Using the dCache-Storage Infrastructure at DESY
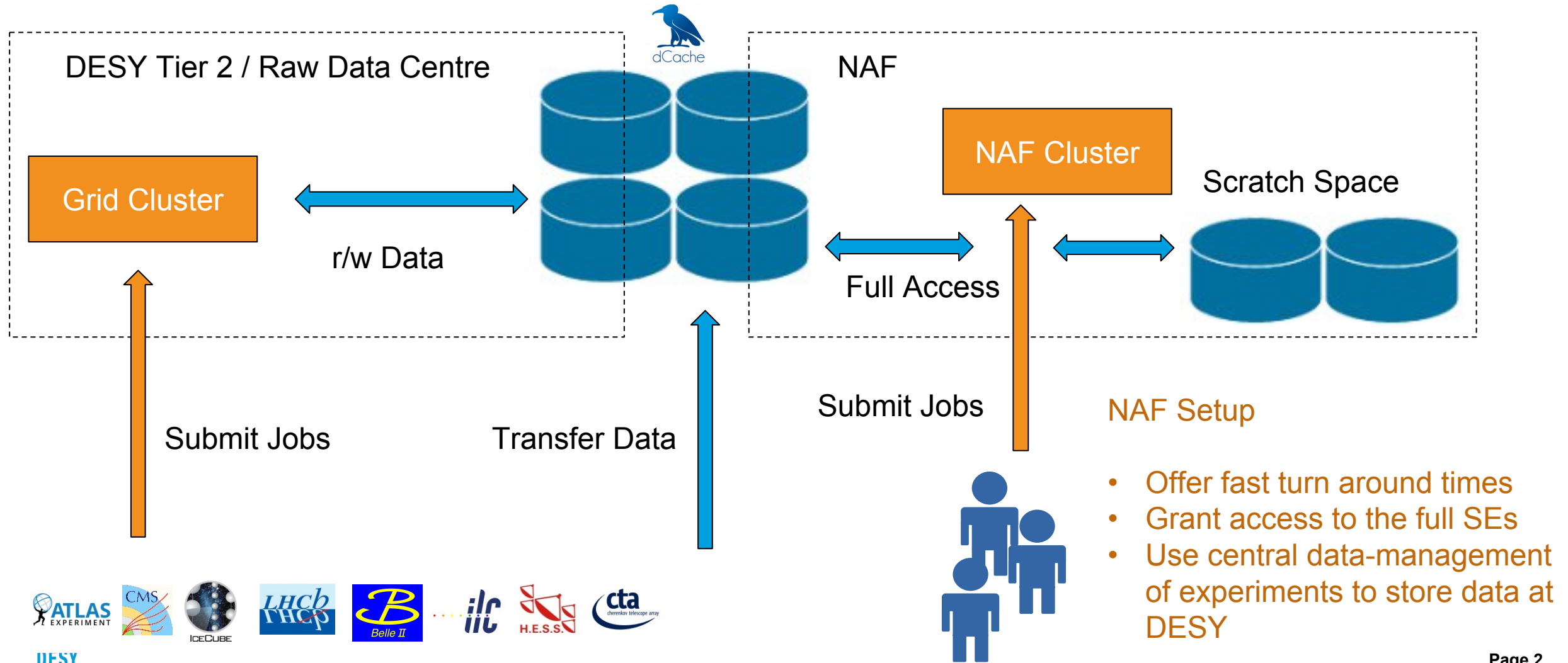
Meeting with Colleagues from the Machine Division

Christian Voß
DESY, 26th May 2023

Paradigm: HEP Analyses are Data Driven

As Underlying Principle for dCache Storage Architecture at DESY

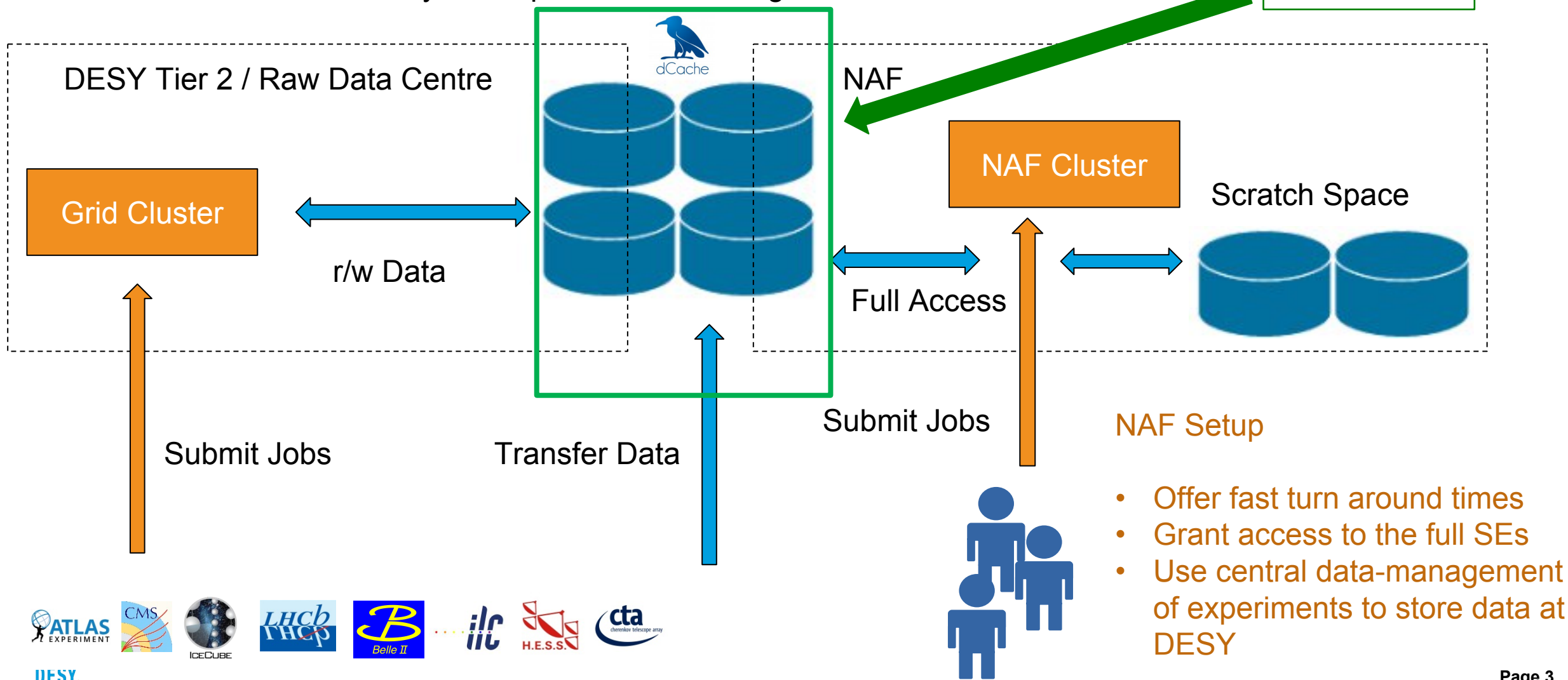- Almost all HEP data analyses require access to large amounts of data

DESY Tier 2 / Raw Data Centre

NAF

Grid Cluster

r/w Data

NAF Cluster

Scratch Space

Full Access

Submit Jobs

Transfer Data

Submit Jobs

NAF Setup

- Offer fast turn around times
- Grant access to the full SEs
- Use central data-management of experiments to store data at DESY

Paradigm: HEP Analyses are Data Driven

As Underlying Principle of the NAF

- Almost all HEP data analyses require access to large amounts of data

Put focus on

DESY Tier 2 / Raw Data Centre

dCache

NAF

Grid Cluster

NAF Cluster

Scratch Space

r/w Data

Full Access

Submit Jobs

Transfer Data

Submit Jobs

NAF Setup

- Offer fast turn around times
- Grant access to the full SEs
- Use central data-management of experiments to store data at DESY

Mass-Storage for Different Communities

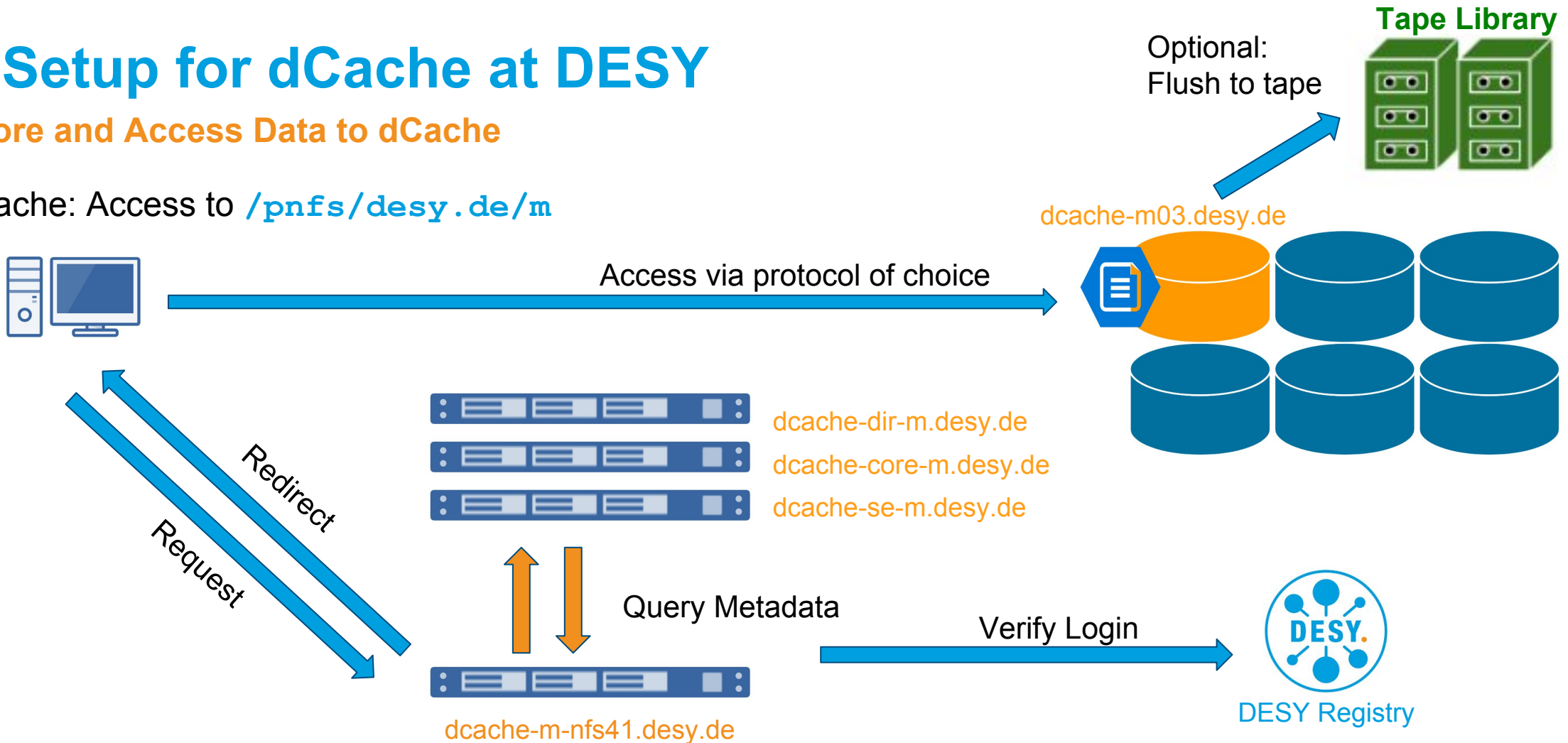## dCache as Central Mass Storage

- Central element in overall storage strategy
- Collaborative development under open source licence by
  - DESY
  - Fermilab
  - Nordic E-Infrastructure Collaboration (inoffically NDGF)
- Particle Physics in general
  - In production at 9 of 13 WLCG Tier-1 centres
  - In use at over 60 Tier-2 sites world wide
  - 75% of all remote LHC data stored on dCache
  - In addition: Tevatron and HERA data
  - All smaller DESY experiments store data in dCache
- Photon Science
  - Raw-Data for all DESY light sources
  - Long-term archival
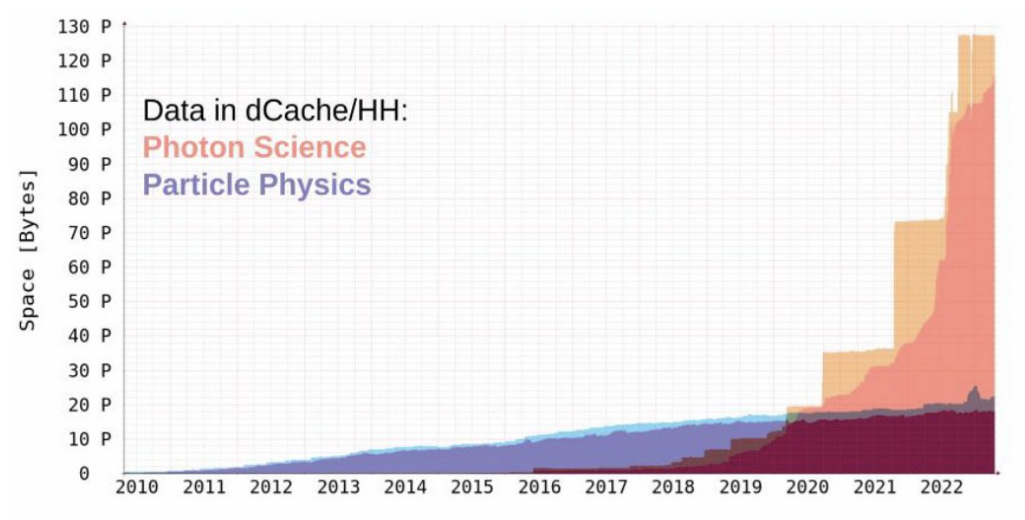
Basic Setup for dCache at DESY

How to Store and Access Data to dCache

Tape Library

Optional:
Flush to tape

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

dcache-m03.desy.de

Access via protocol of choice

Redirect

Request

dcache-dir-m.desy.de
dcache-core-m.desy.de
dcache-se-m.desy.de

Query Metadata

Verify Login

DESY

DESY Registry

dcache-m-nfs41.desy.de

- Access done through doors: several load balanced door for each protocol to ensure availability (M uses NFS)
- Access controlled via Grid-certificates or by tokens and POSIX (NFS@NAF and Maxwell)
- Data streamed to/from pool, never through doors: allows horizontal scaling
- Namespace is uniform and independent of protocol

Access to dCache

Via NAF or Maxwell or DAQ Nodes Using NFS

- Straight forward ➜ login to a NAF WGS and treat it like a local file

- Works on work-group servers and worker nodes

```
[vossc@naf-m01] ~ $ ls /pnfs/desy.de/m
group  mls  mpa  msk  plasma  project  projects  user
[vossc@naf-m01] ~ $ md5sum /pnfs/desy.de/m/project/profile
253089b60a2bd52300223f8dd4680363  /pnfs/desy.de/m/project/profile
```

- Permissions are handled through POSIX definitions

  - You write files with your UID and primary GID (e.g. `16490:1000` for me)

  - POSIX permissions define access: user, group, others; if you're a member of the group, even if not as primary, the group permissions will take effect

```
[vossc@naf-m01] ~ $ ls -l /pnfs/desy.de/m/group/msk/ipc/intern/
total 1
drwxr-xr-x 4 ipcds msk_ipc_ds 512 Apr 19 12:39 DAQ_Data
```

  - Not a member of the group: Permission denied

```
[vossc@naf-m01] ~ $ ls -l /pnfs/desy.de/m/project/petra4/
total 1
drwxr-x--- 3 p4madm p4m 512 Aug 31  2022 intern
drwxr-xr-x 2 p4madm p4m 512 Aug 31  2022 public
[vossc@naf-m01] ~ $ ls /pnfs/desy.de/m/project/petra4/intern/
ls: cannot open directory /pnfs/desy.de/m/project/petra4/intern/: Permission denied
```

Problems with POSIX Permissions

Basic Principles

- Everyone has their UID and their primary GID

```
[vossc@naf-it01] ~ $ id
uid=16490(vossc) gid=1000(it) groups=1000(it),...
[vossc@naf-it01] ~ $ id aeichler
uid=33375(aeichler) gid=1432(msk) groups=1432(msk),9491(msk_ipc_ds),...
```

- So any file anyone writes:

```
[vossc@naf-it01] ~ $ ls /nfs/dust/it/user/vossc/nexcloud-cli.sif  -l
-rwxr-xr-x 1 vossc it 354099200 Feb  9  2022 /nfs/dust/it/user/vossc/nexcloud-cli.sif
[vossc@naf-m01] ~ $ ls /pnfs/desy.de/m/project/iDAS/work/README -l
-rw-r--r-- 1 mahoff idas 719 Feb  9 08:22 /pnfs/desy.de/m/project/iDAS/work/README
```

- Reflected by the DAQ machines

```
[vossc@naf-m01] ~ $ ls /pnfs/desy.de/m/group/msk/lbsync/public/ -l
total 1
drwxrwxr-x 4 lbsyncds msk_lbs_ds 512 Mar 24 16:22 flash
drwxrwxr-x 5 lbsyncds msk_lbs_ds 512 Mar 24 16:21 xfel
```

Problems with POSIX Permissions

Zoo of Primary Groups

- Straight forward ➔ should be applicable to all our workflows? Sadly not:

- Works very well in closed well defined setups

```
[vossc@naf-belle12] ~ $ ls /pnfs/desy.de/belle/local/user/ -l | grep vossc
drwxrwxr-x 32 vossc    af-belle2 512 Feb 18  2021 vossc
[vossc@naf-belle12] ~ $ ls /pnfs/desy.de/belle/local/group/ -l | grep '^d' | head -n 2
drwxr-xr-x  5 belle2   af-belle2 512 Apr  5 15:51 belle2
drwxrwxr-x  5 belle2   af-belle2 512 May  7 21:08 hephy
```

- Here: all Belle II users share the same group: **af-belle2**

- Not the case the case with the M department

```
[root@dcache-door-m01 ~]# id kluth
uid=5252(kluth) gid=1500(mpy) groups=1500(mpy),3651(zemax_user),...
[root@dcache-door-m01 ~]# id tlamb
uid=18284(tlamb) gid=1432(msk) groups=1432(msk),5536(owncloud),...
```

- It would only work if MPY or MSK would only work among themselves

- Primary group structure does not reflect the day-to-day work praxis

- On NAF there comes in another detail

DESY.

Dealing with Permission Issues

NAF Peculiarities

Let us start with the NAF, because it is special:

- Solved half way on the NAF: when logging on or submitting a job to NAF you will be assigned a primary group

```
[vossc@naf-m01] ~ $ id kluth
uid=5252(kluth) gid=7723(af-m) groups=7723(af-m),1000064(helmholtz-member-f),...,1500(mpy)
[vossc@naf-m01] ~ $ id vossc
uid=16490(vossc) gid=7723(af-m) groups=7723(af-m),1000252(hip-storage-hzi-histo-f),...,1000(it)
```

- Carsten and I share the same primary group: **af-m ; but we keep all our groups**

- But the NAF groups are typically not related to the DESY internal groups

- Outside of the NAF everyone writes with their primary DESY group (**true also for Maxwell**)

```
[vossc@naf-m01] vossc $ pwd
/pnfs/desy.de/m/user/vossc
[vossc@naf-m01] vossc $ ls ../ -l | grep vossc
drwx------ 4 vossc    it    512 May 25 13:56 vossc
[vossc@naf-m01] vossc $ ls -l
total 0
-rw-r--r-- 1 vossc it   0 May 25 13:56 test_Maxwell
-rw-r--r-- 1 vossc af-m 0 May 25 13:56 test_NAF
```

- If you stick to NAF only, everything is fine, but that's not the case in reality, e.g. DAQ node are not part of NAF

Dealing with Permission Issues

Choosing your Groups Wisely

- Forego POSIX access, e.g. using token-based access through WebDAV or XrootD

  - Needs support from application, typically not the case

  - Allow some dCache internal group mapping (basically what happens in the LHC compute grid)

- Choose and build groups with the projects in mind: **example PETRA 4**

  - Project involves multiple M groups, at some point colleagues from FS or even external colleagues

  - Regular DESY group setup fails completely unless everything is world readable and even writeable

  - Create a project specific group and add the users you want there yourself

```
[vossc@naf-m01] petra4 $ pwd
/pnfs/desy.de/m/project/petra4
[vossc@naf-m01] petra4 $ ls -l
total 1
drwxr-x--- 3 p4madm p4m 512 Aug 31  2022 intern
drwxr-xr-x 2 p4madm p4m 512 Aug 31  2022 public
```

  - Only group-members can access the internal directory

Dealing with Permission Issues

Changing your Groups Wisely

- Choose and build groups with the projects in mind : still have the issue with the primary groups

- But each user may do a `chgrp` to another group he is a member of

  - Before sharing data with a colleague, check the ownership of files and directories

  - Granting read to all users an option, but makes data de-facto public

```
[vossc@naf-m01] vossc $ ls -l
total 0
-rw-r--r-- 1 vossc it   0 May 25 13:56 test_Maxwell
-rw-r--r-- 1 vossc af-m 0 May 25 13:56 test_NAF
[vossc@naf-m01] vossc $ chgrp it test_NAF
[vossc@naf-m01] vossc $ ls -l
total 0
-rw-r--r-- 1 vossc it 0 May 25 13:56 test_Maxwell
-rw-r--r-- 1 vossc it 0 May 25 13:56 test_NAF
[vossc@naf-m01] vossc $ chgrp msk_lbs_ds test_NAF
[vossc@naf-m01] vossc $ ls -l
total 0
-rw-r--r-- 1 vossc it         0 May 25 13:56 test_Maxwell
-rw-r--r-- 1 vossc msk_lbs_ds 0 May 25 13:56 test_NAF
```

- Needs to be done manual or as part of a workflow (e.g. through functional account)

Dealing with Permission Issues

Write Data Wisely

- Choose and build groups with the projects in mind : still have the issue with the primary groups

- Chose an (ideally) functional account to write data whose primary group matches your project (if possible)

```
[root@dcache-door-m01 ~]# ls -l /pnfs/desy.de/xfel-linac/real/2023/
total 2
drwxr-xr-x. 3 doocsadm doocsadm 512 Jan 13 13:46 linac
drwxr-xr-x. 3 doocsadm doocsadm 512 Jan  9 16:34 linac_test
drwxr-xr-x. 3 doocsadm doocsadm 512 Jan  9 16:54 llrf_gun
```

  - Here only the functional account would have access (neglecting the `755` permissions)

```
[root@dcache-door-m01 ~]# ls -l /pnfs/desy.de/m/group/msk/lbsync/
total 1
drwxr-x---. 3 lbsyncds msk_lbs_ds 512 Apr 28 14:48 intern
drwxrwxr-x. 4 lbsyncds msk_lbs_ds 512 Dec 20 16:28 public
```

  - Here all members have access to the data (without needing it to be `755`)

- Functional account can write with by itself limited permissions

- Might only fit specific use cases, c.f. xfel-linac vs. lb-sync

Access to dCache

Different Systems that Allow Access to the Data

- As mentioned dCache supports a wide area of protocols:

  - NFS: used heavily on NAF and when taking data (DAQ nodes write to dCache via NFS)
  - XrootD: used within the LHC experiments, for local and remote access (ideal for ROOT files)
  - FTP: heavily used by LHC experiments in the past for WAN transfers
  - WebDAV: replaced FTP for LHC experiments; suggested protocol for smaller groups
  - dCap: dCache native format

- NFS Access:

  - Accessible on NAF and Maxwell as well all M internal nodes as requested
  - If node/mount is missing: open a ticket with it-DataManagement in request tracker or via osm.service@desy.de

- Access on Windows via Samba:

  - Can offer an export to Windows via Samba (looks like a network device); let us know if you are interested
  - Needs some work on our end

- Access via WebDAV (independent from M/IT nodes):

  - dCache is available via WebDAV to anyone with the correct permissions (login via DESY credentials)
  - Probably worth a small seminar to cover all the details

- Access through browser UI (independent from M/IT nodes):

  - Feel free to try. https://dcache-se-m.desy.de:3880

DESY.

Access to dCache

How to Create and Manage the Functional Accounts

- Mentioned functional accounts a couple of times:
  - Basically a DESY account with a different password policy that belongs to a number of individuals
  - Can be used to manage data independent of the currently responsible person
  - Hand over to Carsten to show how to do that in the DESY registry and how manage the members
- **Final remark in permissions:**
  - A number of you have asked about future access to colleagues outside of DESY
  - Through WebDAV it's easily done (practise this for the FH colleagues on PiB-level)
  - But: something that has permissions of `755` will be world readable
  - Currently the WebDAV door and Browser UI are still blocked from Non-DESY IPs
  - But: something that has permissions of `755` will be DESY readable
  - Not a topic for today but this would need to be discussed on a larger scale

```
[root@dcache-door-m01 ~]# ls -l /pnfs/desy.de/xfel-linac/real/2023/
total 2
drwxr-xr-x. 3 doocsadm doocsadm 512 Jan 13 13:46 linac
drwxr-xr-x. 3 doocsadm doocsadm 512 Jan  9 16:34 linac_test
drwxr-xr-x. 3 doocsadm doocsadm 512 Jan  9 16:54 llrf_gun
```

- Linac data is save as we only expose `/pnfs/desy.de/m` currently

Thank you, any Questions