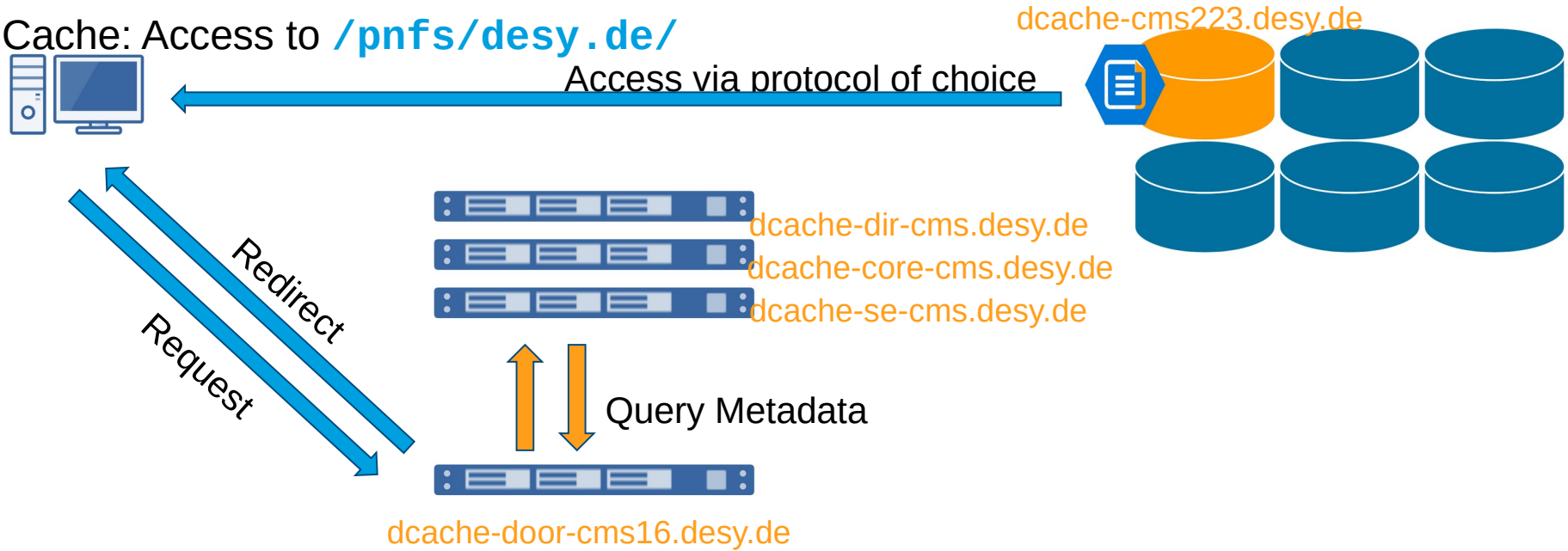
- Storage resources at participating centres running dCache and XrootD Storage Middleware have been interconnected and are accessible via AAI



Basic Setup

Standard Single Site Setup

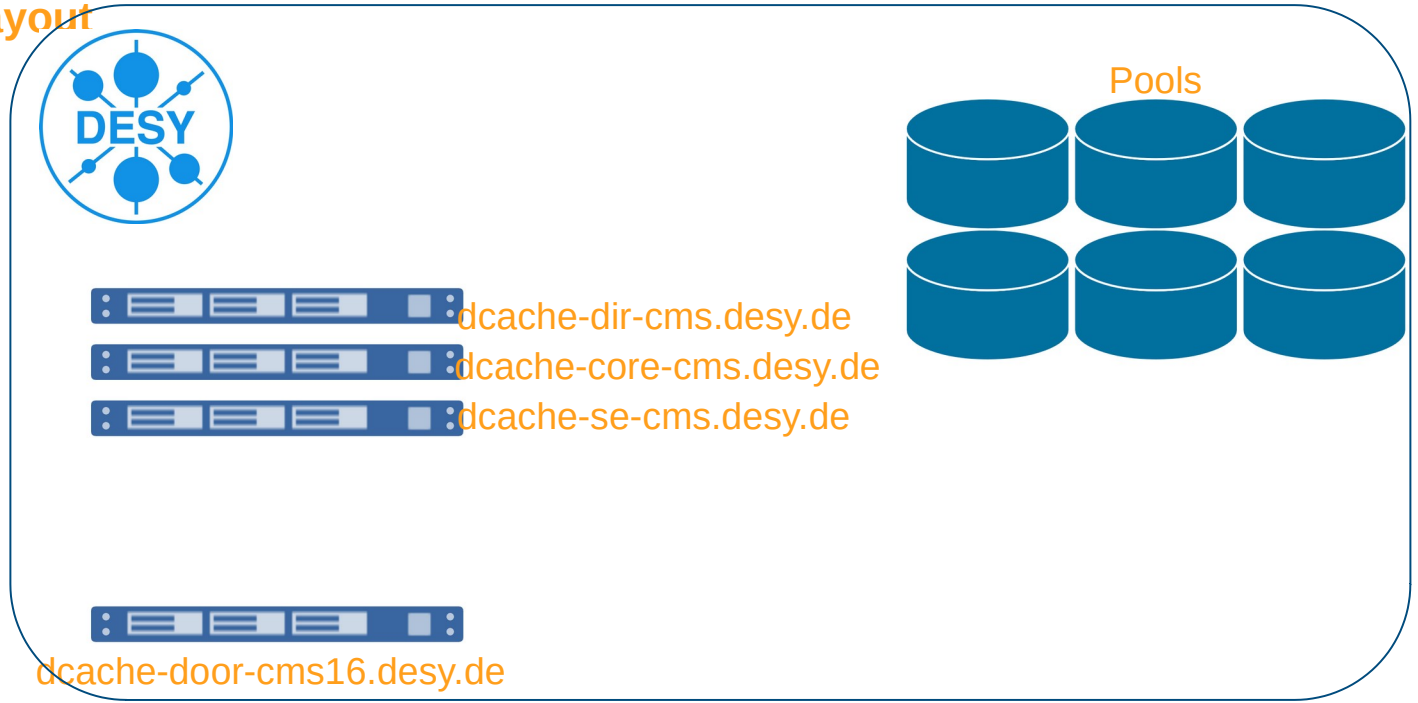
- Use dCache: Access to **`/pnfs/desy.de/`**

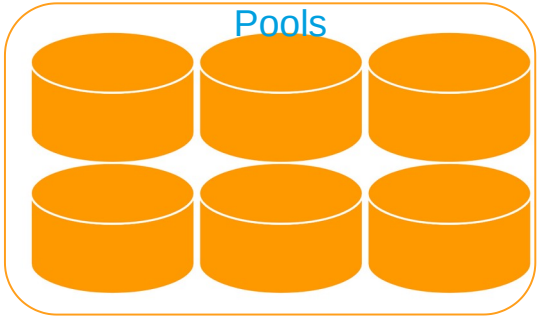


- dCache instances for Photon Science/Machine, European XFEL, ATLAS, CMS, Belle/ILC/DPHEP, Sync&Share
- Simlar layout: three head-nodes, doors for requested protocols and pools nodes
- Scale-out horizontally: 10 pool nodes for Sync&share and 200 for European XFEL with 100 more ordered
- Scale-out horizontally: client always to connect to pools for transfer, no data access through doors

Layout of Federated dCache

Simplest, most Centralised Layout





- At remote site only pools are deployed
- All management services located at central site
- Central accesspoint
- Centralised AAI-interface and namespace

Storage4PUNCH accessible via CLI and Web

Demo dCache at DESY

To support various R&D efforts DESY provides a test system to exercise now functionality. This system can also be used for testing within PUNCH4NFDI. You need to be registered in the PUNCH-AAI.

Note that the test system is configured rather openly. Data can by seen (almost) publicly by anybody. Do not put any non-public data there.

The system is also not meant for production use. There can be not announced downtimes. You might be asked to remove your data (again), because of major reorganisation of the system.

Access via Webbrowser

The system offers a simple file browser:

https://dcache-demo.desy.de/

To log in choose "OpenID-Connect account" and then "Helmholtz" as provider. Perform the login procedure with the credential of your home institution (or social IdP) like you did when registering with the PUNCH-AAI. **Right now every user will be mapped to the user punch**.

Command line access

Prepare your oidc-agent as mentioned above (with the links to documentation).

Get an access token (Note: you need to refresh this regularly, every few hours)

export TOKEN=` oidc-token <short-name>`

You need to have *davix* clients installed. Usually they are provided as packages for all major Linux distributions. Most operations can also be executed using the *curl* tool.

Listing files

```
davix-ls -l -H "Authorization: Bearer ${TOKEN}" https://dcache-demo.desy.de:2443/punch/
```

Downloading files

```
davix-get -H "Authorization: Bearer ${TOKEN}" https://dcache-demo.desy.de:2443/punch/PUNCH_logo.png /tmp/logo.png
```

```
curl -L -X GET -H "Authorization: Bearer ${TOKEN}" --output /tmp/PUNCH.png -H "Authorization: Bearer ${TOKEN}" https://dcache-demo.desy.de:2443/punch/PUNCH_logo.png
```

Uploading files

```
davix-put -H "Authorization: Bearer ${TOKEN}" /bin/bash https://dcache-demo.desy.de:2443/punch/junkdir1/RemoveMe.file
```

```
curl -L -X PUT -H "Authorization: Bearer ${TOKEN}" --upload-file /etc/services https://dcache-demo.desy.de:2443/punch/junkdir1/RemoveMe2.file
```

Make directory

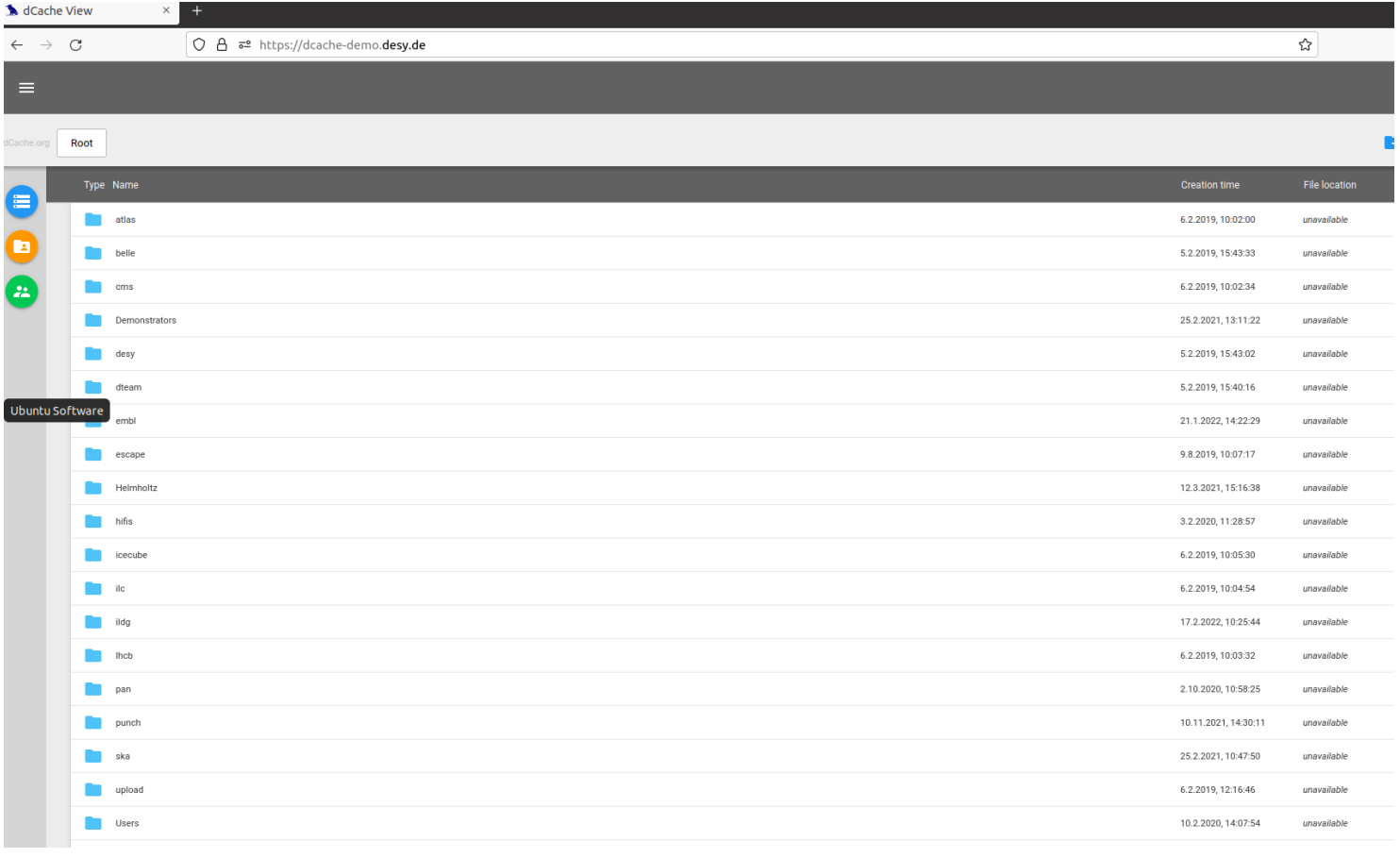
```
davix-mkdir -H "Authorization: Bearer ${TOKEN}" https://dcache-demo.desy.de:2443/punch/junkdir2
```

```
curl -X MKDIR -H "Authorization: Bearer ${TOKEN}" https://dcache-demo.desy.de:2443/punch/junkdir3
```

Delete Files/Directories

```
davix-rm -H "Authorization: Bearer ${TOKEN}" https://dcache-demo.desy.de:2443/punch/junkdir2
```

Storage4PUNCH Web Interface



Identity & Access Management Systems and Storage

PUNCH/Helmholtz AAI together the Storage Prototype at DESY

- Details: see Oliver's talk
- OIDC token bear information about the owner of the token
  - Personal Identifier string: identifies the specific user
  - Eduperson Entitlement: identifies group memberships
- Use the tokens to map to internal UID:GIDs

Storage Perspective

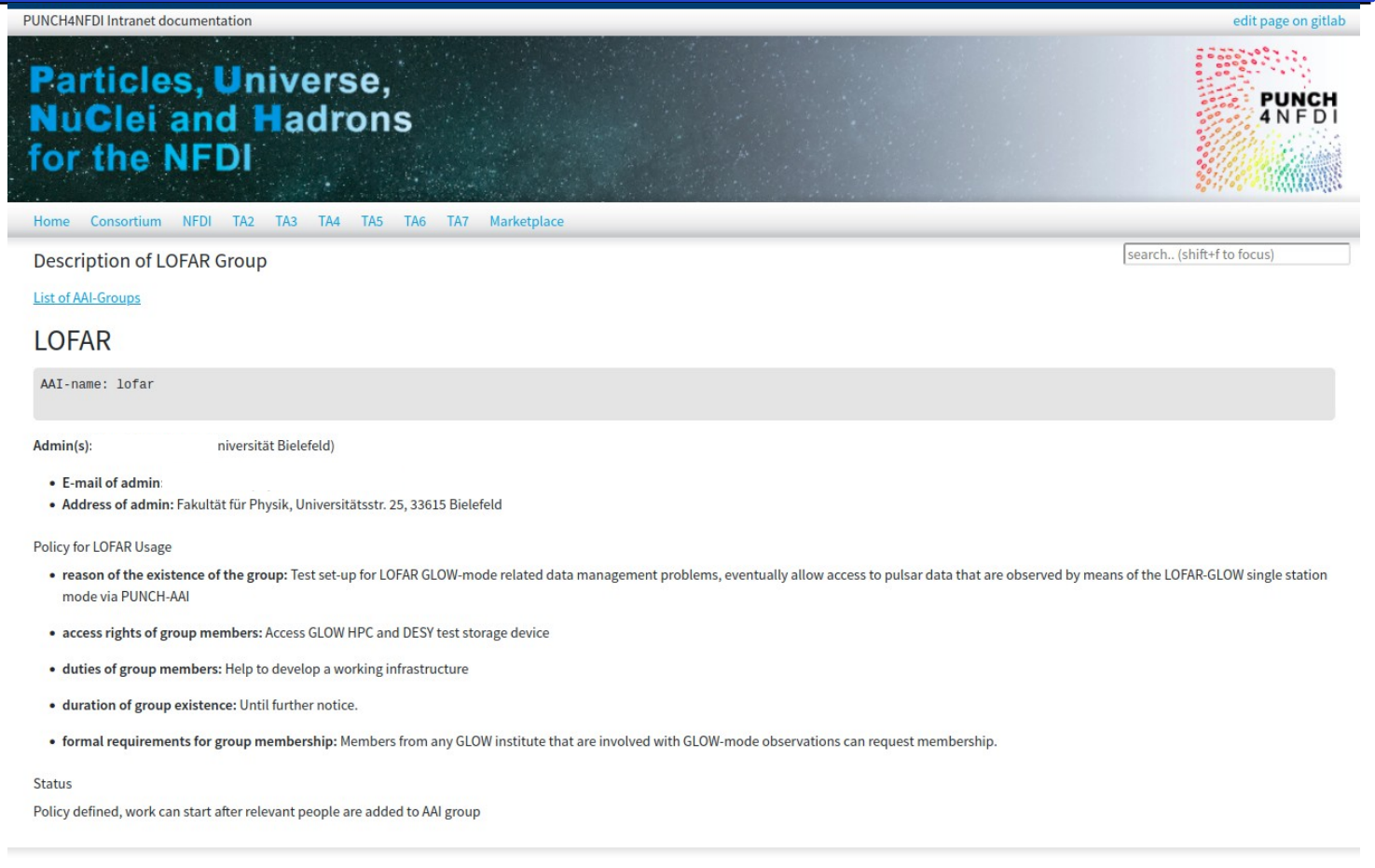
dCache supports

- Classic X.509 certificates and VOMS proxies
- Basic OIDC AAI as done by the PUNCH AAI (based on Helmholtz AAI)
- SciToken close to WLCG Tokens

Client Perspective

- Most commonly used: `gfal` tool kit
  - Support large variety of protocols
  - Current stable release: no token support
  - Authentication via X.509
- Alternatives: native `XrootD` or `curl/davix/rclone` for HTTPS
- `curl/davix` as the most flexible tools w,r.t. authorisation
- `rclone` for using access tokens

Storage4PUNCH dedicated AAI groups for projects



PUNCH4NFDI Intranet documentation

edit page on gitlab

Particles, Universe, NuClei and Hadrons for the NFDI

PUNCH 4NFDI

Home Consortium NFDI TA2 TA3 TA4 TA5 TA6 TA7 Marketplace

search.. (shift+f to focus)

Description of LOFAR Group

List of AAI-Groups

LOFAR

```
AAI-name: lofar
```

Admin(s): niversität Bielefeld)

- **E-mail of admin**:
- **Address of admin:** Fakultät für Physik, Universitätsstr. 25, 33615 Bielefeld

Policy for LOFAR Usage

- **reason of the existence of the group:** Test set-up for LOFAR GLOW-mode related data management problems, eventually allow access to pulsar data that are observed by means of the LOFAR-GLOW single station mode via PUNCH-AAI
- **access rights of group members:** Access GLOW HPC and DESY test storage device
- **duties of group members:** Help to develop a working infrastructure
- **duration of group existence:** Until further notice.
- **formal requirements for group membership:** Members from any GLOW institute that are involved with GLOW-mode observations can request membership.

Status

Policy defined, work can start after relevant people are added to AAI group