

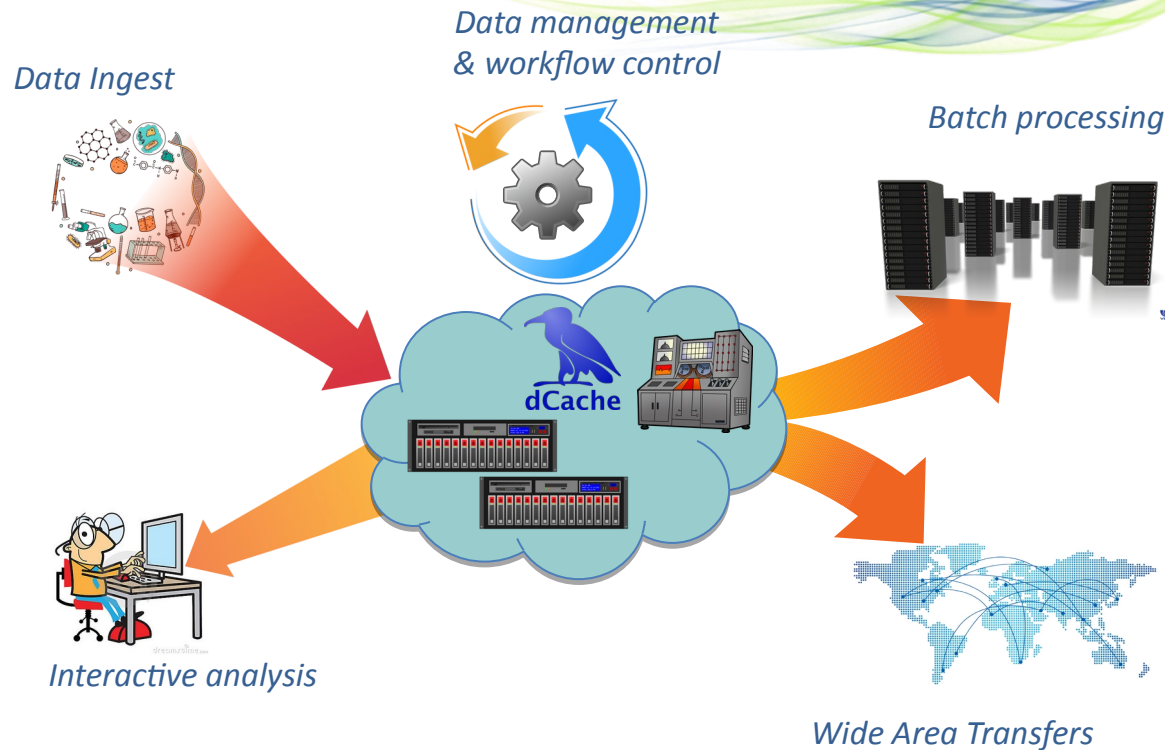
# dCache Roadmap

Tigran Mkrtchyan  
for dCache Team



# Strategic Direction

- WLCG
- DESY/Fermilab
  - Belle II
  - CTA
  - DUNE
  - EuXFEL
- SKA
- Long tail science
  - BioMed, Astro, Neutrino, LifeScience experiments



# Main developments

- Mainline distributed deployments
  - Network topology aware internals communication
  - Disconnected read-only operation
- HA-enhancements
- Storage events
- Data protection
- Standard access protocols and AuthN/Z
- Web admin and REST-api consolidation
- QoS
- Integration into existing infrastructure
  - native cloud integration



# User Workflow Shift

- More non HEP tools and POSIX access

- ROOT  $\Rightarrow$  Jupyter Notebook
- Apache Spark
- HDF5

- Growth of interactive analysis

- Analysis Facilities

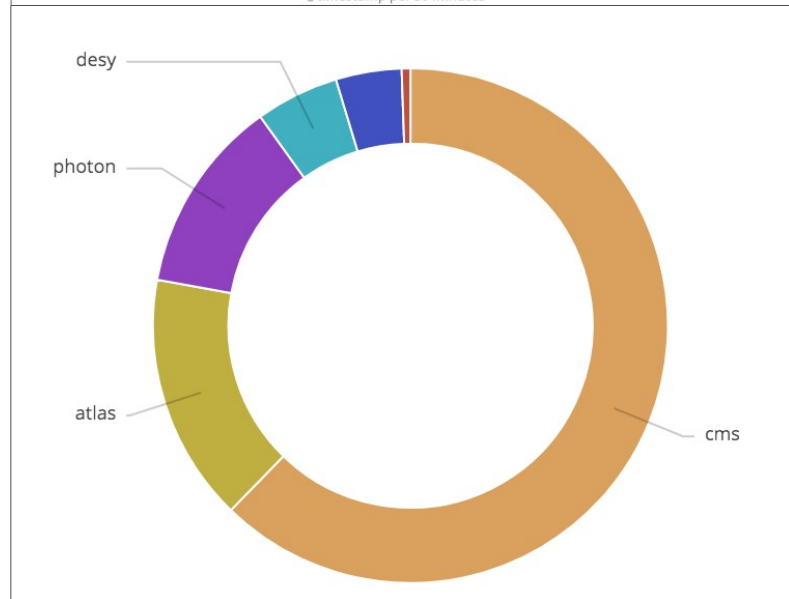
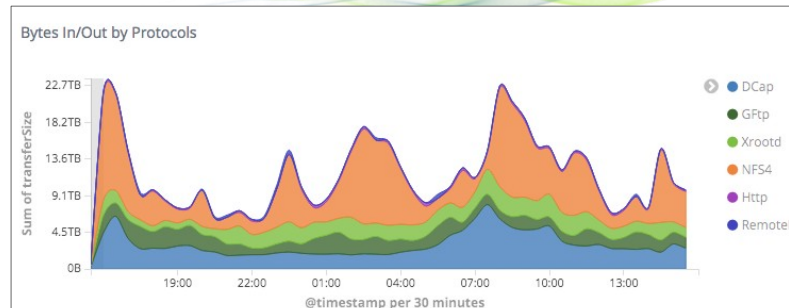
- Industry standard AuthN

- OpenID Connect
- OAuth2
- Federated IDP

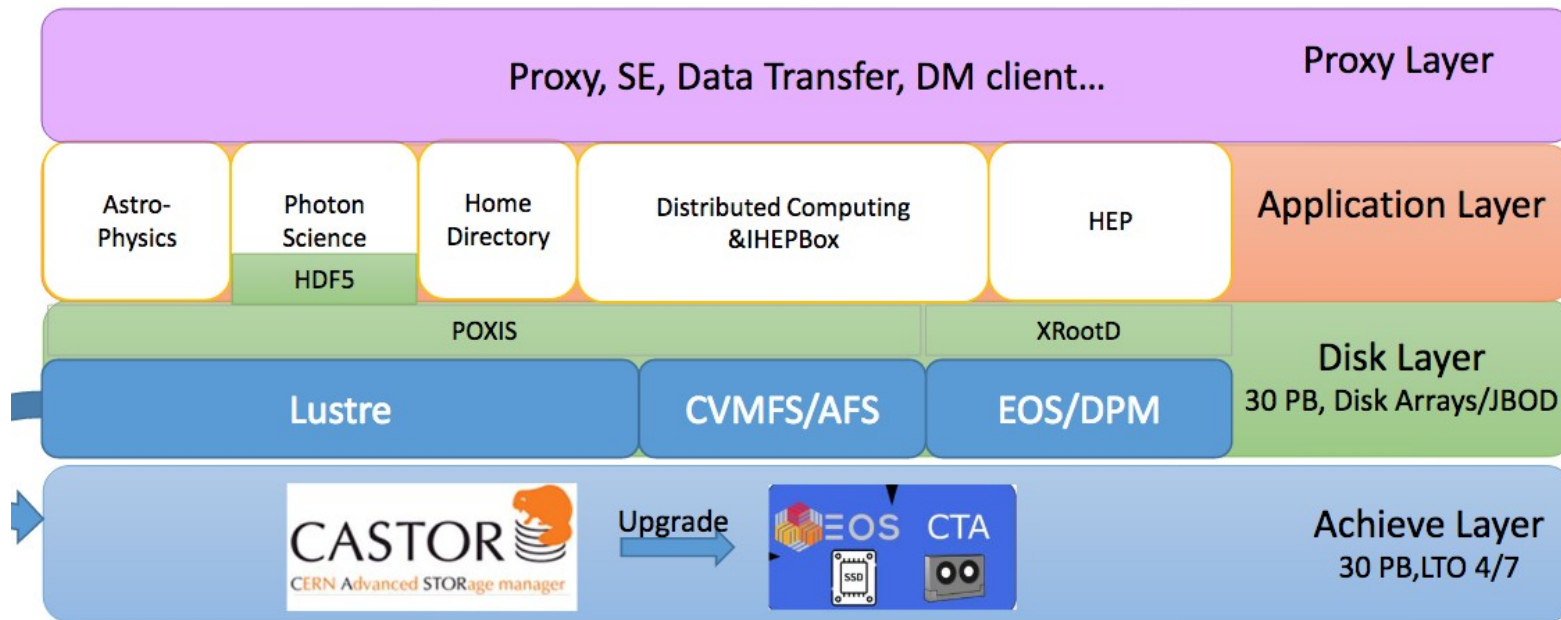
- Hybrid Clouds

- New 3<sup>rd</sup>-party transfers protocols

- Integration with HPC clusters



# We are Not Alone!



Stolen from Wang Lu

<https://indico.cern.ch/event/941278/contributions/4083276/>

# Tape Access

- WLCG
  - Atlas Tape carousel
  - Model considered by other experiments
- EuXFEL
  - Fast data ingest
  - Archival storage
  - Multiple media requirement

# Tape Driver Interface

- Simplify custom driver implementation
- New drivers
  - Native S3
  - Small file aggregation (WIP)
- Functionality enhancements



# Re-calling 80% of a tape in RAO\* hides mount & seek overhead!

\* Recommended Access Ordering

BACHELOR THESIS KOLLOQUIUM



## Improving Tape Restore Request Scheduling in the Storage System dCache

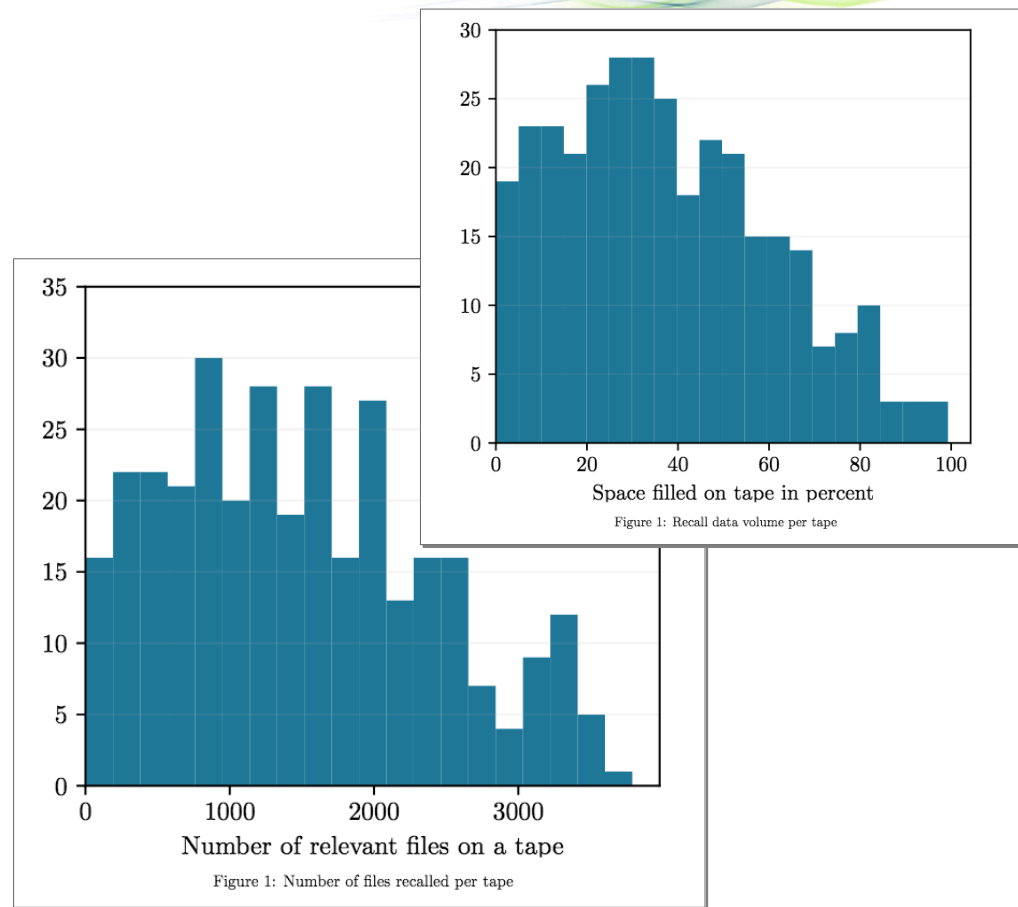
Evaluation of Pre-Scheduling Strategies by Disk Systems in Front of  
Automated Tape Storage in the Context of the ATLAS Experiment

Lea Morschel, March 2020

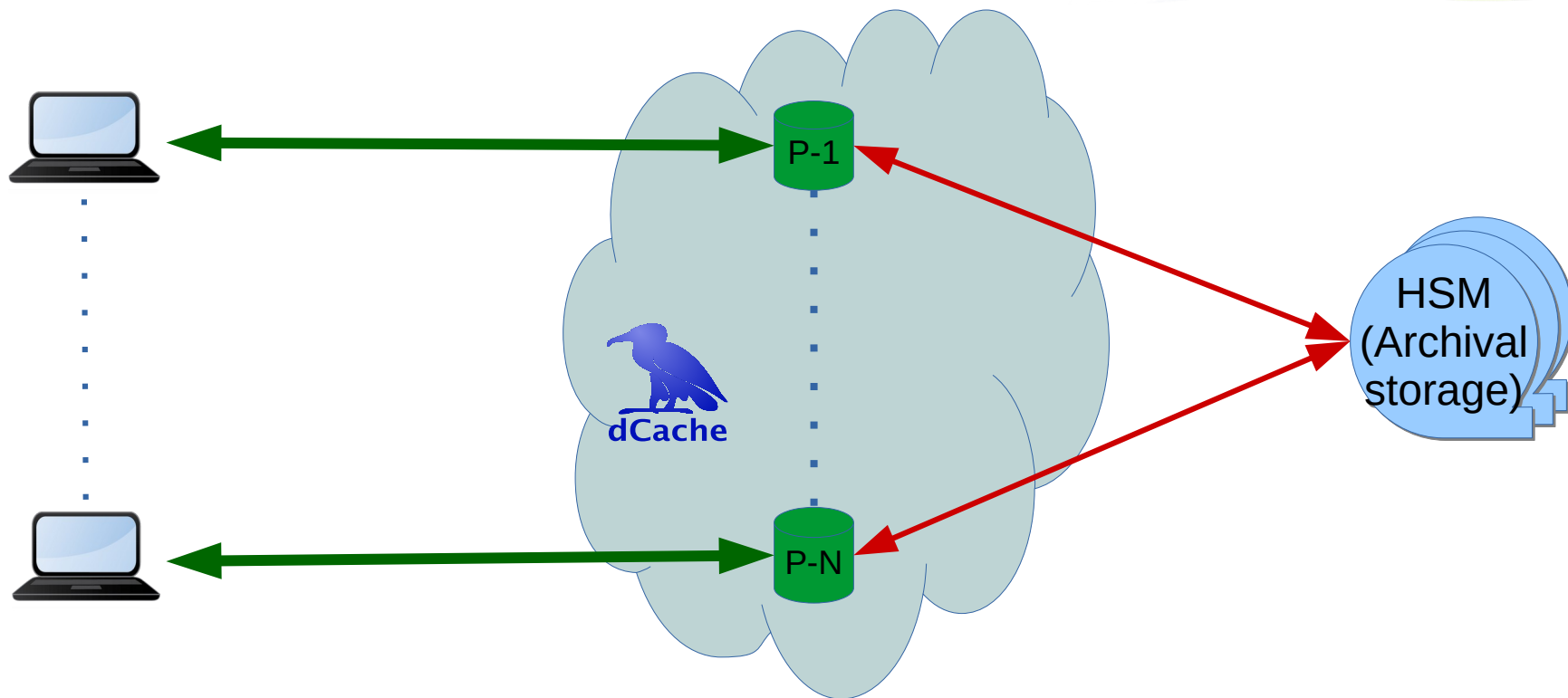


# KIT Under the Spotlight

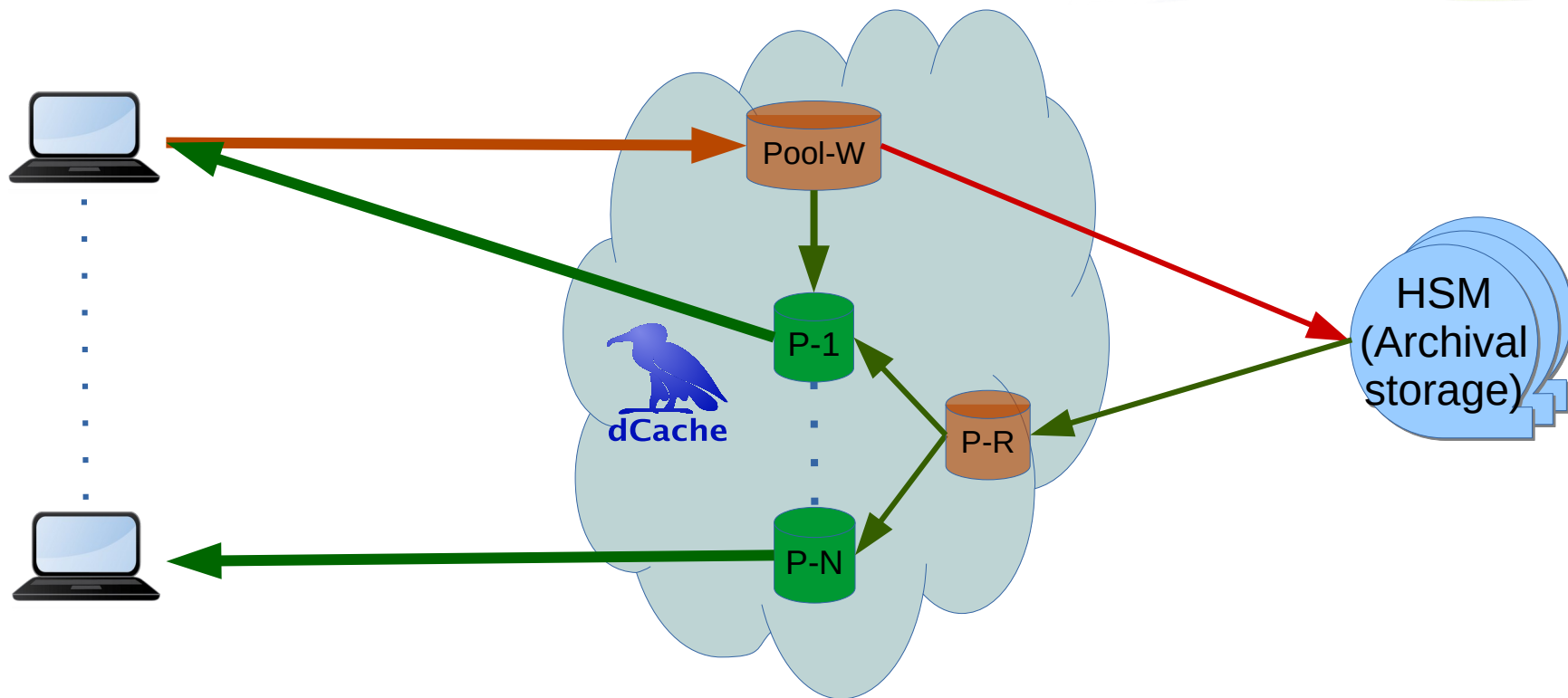
- Majority on dataset use less then 40% of tape
- Majority of tapes have up-to 3K files of a dataset
- To keep all drives busy, request queue should have enough data
  - ✓  $12 * 3K = 36K$



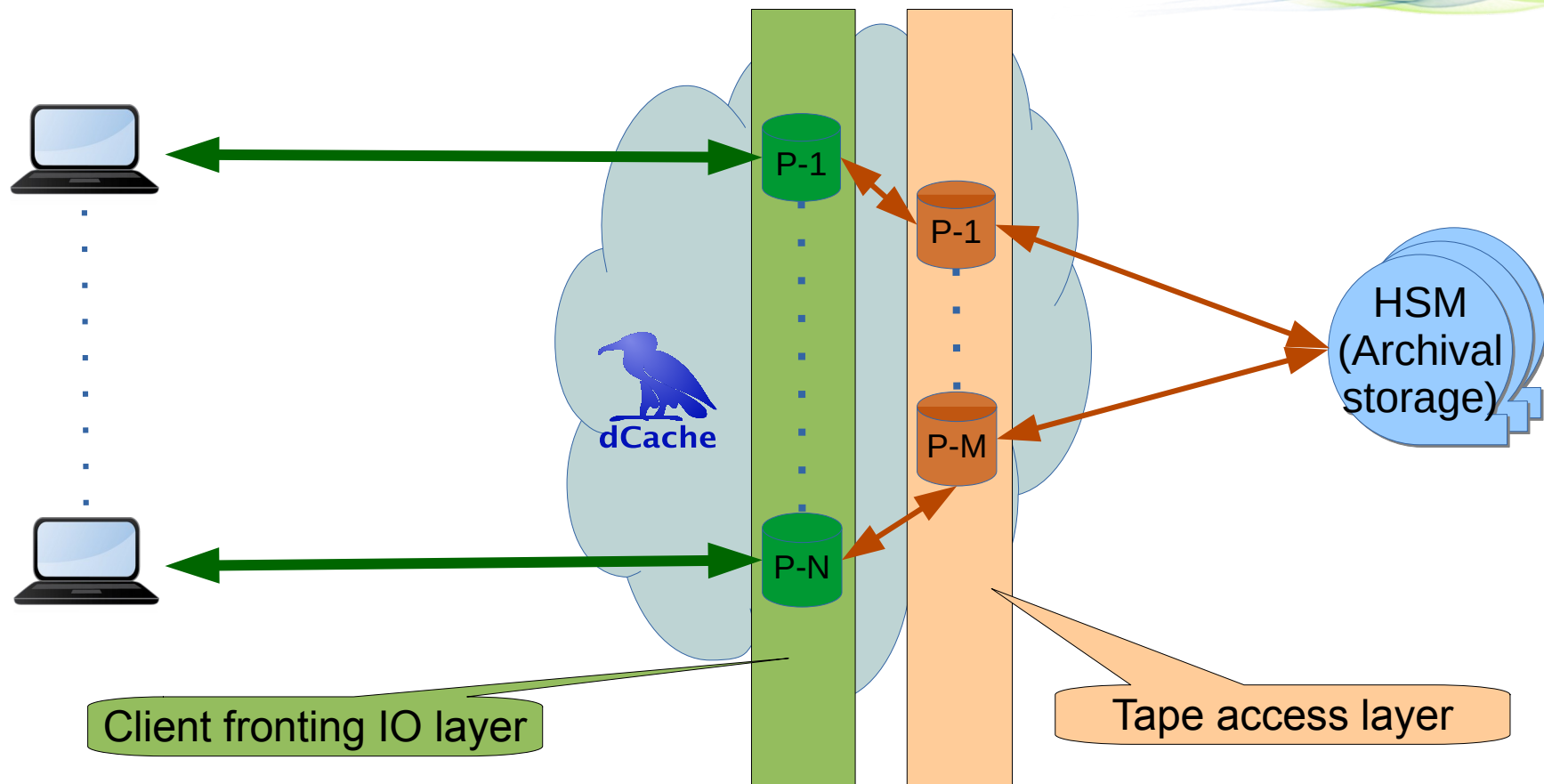
# Typical Tape connection



# Dedicated Write Pool



# Layered Model



# Tape Access Improvements

- Group requests by tapes (on pools)
  - Requests to single tape submitted together
- Group tapes by pools (on PinManager)
  - Access to single tape sent to a single pool
    - Requests from multiple pools to single tape coordinated
  - Some sites `fixed` by using single stage pool
- Multi-layer tape access model (internal QoS)

- More dimensions than latency and price
  - Durability and availability
  - Technology independent
- Not a static parameter
  - Depends on data life-cycle
- Not SRM specific

# QoS in dCache

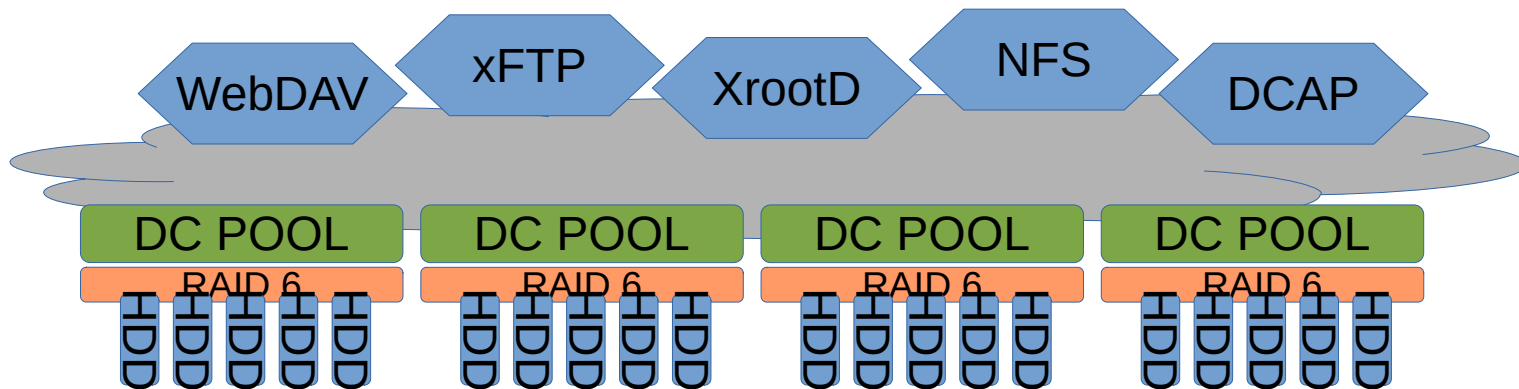
- Disk/Tape migration
  - Automatic, user policy based
  - Manual, user driven
- nDisk replicas
  - Automatic, system policy based



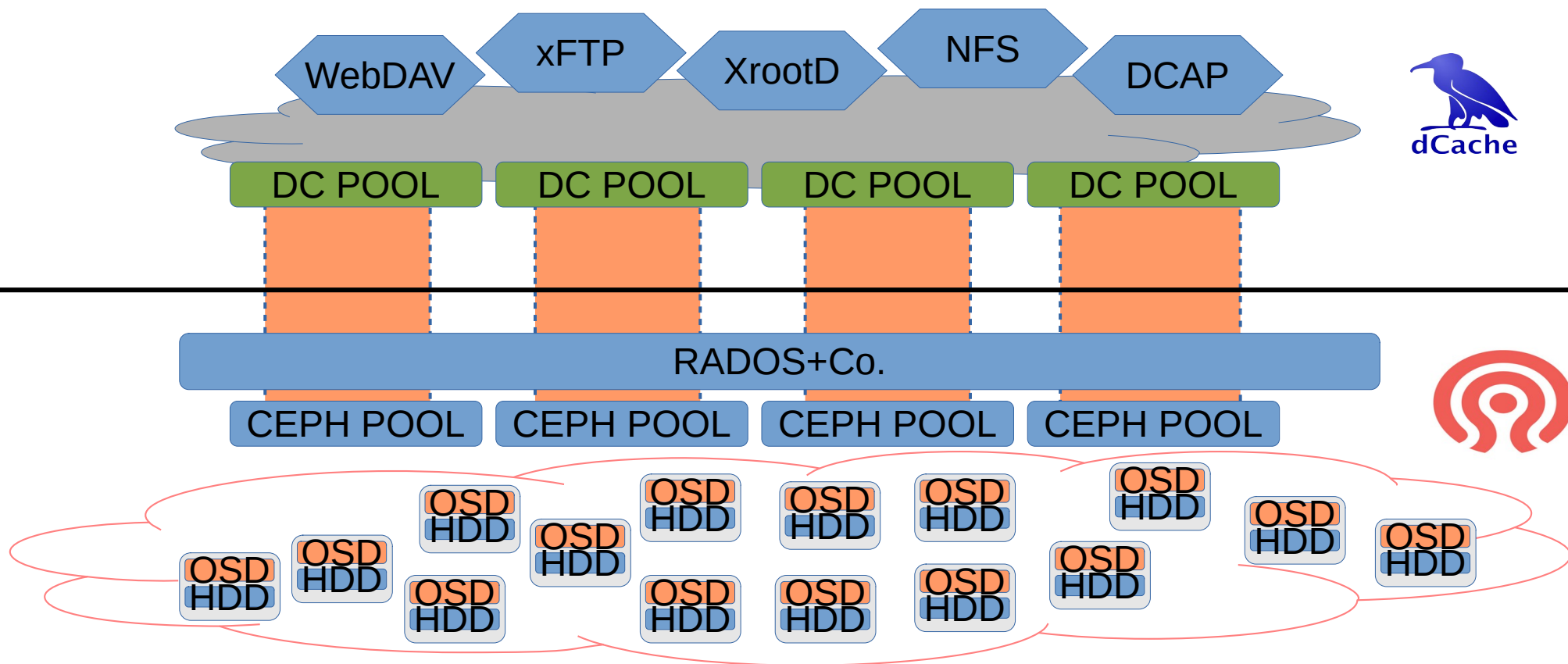
# QoS Changes

- New role for resilience
  - QoS capabilities
  - Tape aware
- Bulk operations in REST API
  - Like SRM, but different
  - QoS transitions directory/storage group based
- Multi-tiering
  - based on hardware capability

# Disk Distribution (classic)



# Disk Distribution (CEHP)



# Datalake *(Multi-Site deployment)*

- dCache designed to federate storage
- Flexible data placement/migration policy
  - On Demand, when requested from remote site
  - Permanent, data protection, location adjustment
  - Manual, for data location optimization, maintenance
- Works for all protocols
- Support HSM connectivity
  - Each site/pool may have it's own tape system
- In-transit data protection

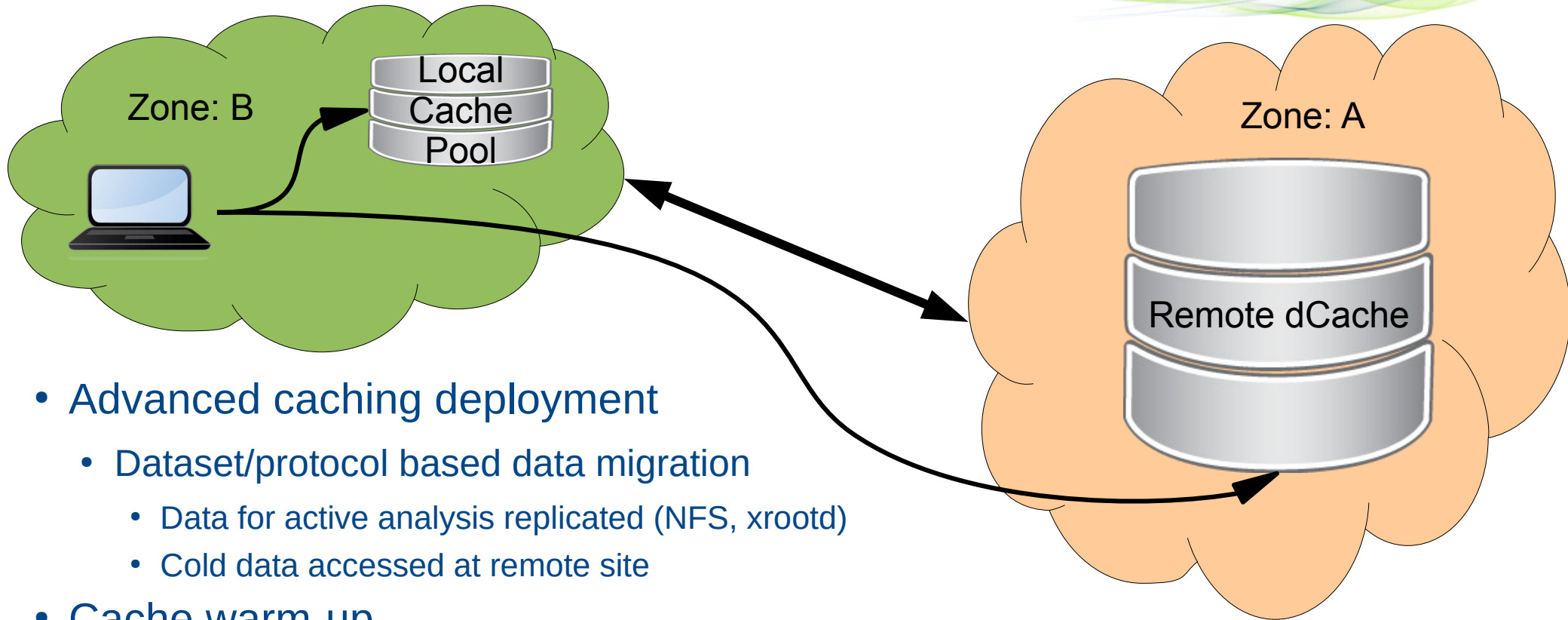
# Zones: Geo-Location

- Geo-location aware unit
- Dynamically groups services together
- Available in replication rules
- Network topology aware internal communication
  - Always prefer local resources
  - Disconnected operation

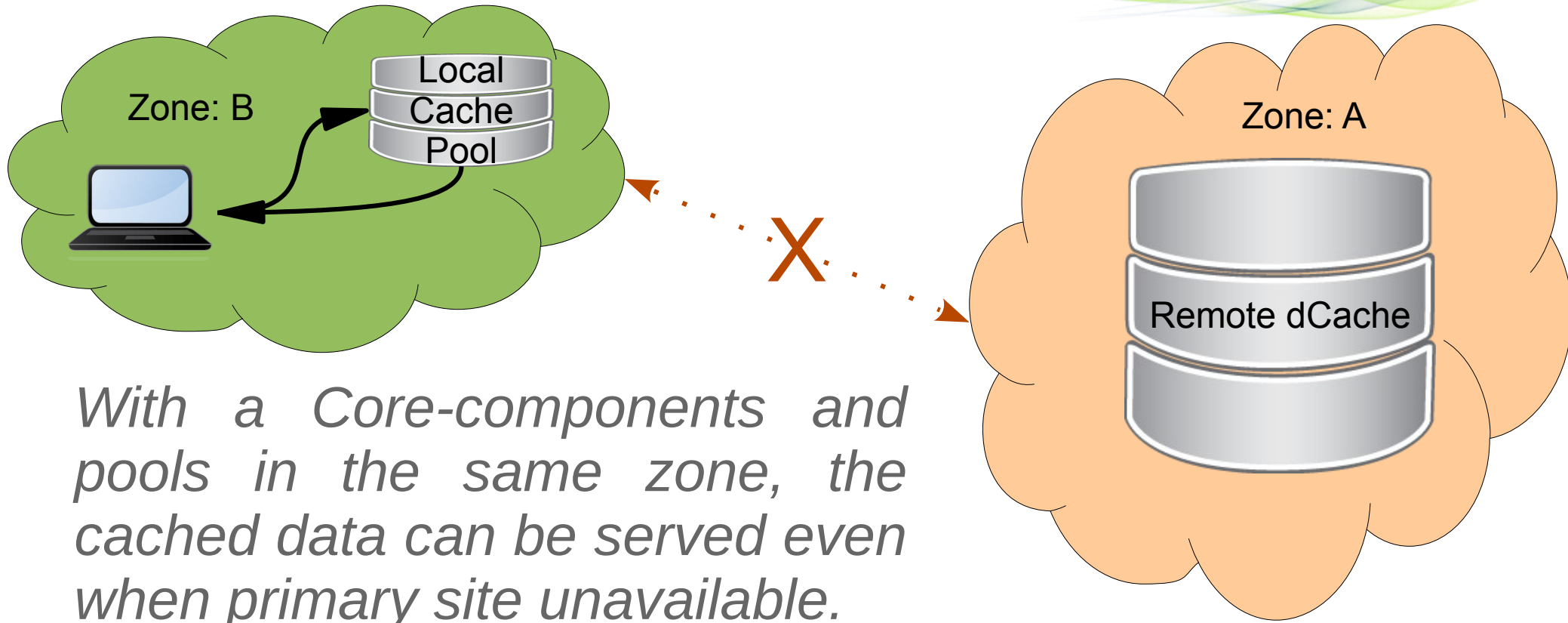
```
set storage unit data:resilient@osm -required=2 -onlyOneCopyPer=zone
```

```
create pgroup caching-pools -dynamic -tags=zone=B
```

# Caching/Cloud Bursting



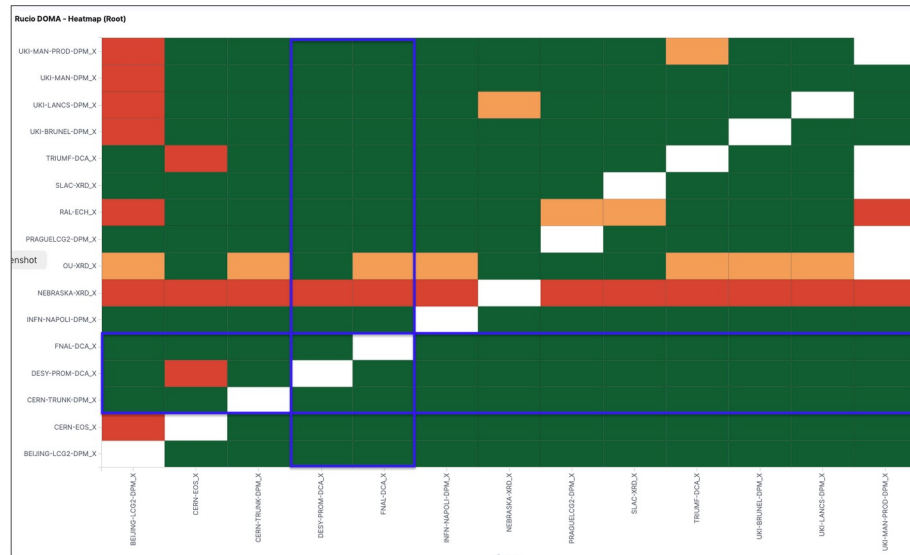
- Advanced caching deployment
  - Dataset/protocol based data migration
    - Data for active analysis replicated (NFS, xrootd)
    - Cold data accessed at remote site
- Cache warm-up





# 3<sup>rd</sup> Party Copy

- XROOTD
  - Source/destination support
  - GSI + delegation, SciToken
  - Inter-op with SLAC xrootd client & server (DPM, EOS)
  - TLS support (dcache-6.2)
- HTTP
  - Source/destination support
  - 3<sup>rd</sup> vendor HTTP server as destination
  - X509, Macaroon and SciToken support
- dCache 5.2.x is the LTS version with all required changes
  - recommended version by DOMA-TPC WG




- dCache as stored storage in kubernetes cluster
  - successful demo by PaNOSC project
- REST API for dynamic access provisioning
- Federated AAI
  - OIDC
  - OAuth2

```
extraVolumes:  
  - name: dcache  
    nfs:  
      server: dcache-door-doma01.desy.de  
      path: /
```

- Improve IO latency
  - New message serialization
- Optimize POSIX (NFSv4.1) interface
  - Continues compatibility tests
  - High-performance byte-range-locking
  - Active collaboration with NFS client development

# Support/Contact Channels

- **support  dcache.org**
  - User request tracking system
  - Place to ask for a help from developers
  - Accessible by all team member
- **security  dcache.org**
  - Request tracking system
  - To report security issues or incidents
  - Restricted access
- **user-forum  dcache.org**
  - Mailing list for sysadmins self-help group
  - To ask for an advice or share experience
  - Used by (almost) all sysadmins and developers
- **dev  dcache.org**
  - Shared mailbox
  - An email to contact developers. Not for support
  - Developers can send e-mail from this address
- **srm-deployment  dcache.org**
  - Tier – 1 coordination mailing list
- **meet.desy.de/xxx**
  - Weekly Tier 1 support video call
- **workshop  dcache.org**
  - Shared mailbox
  - An e-mail used to organize workshops
- **Github issues**
  - Request tracking system
  - To report software defects and feature requests
  - Public
- **Github pull-requests**
  - Request tracking system
  - To provide code changes
  - Public

# Questions?

