

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**

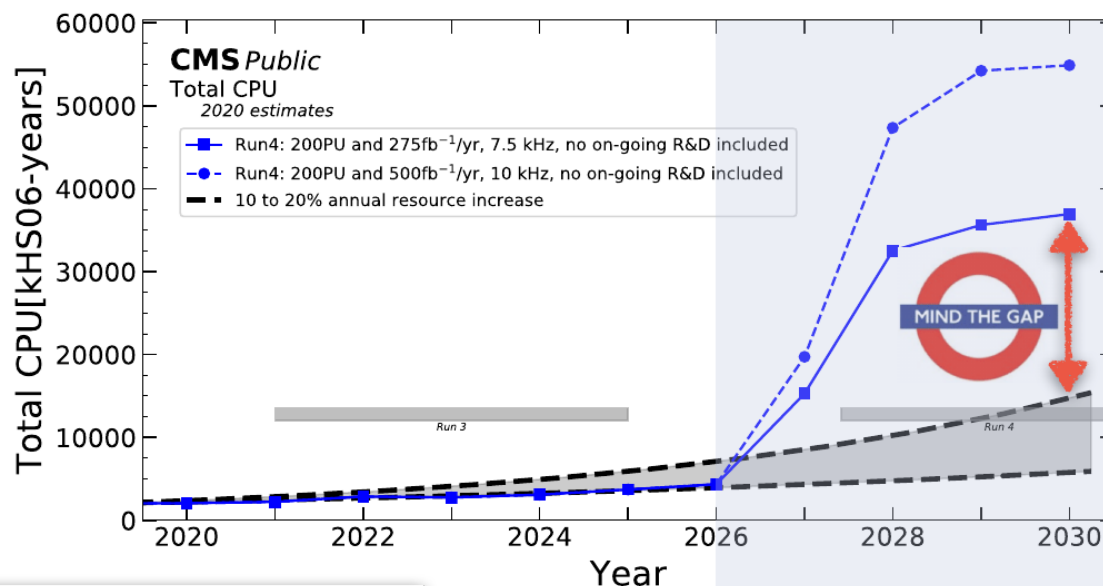
- real-time monitoring
- dynamic data placement and replication

**Compute Workflows**

- multi Cloud scenarios
- JupyterHub workflow templates

# Upcoming Computing Challenges in HEP

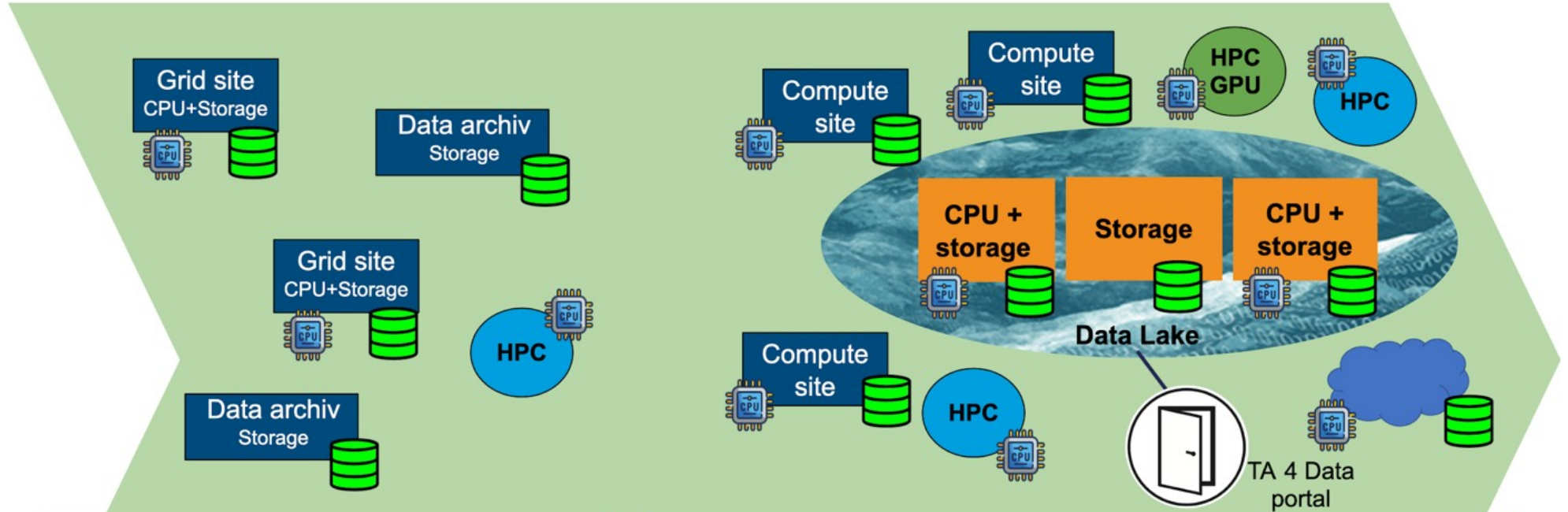
- HL-LHC poses unprecedented challenges to HEP computing
- Assuming flat budget and 10-20% technology advance per year
- CPU shortcoming of factor 2 - 3 estimated in 2027
- ➔ Critical conditions for HL-LHC (Run 4)



## Potential Measures:

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

# TA 2: Data Management



Now: very heterogeneous; different approaches for communities  
HEP: >170 HTC-based grid centres  
very community-specific  
Astro: local, isolated data archives

PUNCH: Generic solutions with standardised protocols for archive / compute sites, suited for "all" communities  
Globally distributed data lake with large storage and compute resources and portal access  
Opportunistic resources in federated science cloud

# 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 through 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

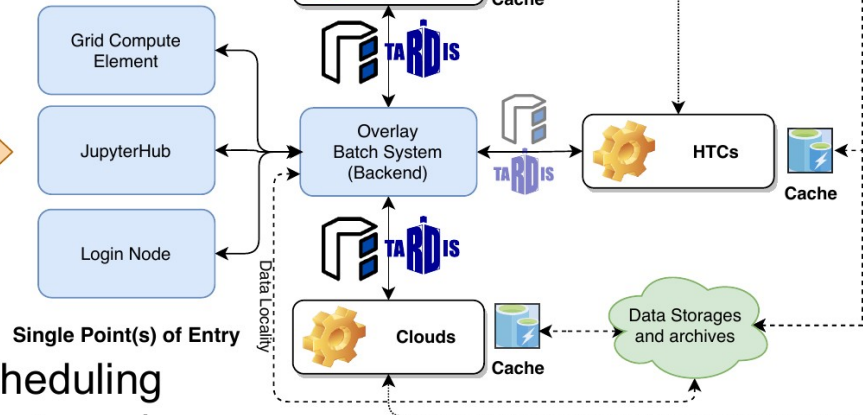
• Container registry

# Federated Compute Infrastructure Compute4PUNCH



## Outlook: Towards the Compute4PUNCH Infrastructure

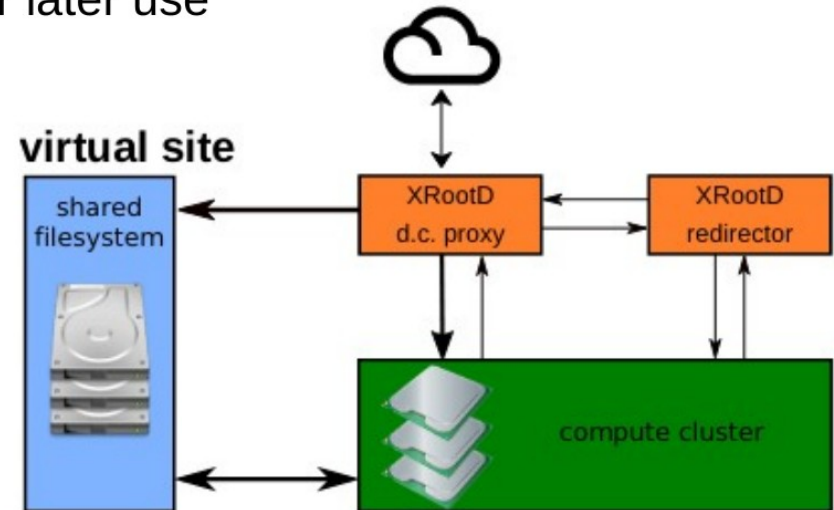
- Establish a federated heterogeneous compute infrastructure for PUNCH
- Integrate data storages, archives and opportunistic caches



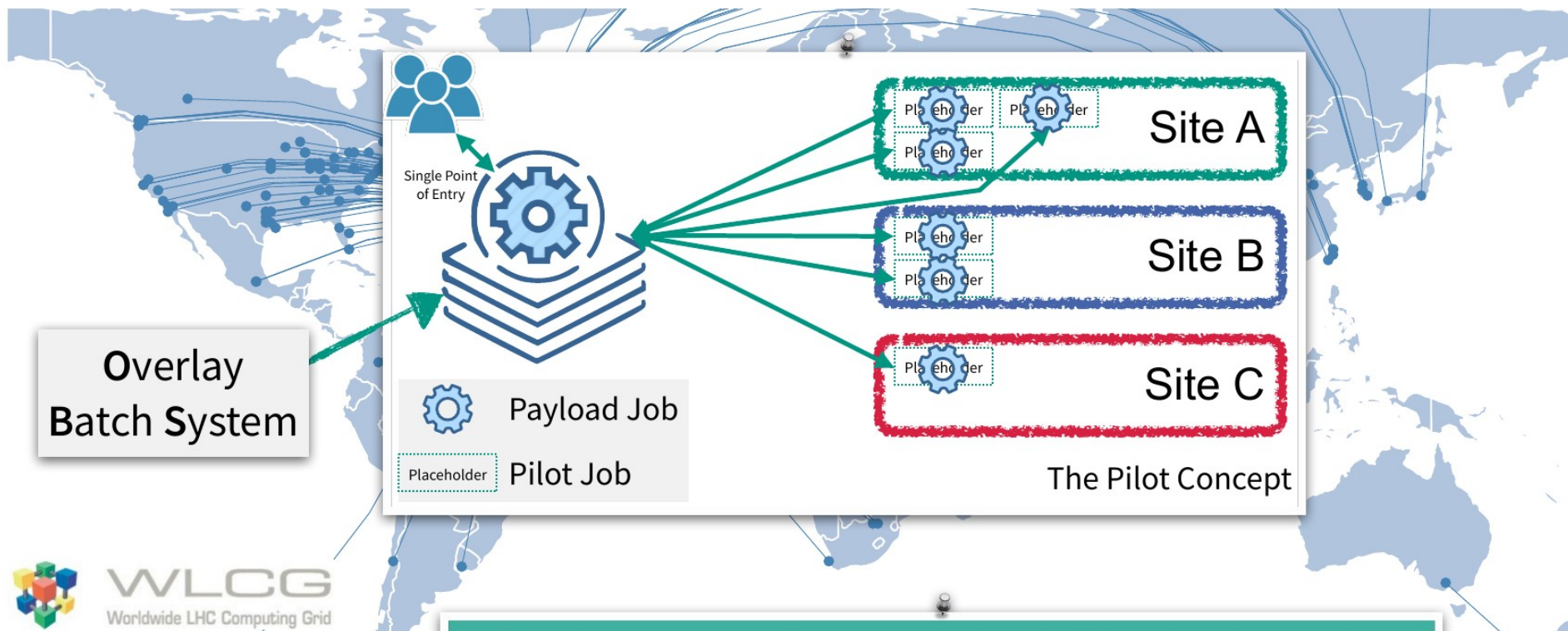
- Introduce data-locality aware scheduling
- Benefit from experiences, concepts and tools available in HEP community

# 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



Integrate resources into a globally distributed batch system and remove a significant fraction of the initial Grid middleware



# The COBaID/TARDIS - Resource Scheduler

[COBaID - the Opportunistic **B**alancing **D**aemon] [Transparent **A**daptive **R**esource **D**ynamic Integration **S**ystem]

## COBaID:

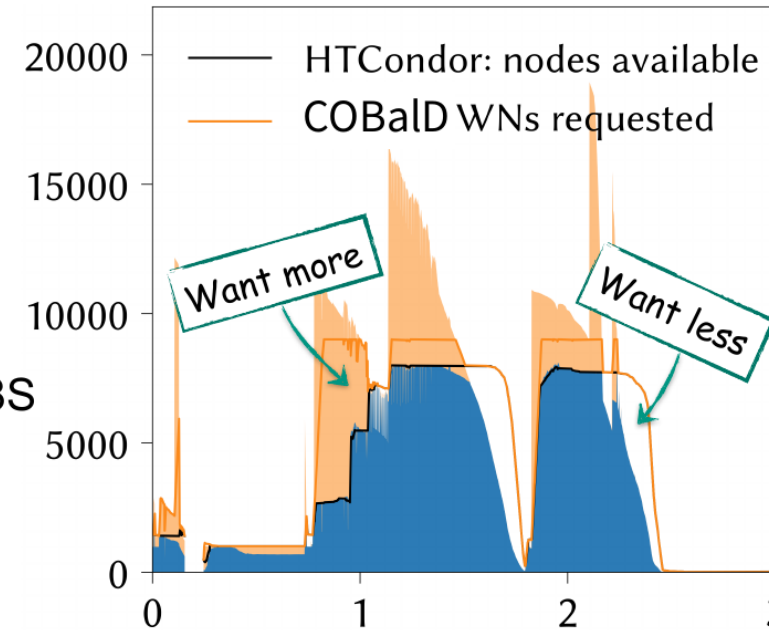
- Look at what is used, not what is requested
  - Simple logic: **more used, less unused** resources
  - COBaID acquires/releases resources
  - Batch system scheduler handles jobs

## TARDIS (a COBaID 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



COBaID



# 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**

## 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)

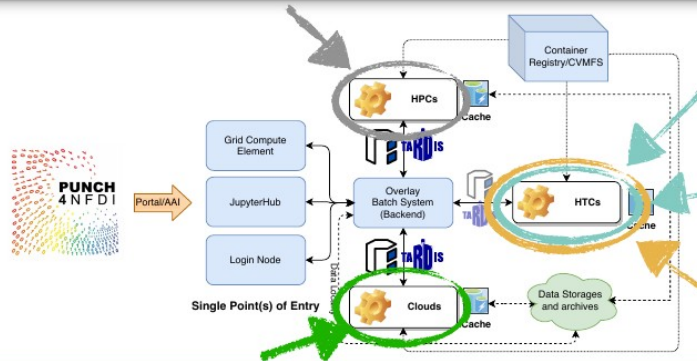
- Up to 2000 cores, 2-3 GB RAM per core (shared)
- Dynamically managed by COBaLD/TARDIS
- To be integrated into Compute4PUNCH
- Status: **Work ongoing**

## 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**

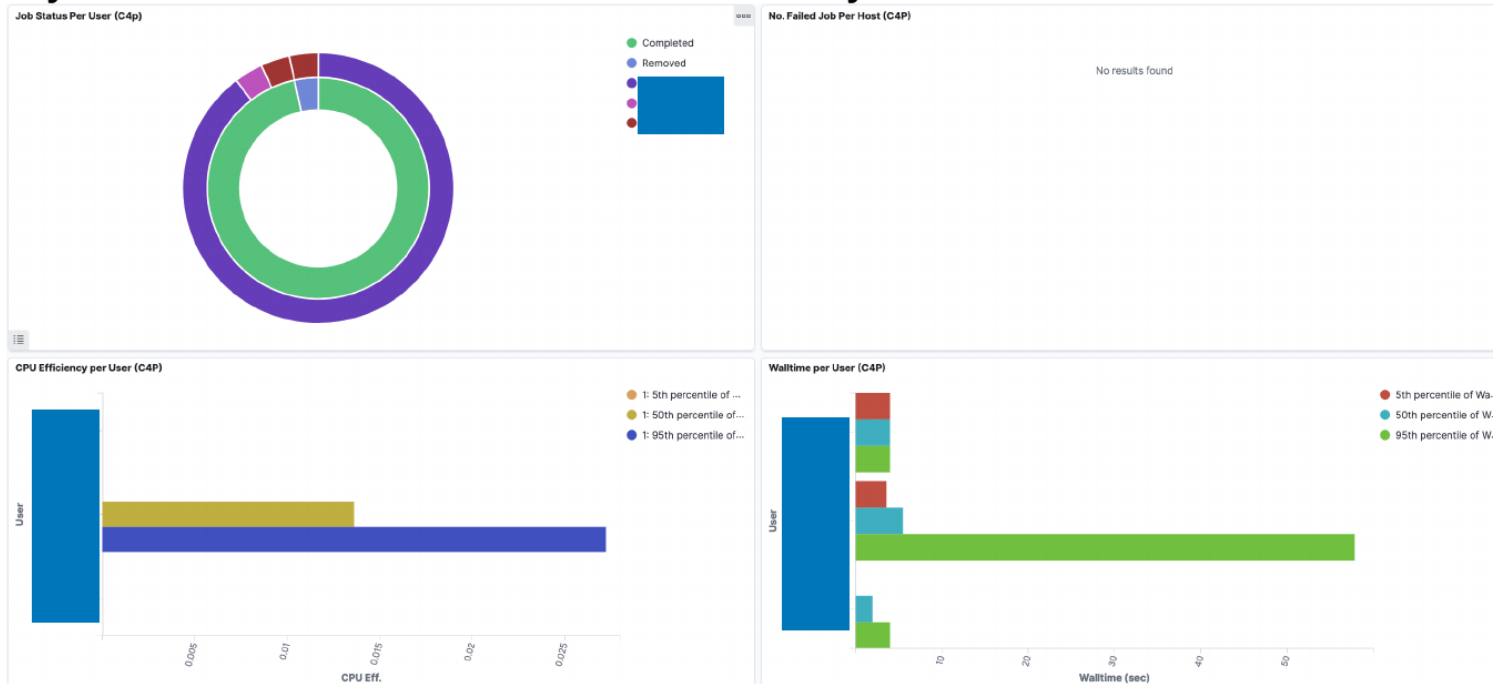
## OpenStack Cluster (WWU Münster)

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



# 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

🔍 🌐 📄 https://intra.punch4nfdi.de/?md=/docs/TA2/WP2/Compute4PUNCH\_Documentation\_Users.md

PUNCH4NFDI Intranet documentation [edit page on gitlab](#)

**Particles, Universe,  
NuCleI and Hadrons  
for the NFDI**

**PUNCH  
4 N F D I**

[Home](#) [Consortium](#) [NFDI](#) [TA2](#) [TA3](#) [TA4](#) [TA5](#) [TA6](#) [TA7](#) [Marketplace](#)

Documentation for Compute4PUNCH Users

This documentation is meant for regular Compute4PUNCH users that would like to use the Compute4PUNCH infrastructure.

## Setup the environment the first time

If you are using Compute4PUNCH for the first time and have not used Storage4PUNCH previously, you need to setup the OIDC Agent. OIDC Agent is necessary to use the PUNCH AAI in combination with command line tools like ssh. OIDC Agent was developed to work similar as an ssh agent.

### Install OIDC Agent

Mac OS X

Install [homebrew](#) if not already done.

```
/bin/bash -c "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)"
```

Install OIDC Agent.

```
brew tap indigo-dc/oidc-agent
brew install oidc-agent
```

Red Hat Enterprise Linux or Derivates

```
# RHEL 7
yum install epel-release
cd /etc/yum.repos.d; curl -O https://repo.data.kit.edu/data-kit-edu-centos7.repo
...

# RHEL 8
yum install epel-release
cd /etc/yum.repos.d; curl -O https://repo.data.kit.edu/data-kit-edu-centos8.repo

yum install oidc-agent
```

# News: Container Registry / CVMFS

- Utilize the docker registry at AIP ([gitlab-p4n.aip.de](https://gitlab-p4n.aip.de))
- Dockerfiles managed in directories on GitLab Project ([compute4punch/container-stacks](https://gitlab.com/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](https://unpacked.cern.ch))

## Ongoing/ToDo:

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

Ready to give it a try!

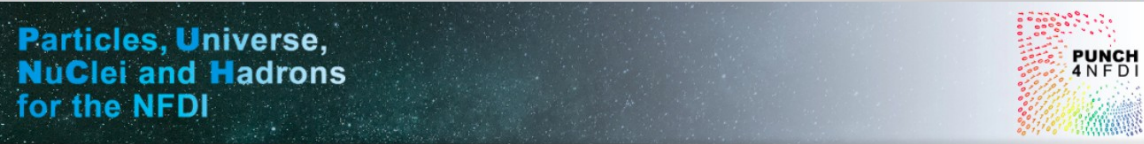
# Federated Storage Infrastructure

## Storage4PUNCH

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

https://intra.punch4nfdi.de/?md=/docs/TA2/WP1/StoragePrototyping.md

PUNCH4NFDI Intranet documentation [edit page on gitlab](#)



Home Consortium NFDI TA2 TA3 TA4 TA5 TA6 TA7 Marketplace

### StoragePrototyping

Here some information about the storage prototypes are being collected: How to use them, etc.

All storages should support the PUNCH-AAI. Users need to be registered in the PUNCH-AAI. (This is likely the case, if you can read this.)

### OIDC Agent setup

To use storages also via the command line one needs to setup an oidc-agent. If you have a Linux machine (or a Mac) with root access, you can install the oidc-agent directly. Recent Linux distributions include oidc-agent packages already. Alternatively there is a [repository provided by KIT](#).

There also [some instructions](#) that can be followed, if you have graphical access to the Linux machine, since you need to open a web browser. There is [a bit of a hack](#) available, when you need to ssh into the Linux machine and cannot install packages as root.

### Start OIDC Agent and Store Token

The initial setup of AAI account itself as described above has to be done only once.

However to get an access token the OIDC-agent needs to be started before. (Here only the manual steps are described, there are methods to do it more automatically, but they depend on the exact environment.)

Employing locally installed OIDC-Agent

If the oidc-agent is installed locally, e.g. via RPM, the following command sequence provides an access token in the shell variable TOKEN:

```
eval `oidc-agent-service start`
oidc-add <Short-Name> # In the example Short-Name was punch-aa1
export TOKEN=`oidc-token <Short-Name>`
```

Employing Singularity Containers

For machines that do not have oidc-agent installed but Singularity, e.g. DESY NAF WGS or lxplus at CERN, the following sequence can be used to get a token stored in the environment variable TOKEN. The [little hack here](#) contains a download link for a prepared Singularity container.

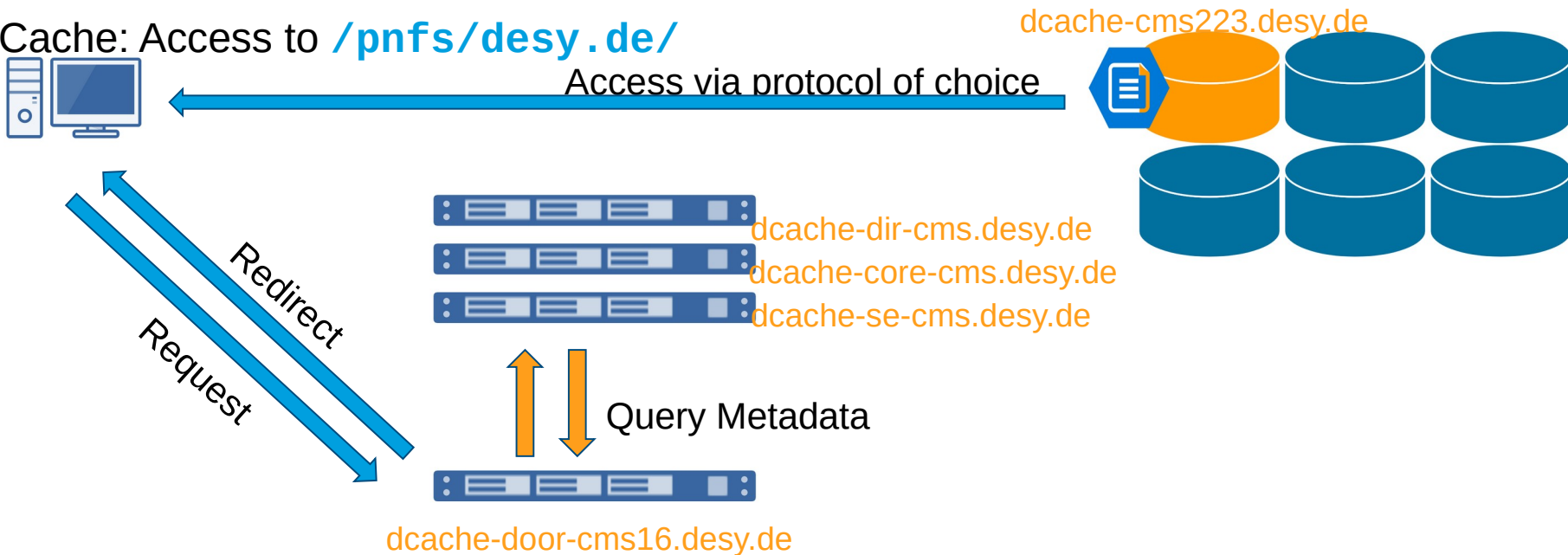
```
eval `singularity exec myoidc-tools8.sif oidc-agent-service start`
singularity exec myoidc-tools8.sif oidc-add <Short-Name> #We used punch-aa1 as Short-Name
export TOKEN=`singularity exec myoidc-tools8.sif oidc-token <Short-Name>`
```

### Demo dCache at DESY

# Basic Setup

## Standard Single Site Setup

- Use dCache: Access to [pnfs/desy.de/](https://pnfs.desy.de/)



- dCache instances for Photon Science/Machine, European XFEL, ATLAS, CMS, Belle/ILC/DPHEP, Sync&Share
- Similar 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



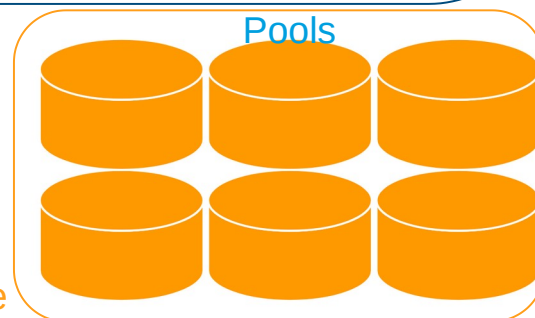
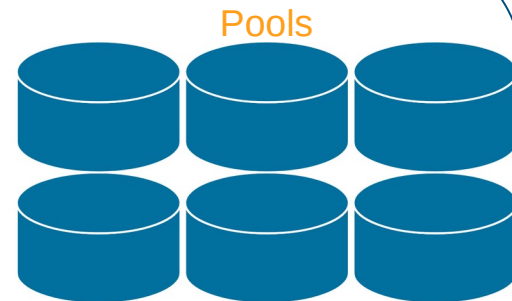
dcache-dir-cms.desy.de

dcache-core-cms.desy.de

dcache-se-cms.desy.de



dcache-door-cms16.desy.de



- 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 be 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



The screenshot shows the dCache View web interface. The browser address bar displays <https://dcache-demo.desy.de>. The interface shows a directory listing for the 'Root' path. The table below lists the contents of the directory.

Type	Name	Creation time	File location
Folder	atlas	6.2.2019, 10:02:00	unavailable
Folder	belle	5.2.2019, 15:43:33	unavailable
Folder	cms	6.2.2019, 10:02:34	unavailable
Folder	Demonstrators	25.2.2021, 13:11:22	unavailable
Folder	desy	5.2.2019, 15:43:02	unavailable
Folder	dteam	5.2.2019, 15:40:16	unavailable
Folder	embi	21.1.2022, 14:22:29	unavailable
Folder	escape	9.8.2019, 10:07:17	unavailable
Folder	Helmholtz	12.3.2021, 15:16:38	unavailable
Folder	hifs	3.2.2020, 11:28:57	unavailable
Folder	icecube	6.2.2019, 10:05:30	unavailable
Folder	ilc	6.2.2019, 10:04:54	unavailable
Folder	ildg	17.2.2022, 10:25:44	unavailable
Folder	lhcb	6.2.2019, 10:03:32	unavailable
Folder	pan	2.10.2020, 10:58:25	unavailable
Folder	punch	10.11.2021, 14:30:11	unavailable
Folder	ska	25.2.2021, 10:47:50	unavailable
Folder	upload	6.2.2019, 12:16:46	unavailable
Folder	Users	10.2.2020, 14:07:54	unavailable

# 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



[Home](#) [Consortium](#) [NFDI](#) [TA2](#) [TA3](#) [TA4](#) [TA5](#) [TA6](#) [TA7](#) [Marketplace](#)

## 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