



dCache - Inter-Disciplinary Storage

29 April 2024

Marina Sahakyan for the dCache collaboration



HELMHOLTZ

RESEARCH FOR
GRAND CHALLENGES



“... to provide a system for storing and retrieving huge amounts of data, distributed among a large number of heterogeneous server nodes, under a single virtual filesystem tree with a variety of standard access methods.”

<https://dcache.org/about/>

Scientific Data Challenges



Ingest

- High data ingest rate
- Multiple parallel streams
- High durability
- Effective handling of large number of files

Analysis

- High CPU efficiency
- Chaotic access
- Standard access protocols
- Access control
- Local user management

Sharing & Exchange

- 3rd party copy
- Effective WAN Access
- In-flight data protection
- Identity federation
- Access control

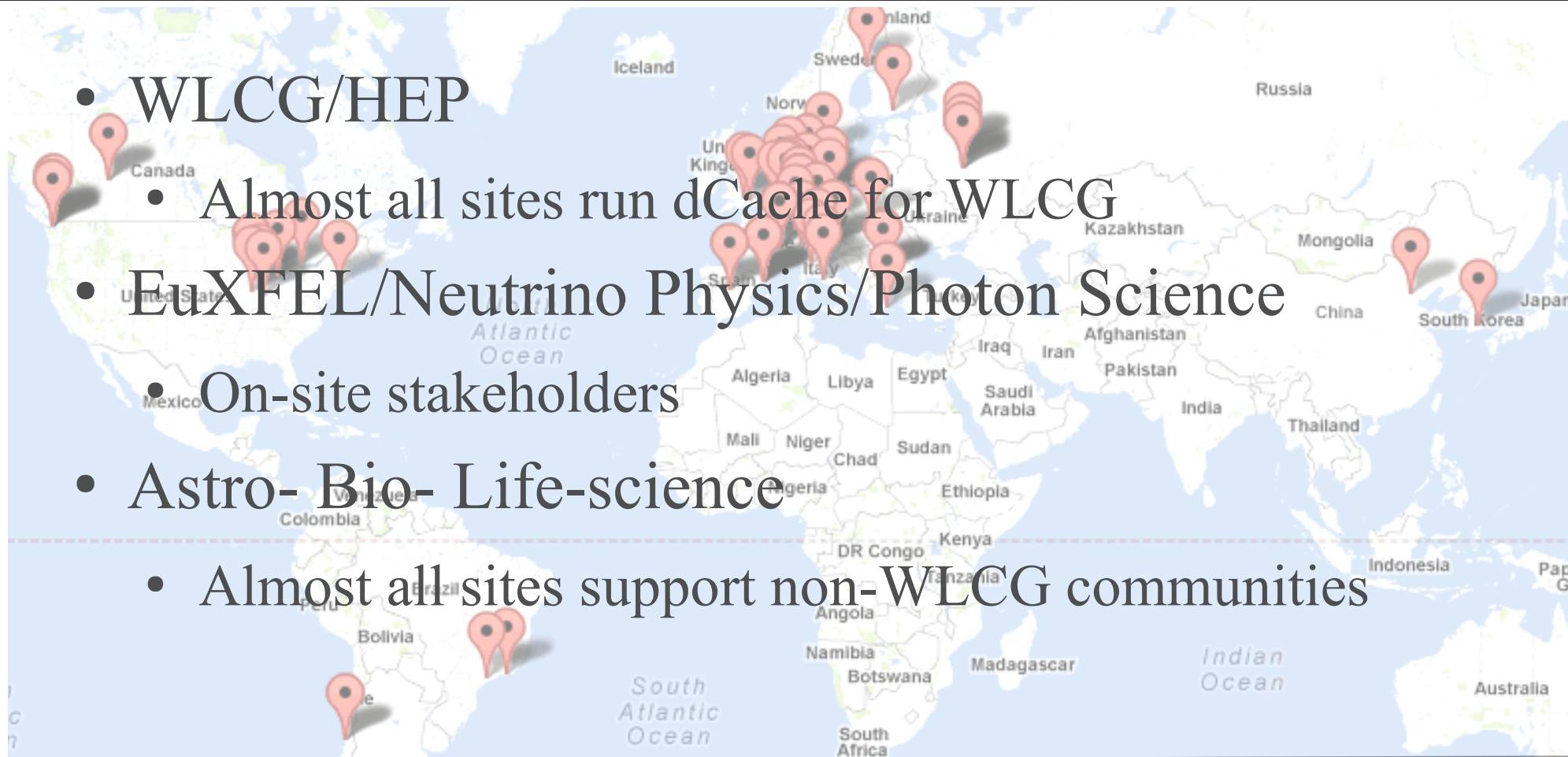
Long Term Preservation

- High Reliability
- Self-healing
- Automatic technology migration
- Persistent identifier

Strategic Communities



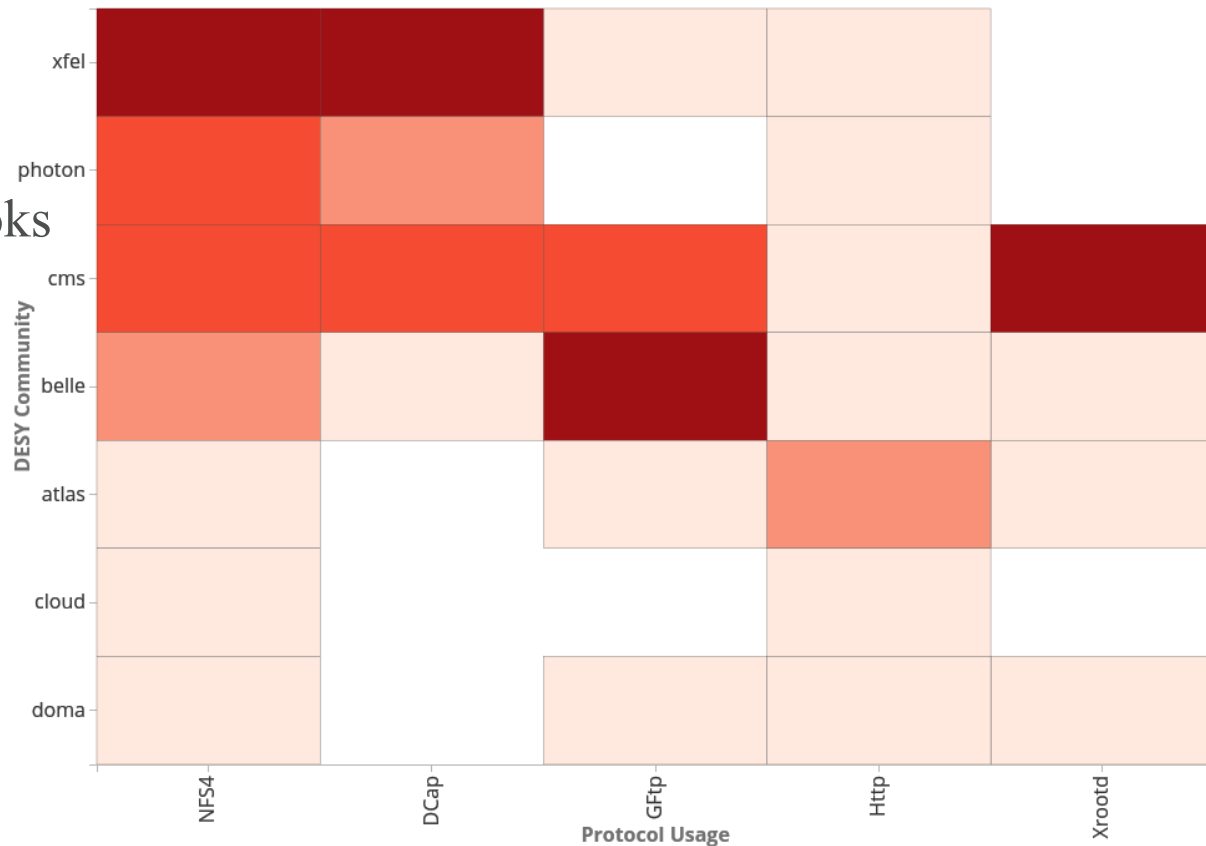
- WLCG/HEP
 - Almost all sites run dCache for WLCG
- EuXFEL/Neutrino Physics/Photon Science
 - On-site stakeholders
- Astro- Bio- Life-science
 - Almost all sites support non-WLCG communities



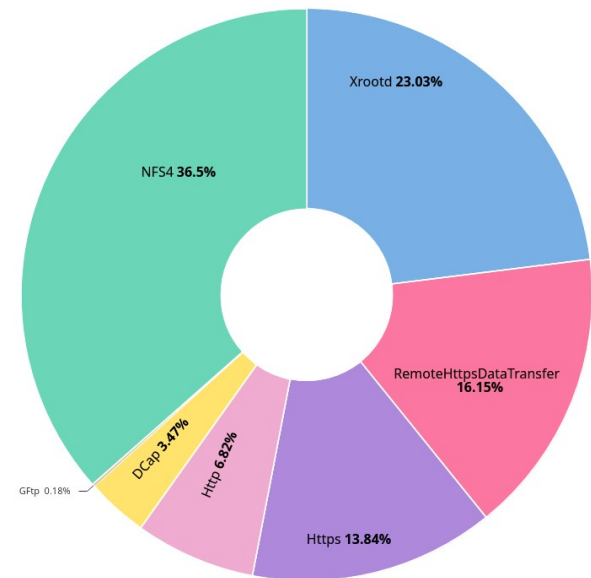
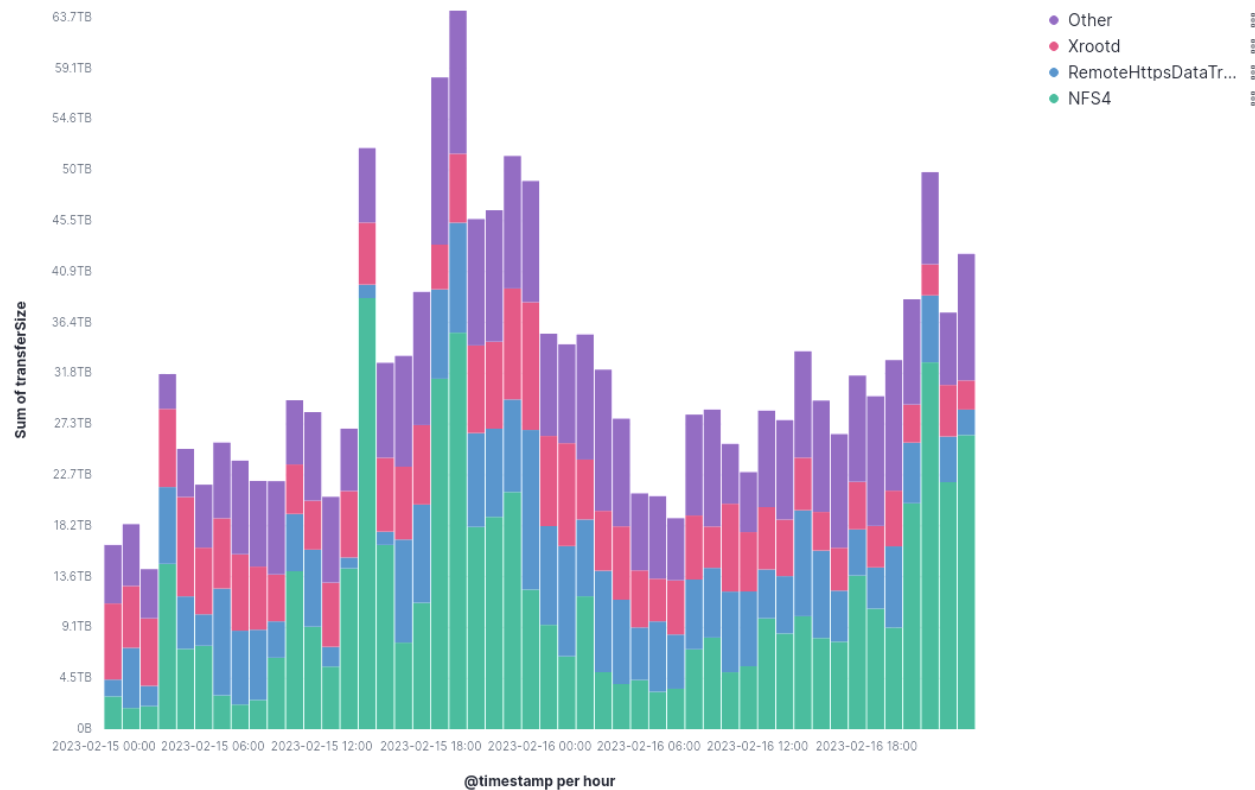
Data Access Variety



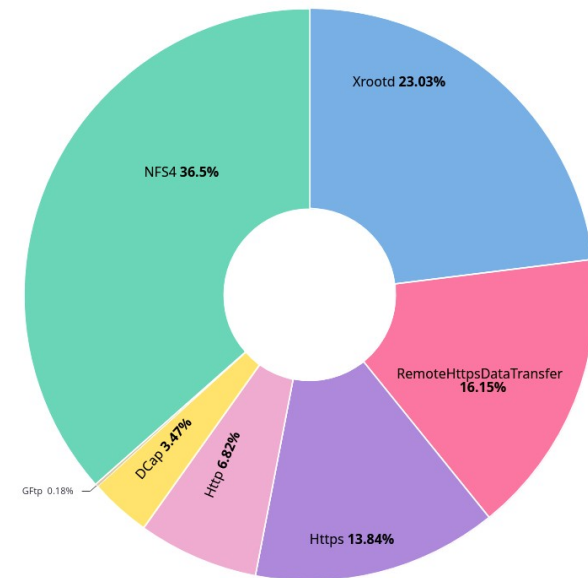
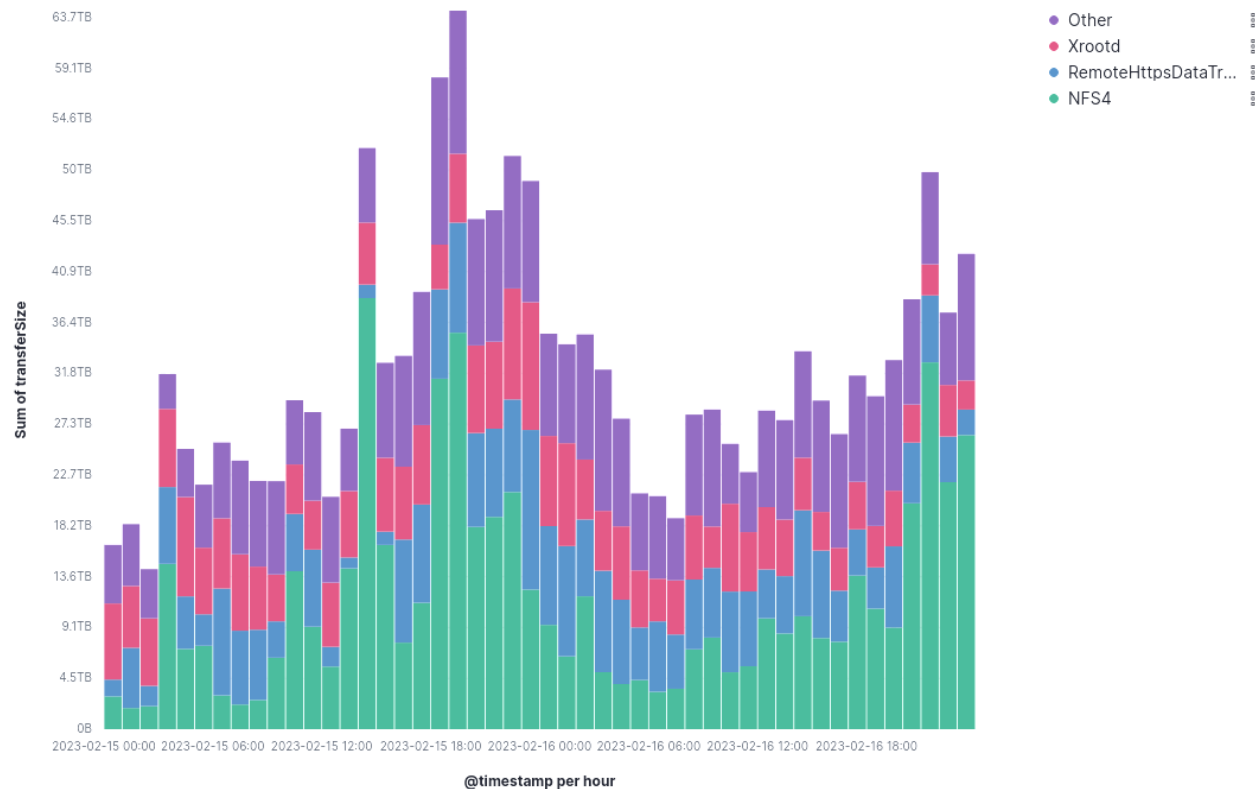
- ROOT-IO
- Non-HEP tool chain
 - Active use of Jupyter Notebooks
 - Non-ROOT data formats
- Industry standard AuthN
 - Tokens based authentication
 - Federated IdP
- Use of private clouds
 - Data access from a container
- Use of HPC resources



CMS Access Profile



CMS Access Profile



resources at DESY

- grid cluster
- workgroup servers
- analysis facility

Bulk REST-API (like SRM, but different)



STAGE

- Request to stage many files at once

CANCEL

- Cancel bulk request

DELETE

- Cancel bulk request + clear history/status

EVICT

- unpin cached copy

PIN

- Pin cached copies with a lifetime

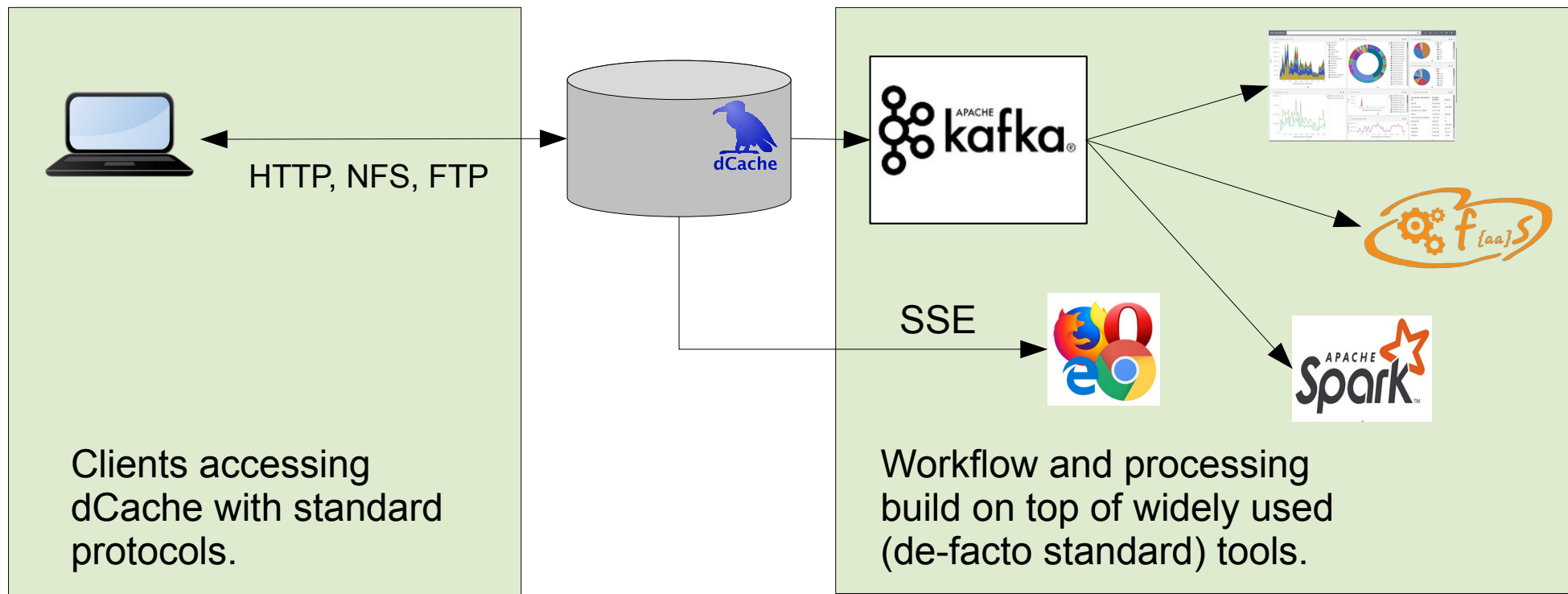
FILEINFO

- Request status many files at once (locality, checksum)

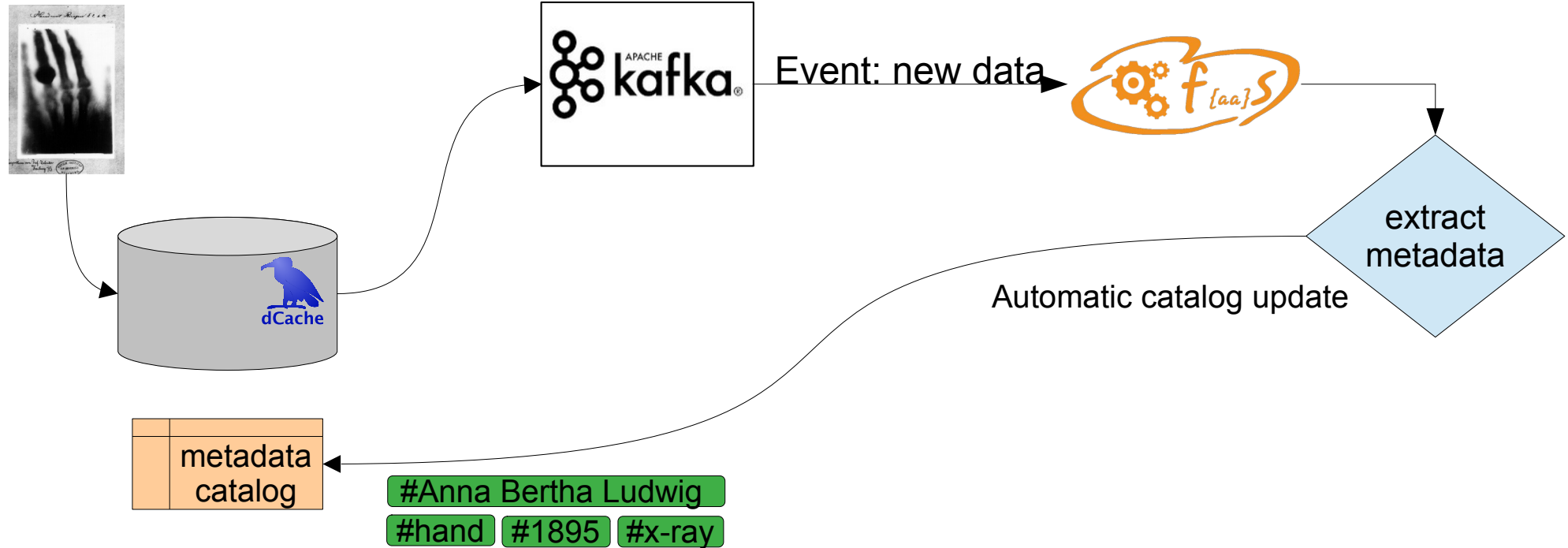


StoRM

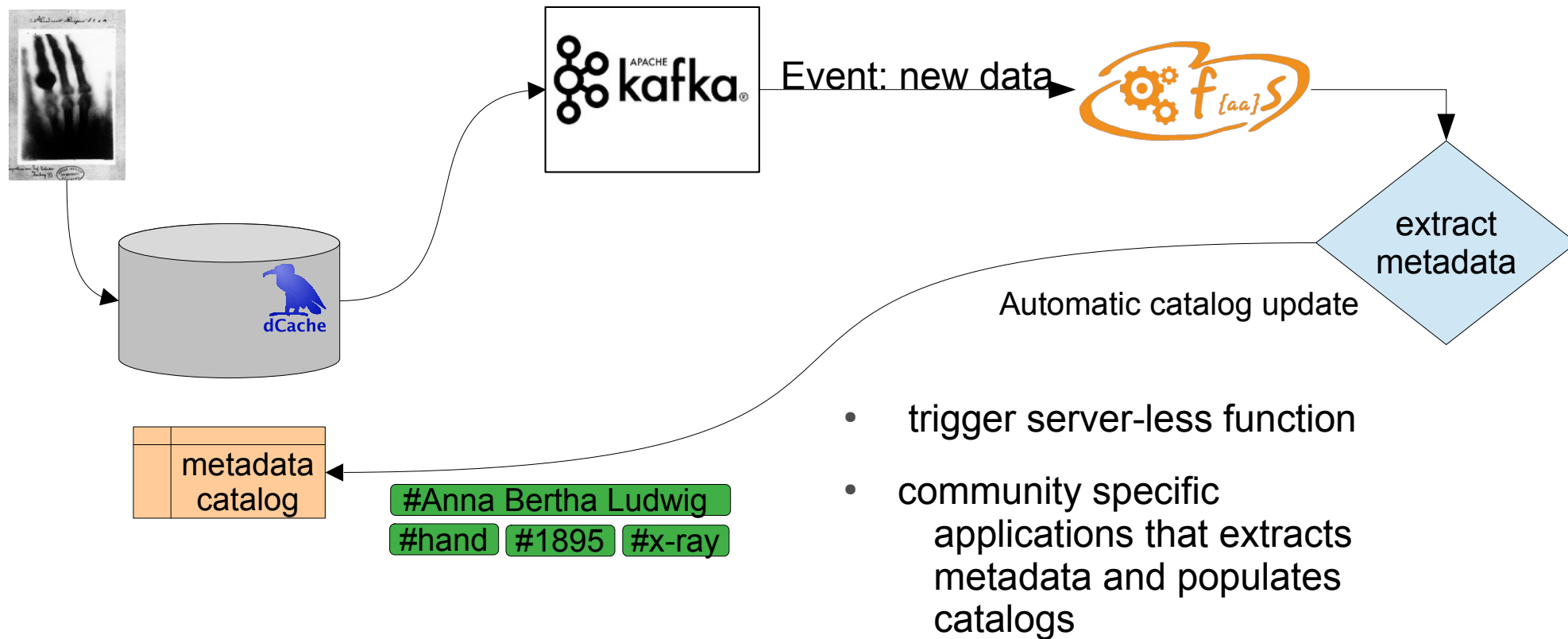
Standards Everywhere...



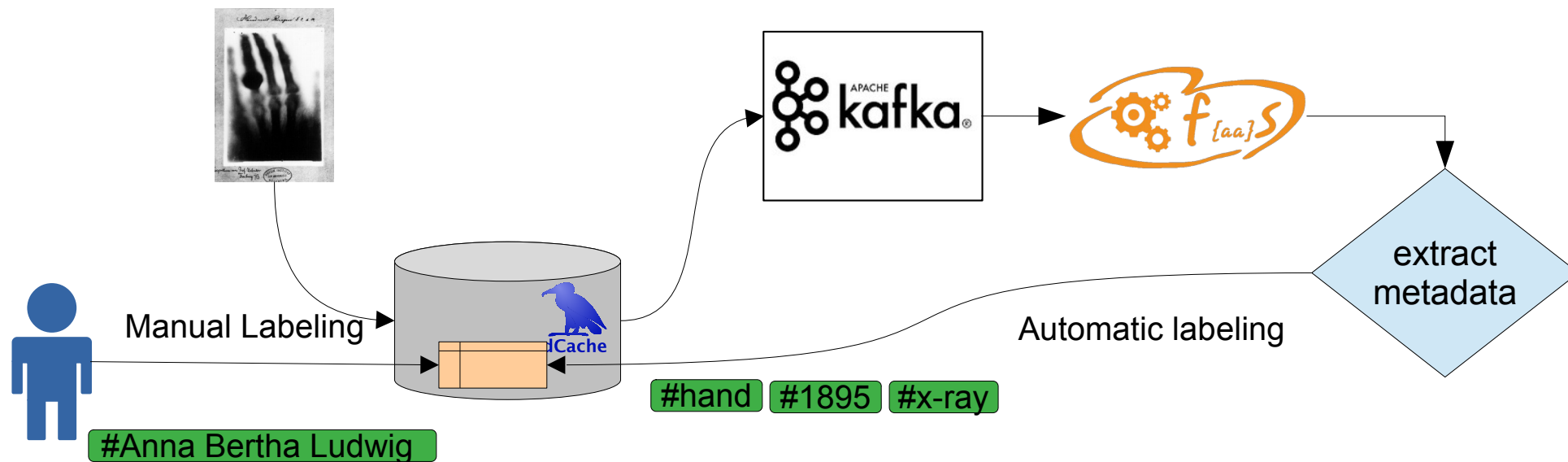
Automatic Metadata Population



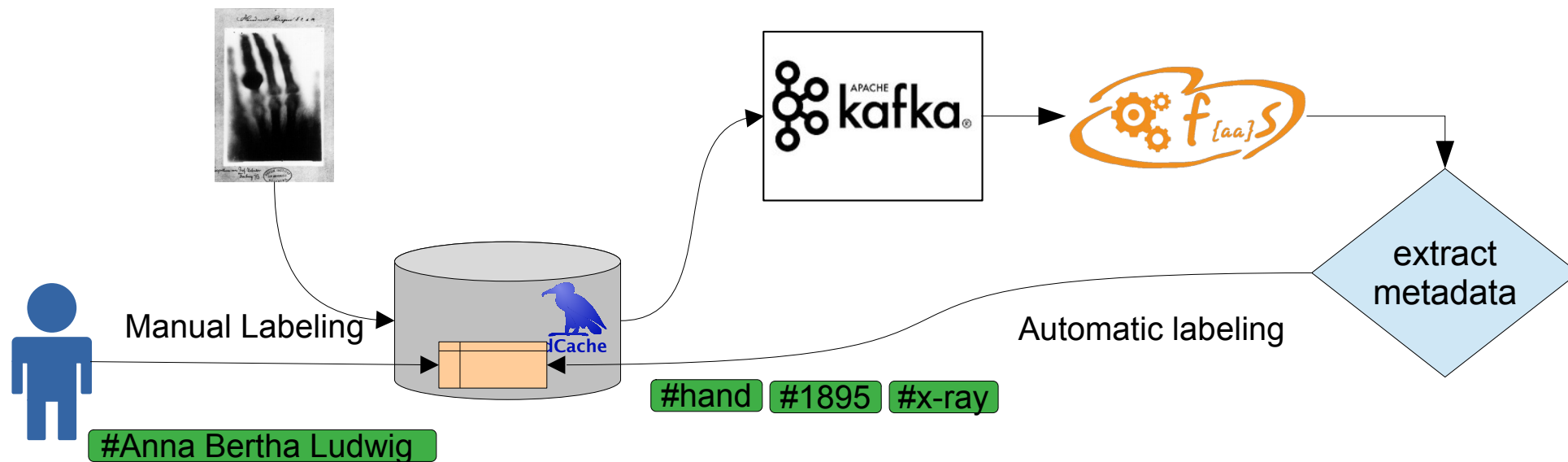
Automatic Metadata Population



Metadata Population



Metadata Population

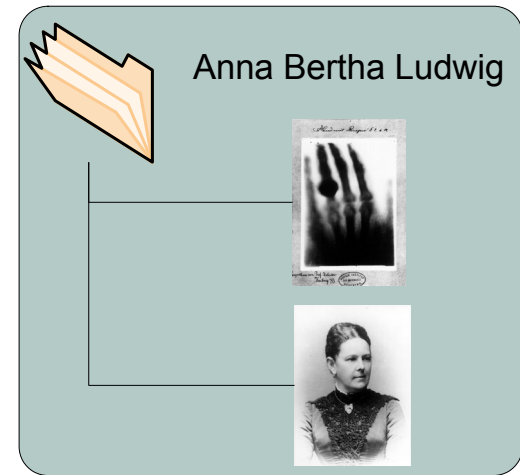


Up to now such catalogs were external to dCache.

User Metadata/Labeling in dCache



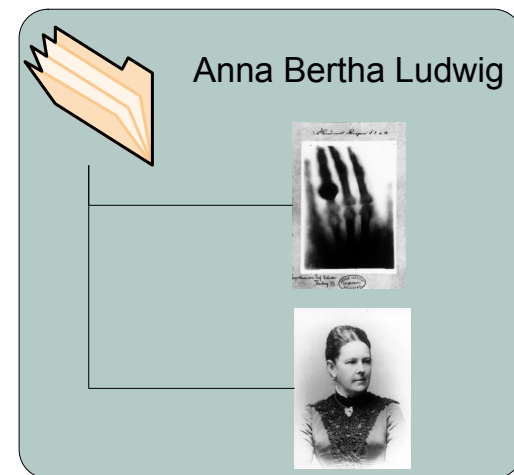
- Extended attributes
 - Exposed via NFS, WebDAV, REST
- Label-based virtual **read-only** directories (WIP)
 - List all files with a given label
- dCache rules applies
 - Visible through all protocols
 - Respect file/dir permissions



User Metadata/Labeling in dCache



- Extended attributes
 - Exposed via NFS, WebDAV, REST
- Label-based virtual **read-only** directories (WIP)
 - List all files with a given label
- dCache rules applies
 - Visible through all protocols
 - Respect file/dir permissions



Setting labels



```
[marina@zitpcx35188 dcache]$ curl -X POST -u admin:dickerelch -H  
"Accept: application/json" -H "Content-Type: application/json"  
http://localhost:3880/api/v1/namespace/tape/file_1.log/ -d '{"action" : "set-label",  
"label" : "beamline12" }'
```

```
[marina@zitpcx35188 dcache]$ curl -X POST -u admin:dickerelch -H  
"Accept: application/json" -H "Content-Type: application/json"  
http://localhost:3880/api/v1/namespace/tape/file_1.log/ -d '{"action" : "set-label",  
"label" : "proposal231" }'
```




```
[marina@zitpcx35188 dcache]$ curl -u admin:dickerelch
http://localhost:3880/api/v1/namespace/tape/file_1.log\?labels=true
{
  "fileMimeType" : "application/octet-stream",
  "labels" : [ "proposal231", "beamline12" ],
  "pnfsId" : "0000323FF30B54DF42E782750E8660B13D6D",
  "fileType" : "REGULAR",
  "nlink" : 1,
  "mtime" : 1714332844117,
  "mode" : 420,
  "size" : 387,
  "creationTime" : 1714332843857
}
```





Query labels/ nfs mount



```
[marina@zitpcx35188 dcache]$ ls -la /mnt/".(collection)(beamline12)"
total 2
drwxr-xr-x. 13 root  root  512 Apr 28 21:44 .
-rw-r--r--.  1 marina marina 387 Apr 28 21:34 file_1.log-2
-rw-r--r--.  1 marina marina 387 Apr 28 21:34 file_1.log-7
-rw-r--r--.  1 marina marina 387 Apr 28 21:34 file_2.log-2
```

Listing files webdav














localhost:2880/.(collection)(beamline12)/

dCache

system-test (built from 2b10736)



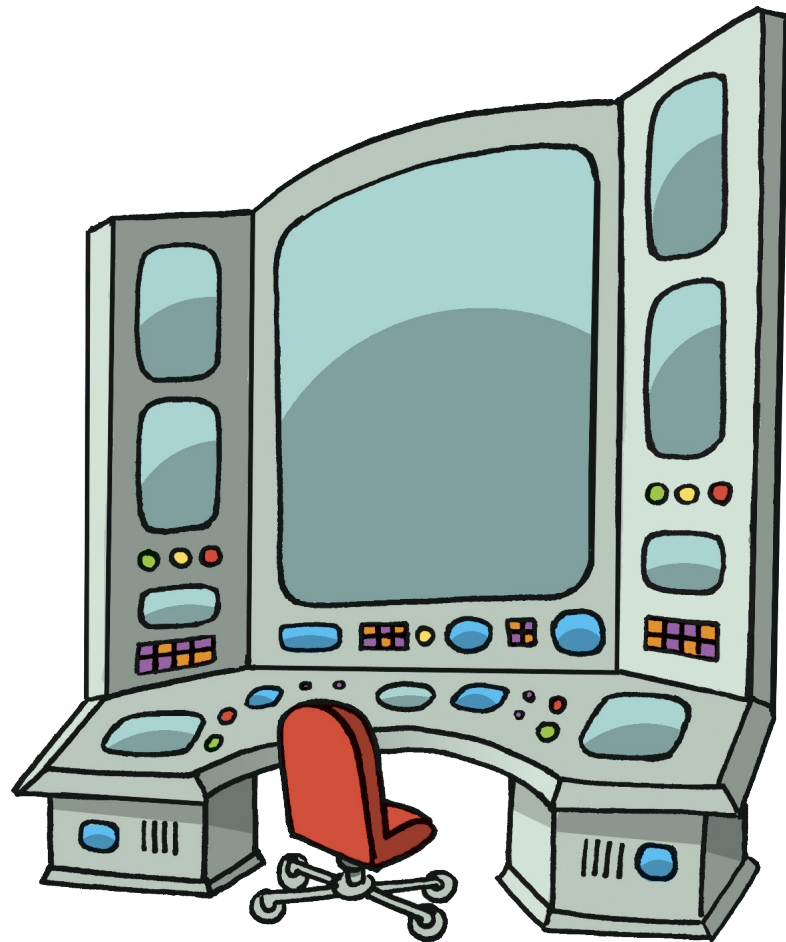
 / [.\(collection\)\(beamline12\)](#)

	Name	 Size	 Last Modified
	file_1.log-2 	387	Sun Apr 28 21:34:04 CEST 2024
	file_1.log-7 	387	Sun Apr 28 21:34:12 CEST 2024
	file_2.log-2 	387	Sun Apr 28 21:34:04 CEST 2024

Some Numbers



- XFEL
 - Total capacity ~120 PB
 - ~400 physical hosts (~4000 dCache pools)
 - 20-40 GB/s inject
- Photon
 - DB size – 2.5TB
 - ACL table 600GB
 - Directories with $3 \cdot 10^6$ files
 - $1.2 \cdot 10^9$ file system objects
 - 100K files in the flush queue
 - Two tape copies, different media type
- ATLAS
 - dir/file → 1/3
- NextCloud
 - File lifetime < 1s



Supported OS platforms



- 8.2
 - RHEL 7, 8, 9
 - JVM 11
- 9.0 (Feb. 2023)
 - RHEL 7, 8, 9
 - JVM 11, 17
- 10.0 (~ 1Q 2024)
 - RHEL 8, 9
 - JVM 17

The screenshot shows the dCache.org website with the URL <https://www.dcache.org/downloads/> in the browser address bar. The page features a dark header with the dCache logo and navigation links: Main, Posts, Downloads (highlighted), Releases, Documentation, Support, About Us, and Developer's Corner. The main content area is titled "Downloads" and is divided into two sections: "Binary packages" and "Unsupported releases".

Binary packages

- **v9.2.x** Latest Golden Release
- **v9.1.x** Feature Release
- **v9.0.x** Feature Release
- **v8.2.x** Golden Release

Unsupported releases

- **v8.1.x** Feature Release
- **v8.0.x** Feature Release
- **v7.2.x** Golden Release
- **v7.1.x** Feature Release
- **v7.0.x** Feature Release
- **v6.2.x** Golden Release
- **v6.1.x** Feature Release
- **v6.0.x** Feature Release
- **v5.2.x** Golden Release
- **v5.1.x** Feature Release
- **v5.0.x** Feature Release
- **v4.2.x** Golden Release

RECENT POSTS

- 18th International dCache Workshop
- 17th International dCache Workshop
- Vulnerability in PostgreSQL server
- 16th International dCache Workshop
- Log4j 1.2 Vulnerability

CATEGORIES

- Info
- workshop

TAGS

- dcache.org
- security
- web
- workshop

Summary & Conclusions



- The dCache team has been providing a reliable software to manage scientific data for over 20 years.
- Seamless integration into the site's infrastructure makes dCache a natural part of any data center.
- Multi-protocol and authentication scheme capabilities allow to support multiple communities even on a single instance.
- In close cooperation with experiments we address today's and future data management challenges.



Thank You!

18th International
dCache Workshop
June 6-7, DESY-Hamburg

More info:

<https://dcache.org>

To steal and contribute:

<https://github.com/dCache/dcache>

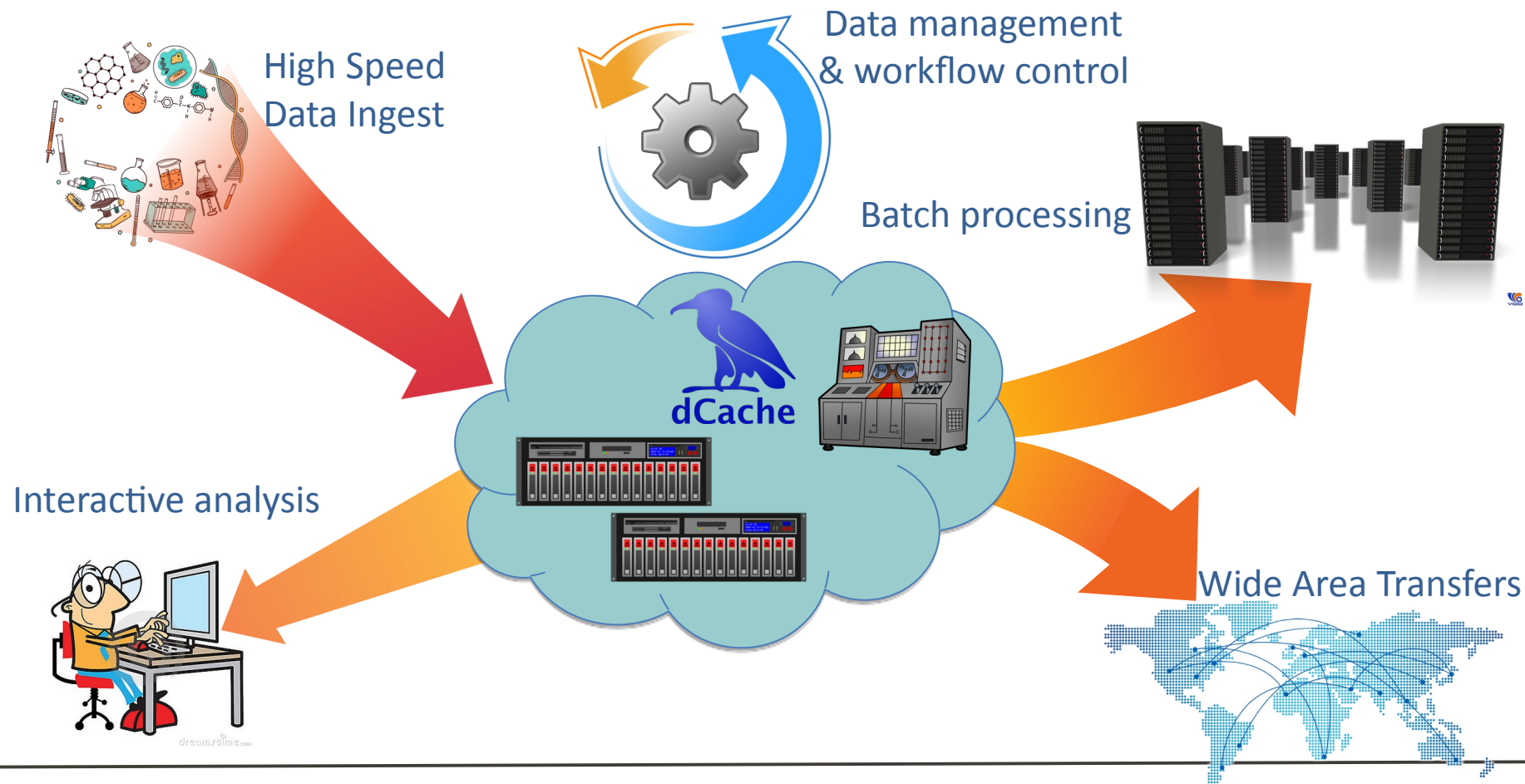
Help and support:

support@dcache.org, [user-forum@dcache.org](https://user-forum.dcache.org)

Developers:

dev@dcache.org







- ATLAS “Tape carousel” => WLCG “Data carousel”
 - Lea’s presentation: Improving Performance of Tape Restore
- High number of small files by Photon Science
 - ~4MB, 10^6 files per directory
 - <https://indico.cern.ch/event/995485/contributions/4256474/>
- Multi-media copy guarantees
 - QoS
- Integration with CERN Tape Archive (CTA)
 - Close work with CERN team to bring CTA support to dCache

Technical Directions



- Scaleout
 - Namespace
 - Number of pools (cells)
- Token-based Authentication
- Better *Analysis Facility* support
 - POSIX access and compliance
 - HPC workload support (DDoS protection)
- QoS
- Tape integration
- Green IT





- Two main gaps to fill
 - Space allocation
 - Tape operation
- Two alternatives to replace
 - User and Group based Quota system
 - WLCG tape recall API



- Quota != Space reservation
- Lazy, based on periodic scans
 - Users might overrun
 - Removed space not reclaimed immediately
- Global per file system
 - No quota per directories
- Respects Files Retention policy
 - Separate for 'disk' and 'tape' files
- Available since 7.2, enabled by default since 8.2

Tape rest API



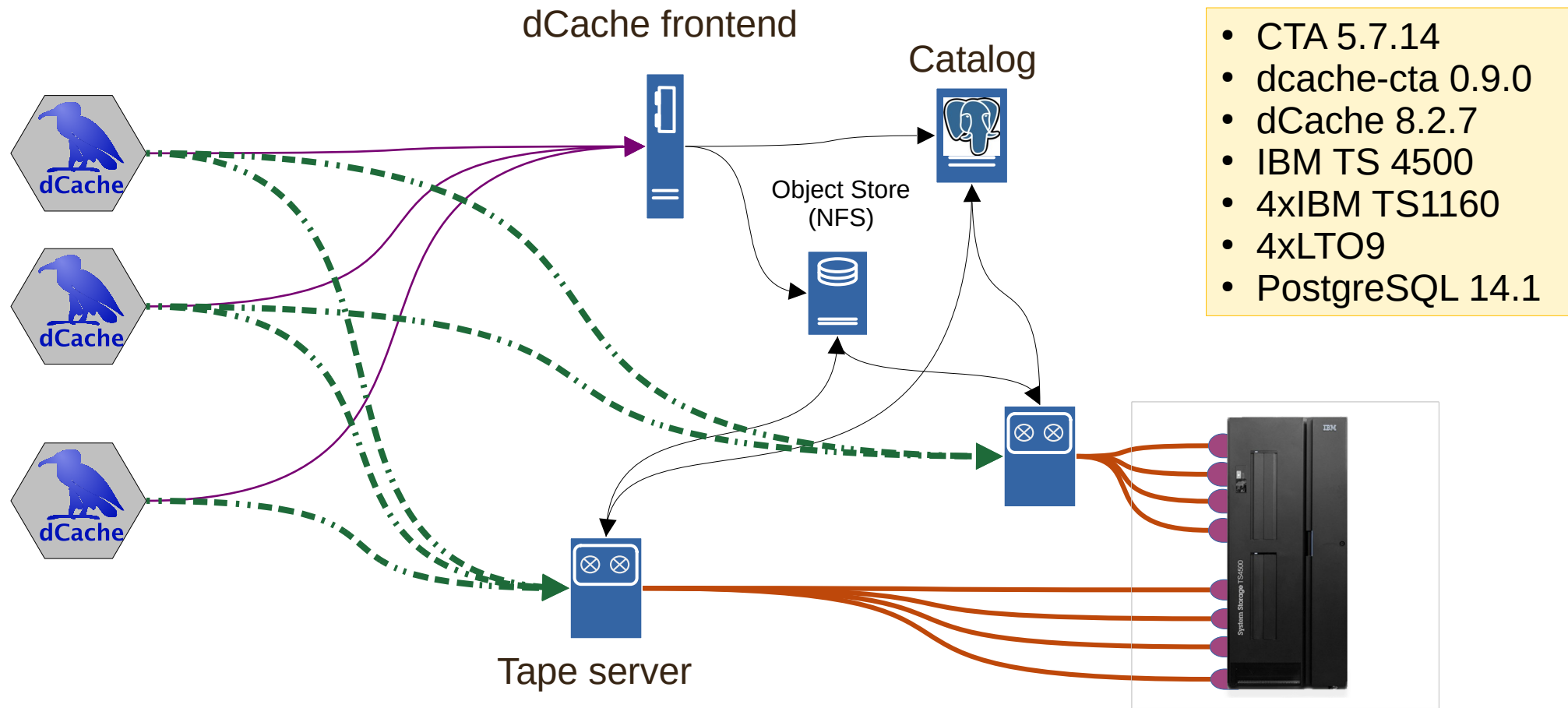
<https://example.org:3880/api/v1>

bulk-requests ▾		
GET	/bulk-requests/{id}	Get the status information for an individual bulk request.
DELETE	/bulk-requests/{id}	Clear all resources pertaining to the given bulk request id.
PATCH	/bulk-requests/{id}	Take some action on a bulk request.
GET	/bulk-requests	Get the status of bulk operations submitted by the user.
POST	/bulk-requests	Submit a bulk request.
archiveinfo ▾		
POST	/archiveinfo	Return the file locality information for a list of file paths.
release ▾		
POST	/release/{id}	RELEASE files associated with a STAGE request.
stage ▾		
POST	/stage/{id}/cancel	Cancel a STAGE request.
POST	/stage	Submit a STAGE request.
GET	/stage/{id}	Get the status information for an individual stage request.
DELETE	/stage/{id}	Clear all resources pertaining to the given stage request id.

dCache bulk API

WLCG Tape API

Production Deployment at DESY





PaNOSC

Laser-Driven Proton Acceleration from Cryogenic Hydrogen Target

Description

2D particle-in-cell simulation of the interaction of high-intensity laser pulse (parameters are relevant to L4 laser) with a cryogenic hydrogen target. Only protons with energy above 300 MeV at the end of the simulation are tracked and their position and energy are visualized. Two different groups of protons accelerated by different mechanisms can be distinguished from each other in space: Protons originated from the target interior and from the target rear side.

Citation	Dana Scully; (2020), Re-polarization of the aft quantum plasma collector, DOI:10.9563/if.2015.87.012
Keywords	X-ray excited optical luminescence,
Type	Proposal
Author	Laima Reinhold
Other	Stuff

Stolen from Michael Schuh

Datasets

PaNOSC Test Dataset 11

HEIMDAL @ ESS

Name

Description

Flavour

Spawn

Preview Visualization

