dCache.org

dCache in a nutshell

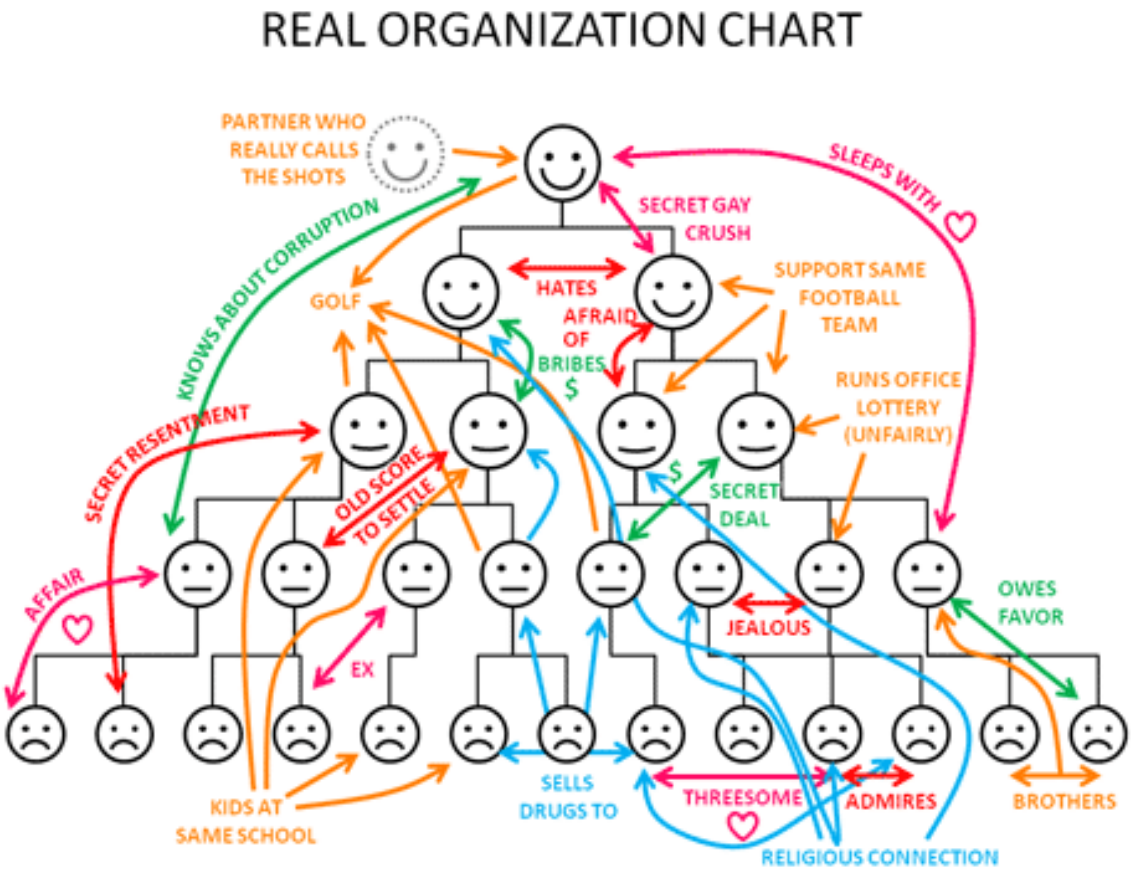
Patrick Fuhrmann

On behave of the project team

INDIGO DataCloud

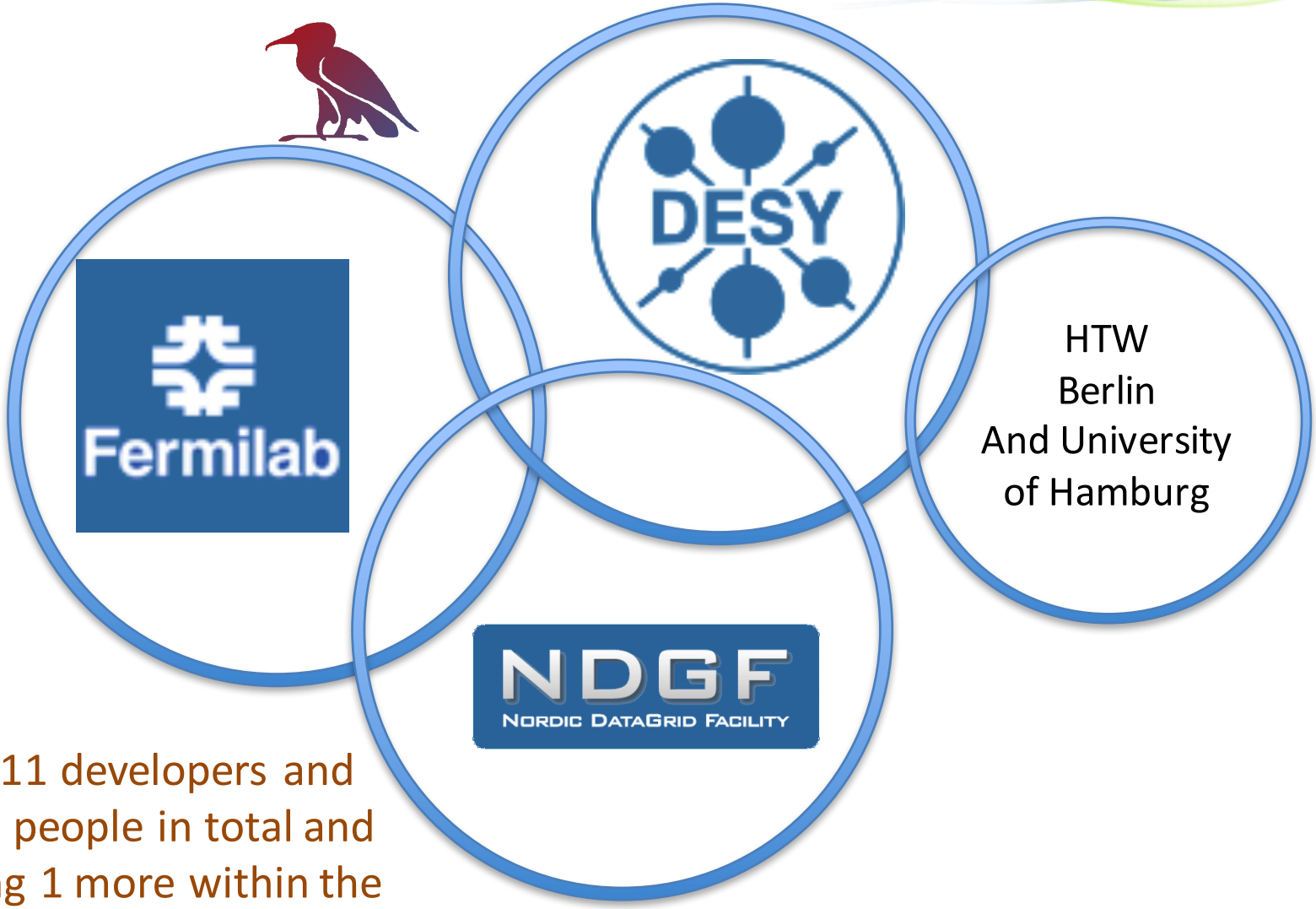
Fermilab NDGF Nordic DataGrid Facility DESY Helmholtz Association Federal Ministry of Education and Research LSDMA

What's dCache.org, how are we organized ?



What's dCache.org

dCache.org



About 11 developers and support people in total and expecting 1 more within the next 3 months.

dCache.org networking

dCache.org

OSG
Open Science Grid (US)

EGI
European Grid Infrastructure

DCORE
Swiss Company

INDIGO-DataCloud
European Grid Infrastructure

NeiC
Nordic e-Infrastructure Collaboration

RDA
Research Data Alliance

WLCG
World Wide LHC Computing Group

LSDMA
Large Scale Data Management And Analysis

Funding and Objectives

dCache.org

LSDMA

2010 — 2013 — 2015 — 2018

Standardization

NFS 4.1 / pNFS

HTTP / WebDAV

Contributing to the Dynamic Federation

Deploying new technologies into Production and exploring new communities

INDIGO DataCloud

Data Life Cycle

Multi Tier Storage
Quality of Service
Migration Archiving

AAI → *AARC*

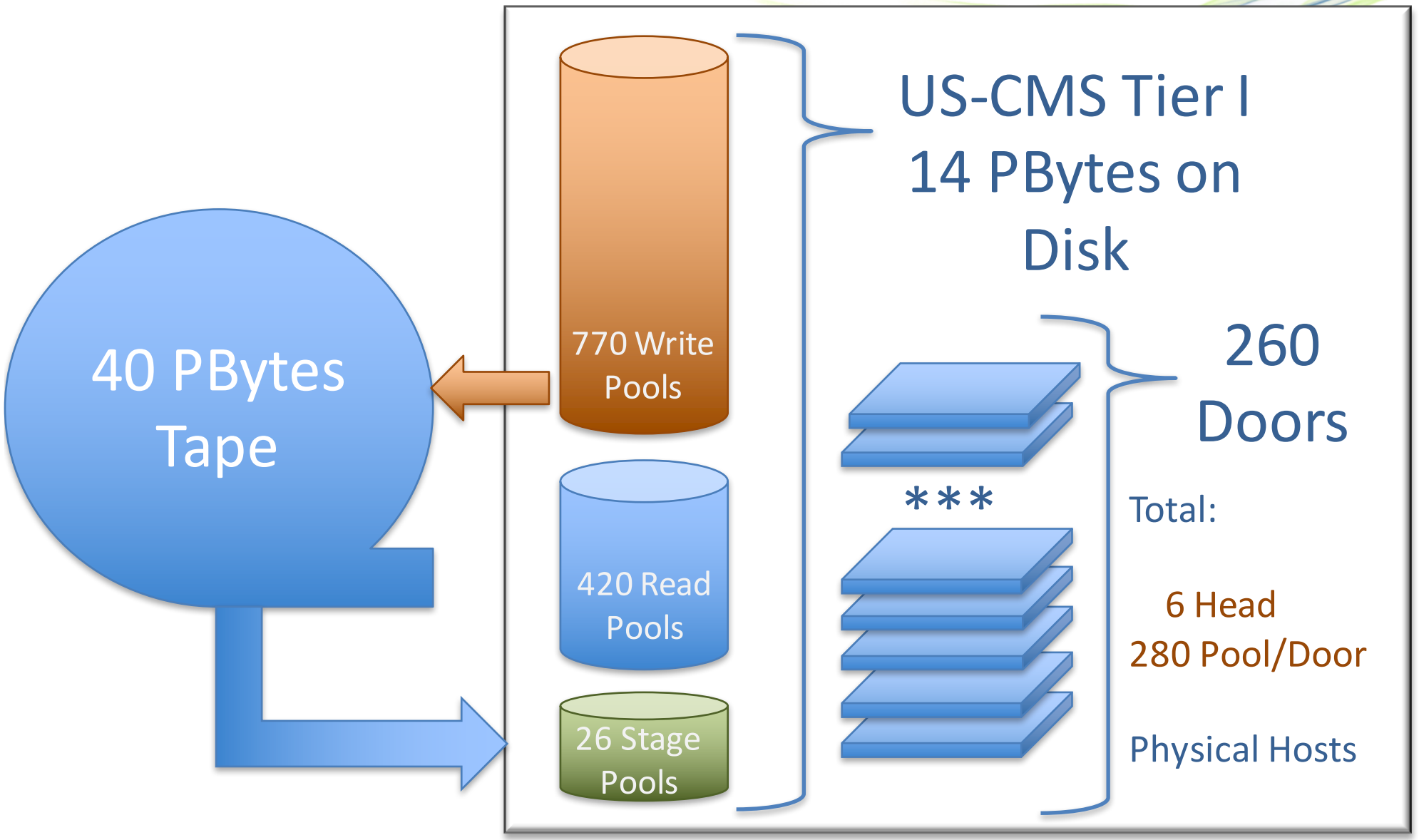
Improve Interoperability of R&E AAI

Who/Where are our customers (users)

Worldwide distribution

Starting with possibly the biggest

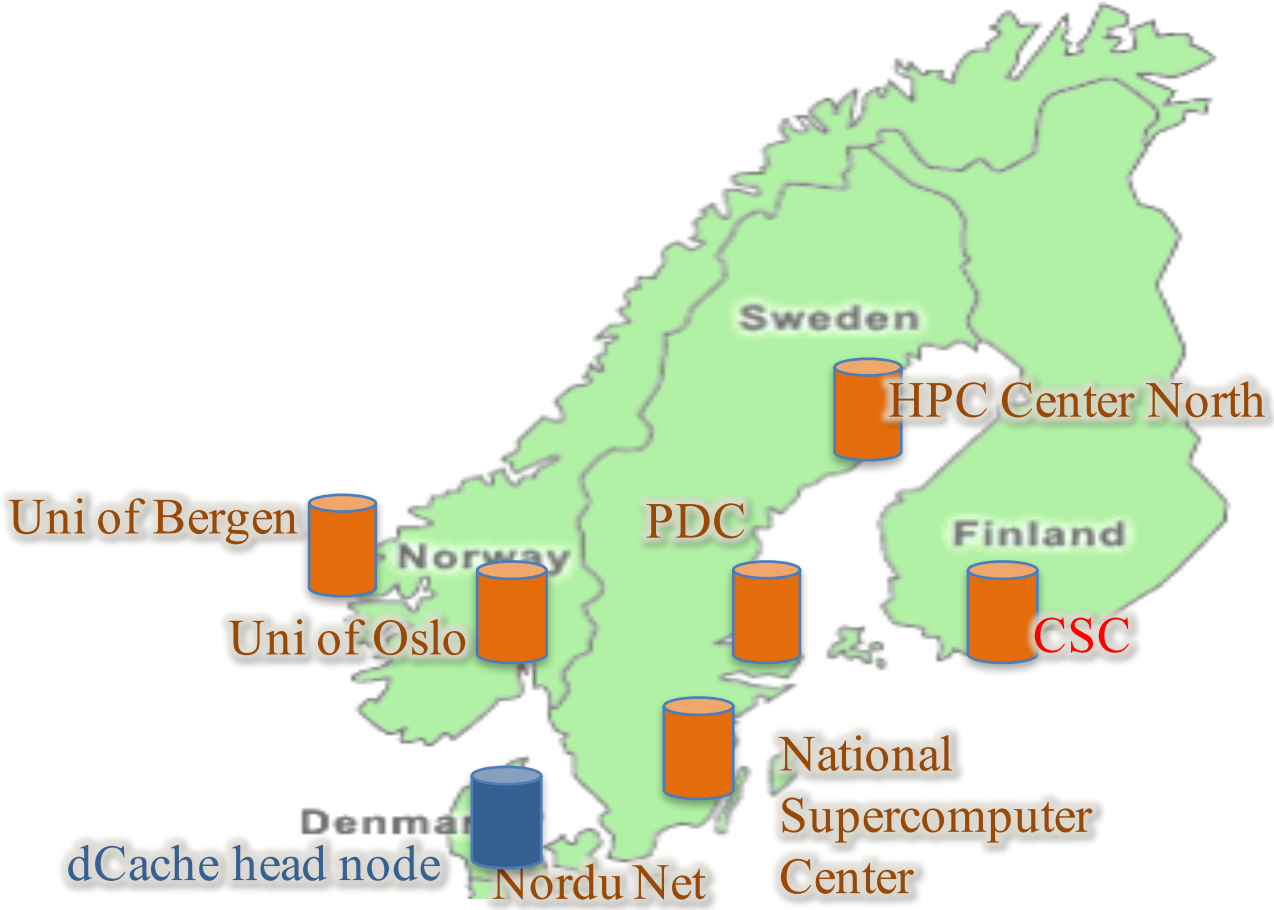
dCache.org



Information provided by Catalin Dumitrescu and Dmitry Litvintsev

To certainly the most widespread

dCache.org



4 Countries

One dCache

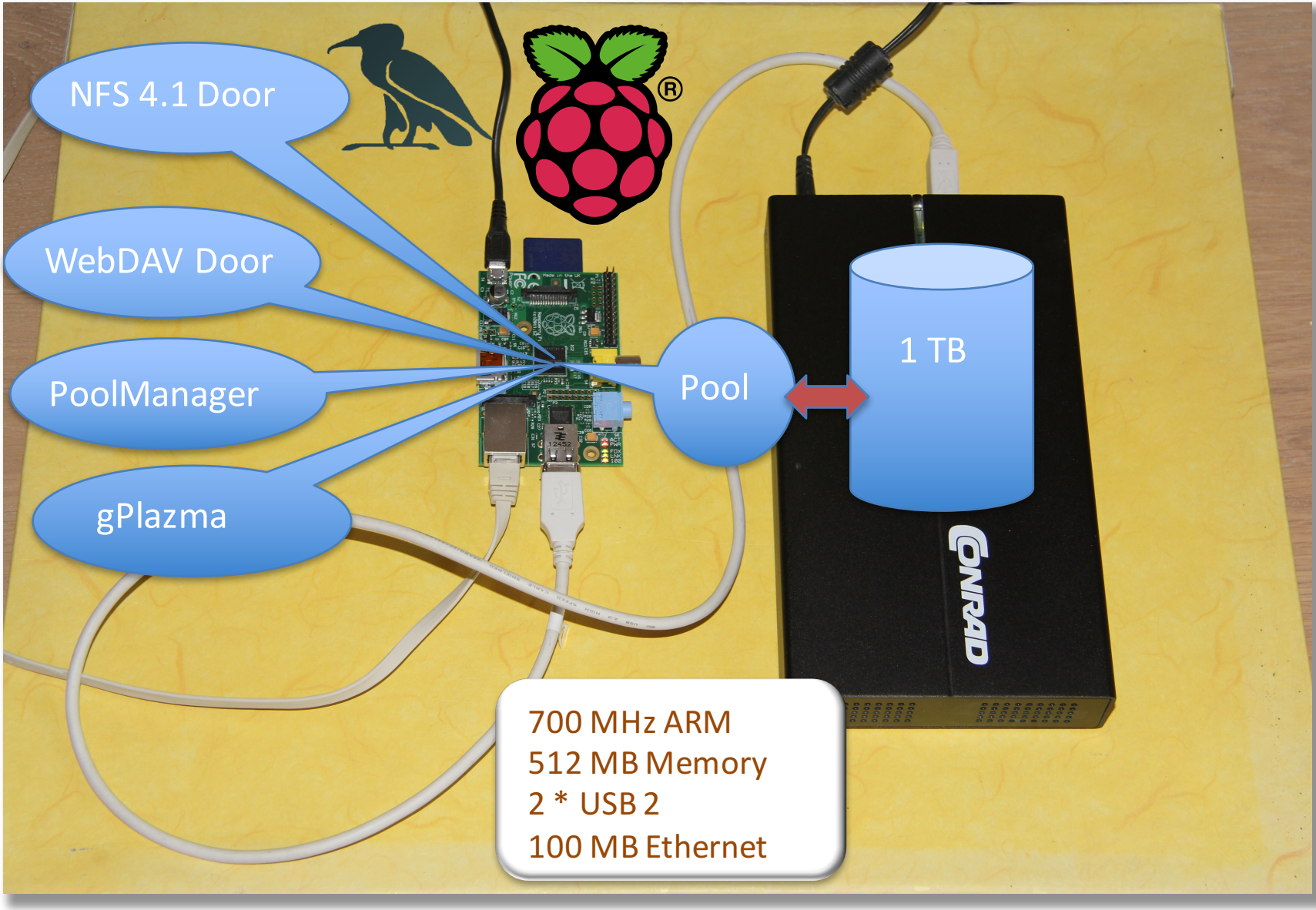
Slowenia

Slide stolen from Mattias Wadenstein, NDGF

To very likely the smallest

One Machine – One Process

dCache.org



Communities

- Biggest : WLCG
- Photon Science (here at DESY)
- LOFAR: Amsterdam, Juelich (Poland im prep.)
- Intensity Frontier (Chicago): Neutrino and Myon
- Others
  - Federated dCache
    - Uni Michigan
    - CESNET
  - People often use dCache for non WLCG science, as they already have an WLCG dCache at home.

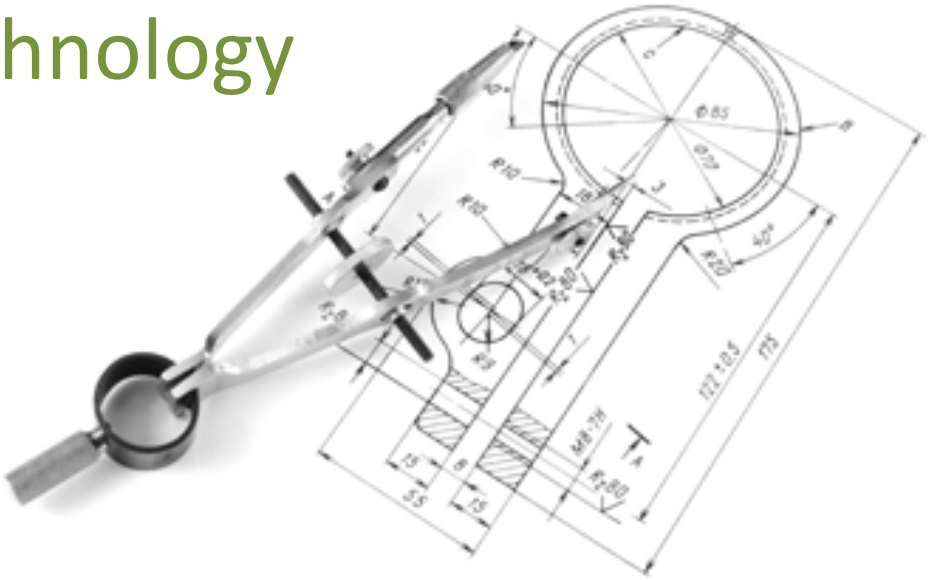
Now … what's a dCache

dCache

dCache Cheat - sheet

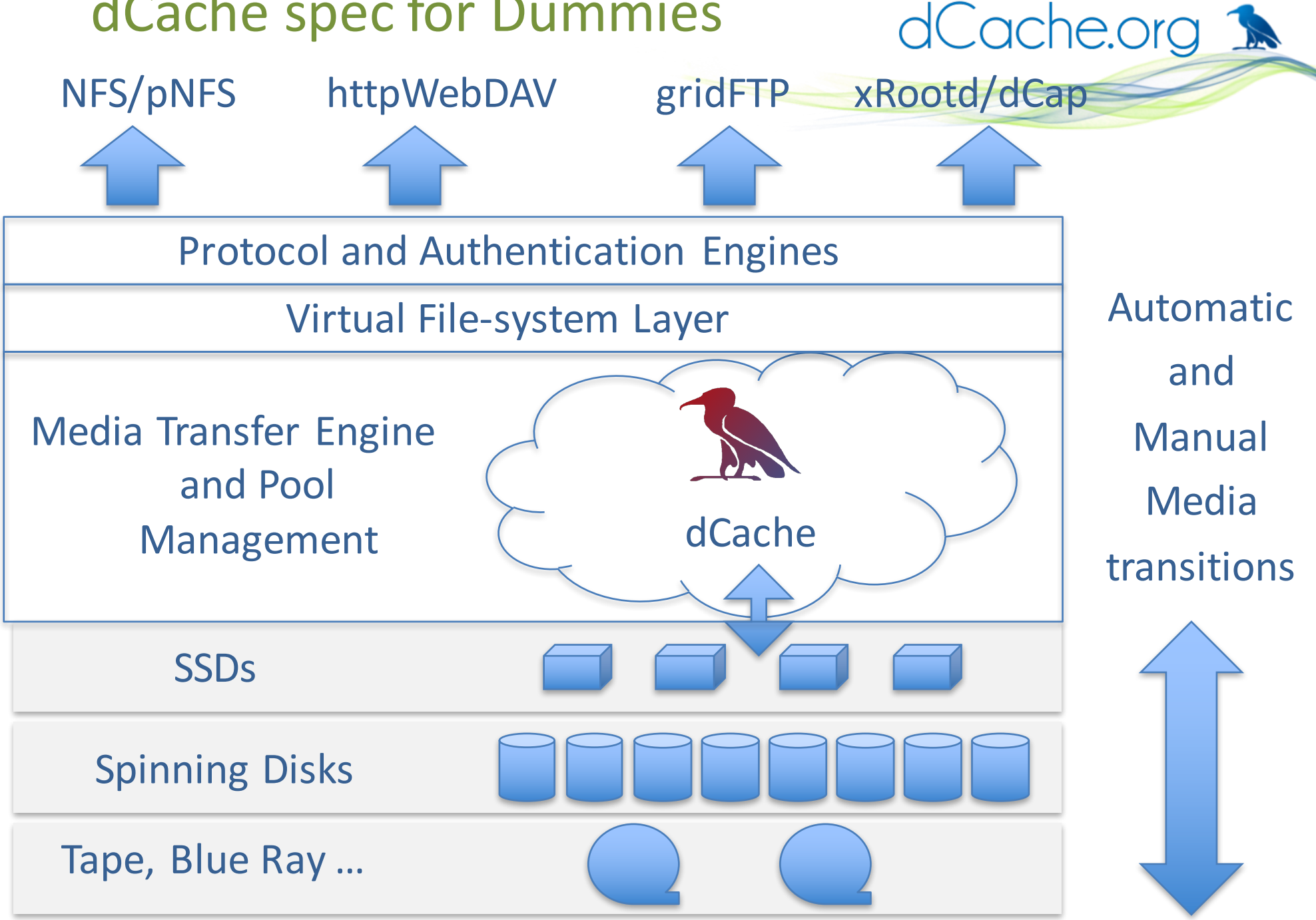
- dCache is a horizontally scaling storage management technolgy.
- Exposes its file system via
  - NFS
  - GridFTP
  - http(WebDAV)
- Provides a variaty of authentication mechanisms
- Fine grained Authorization (by POSIC ACL's)
- Allows fine grained "Storage Management"
  - No system interruption or user notice for
  - Moving data around, between storage
  - Adding or decommisioning of storage units
- Supports Tiered Storage (Tape, Disk, SSD) and Software Defined Storage

The Technology



dCache spec for Dummies

dCache.org

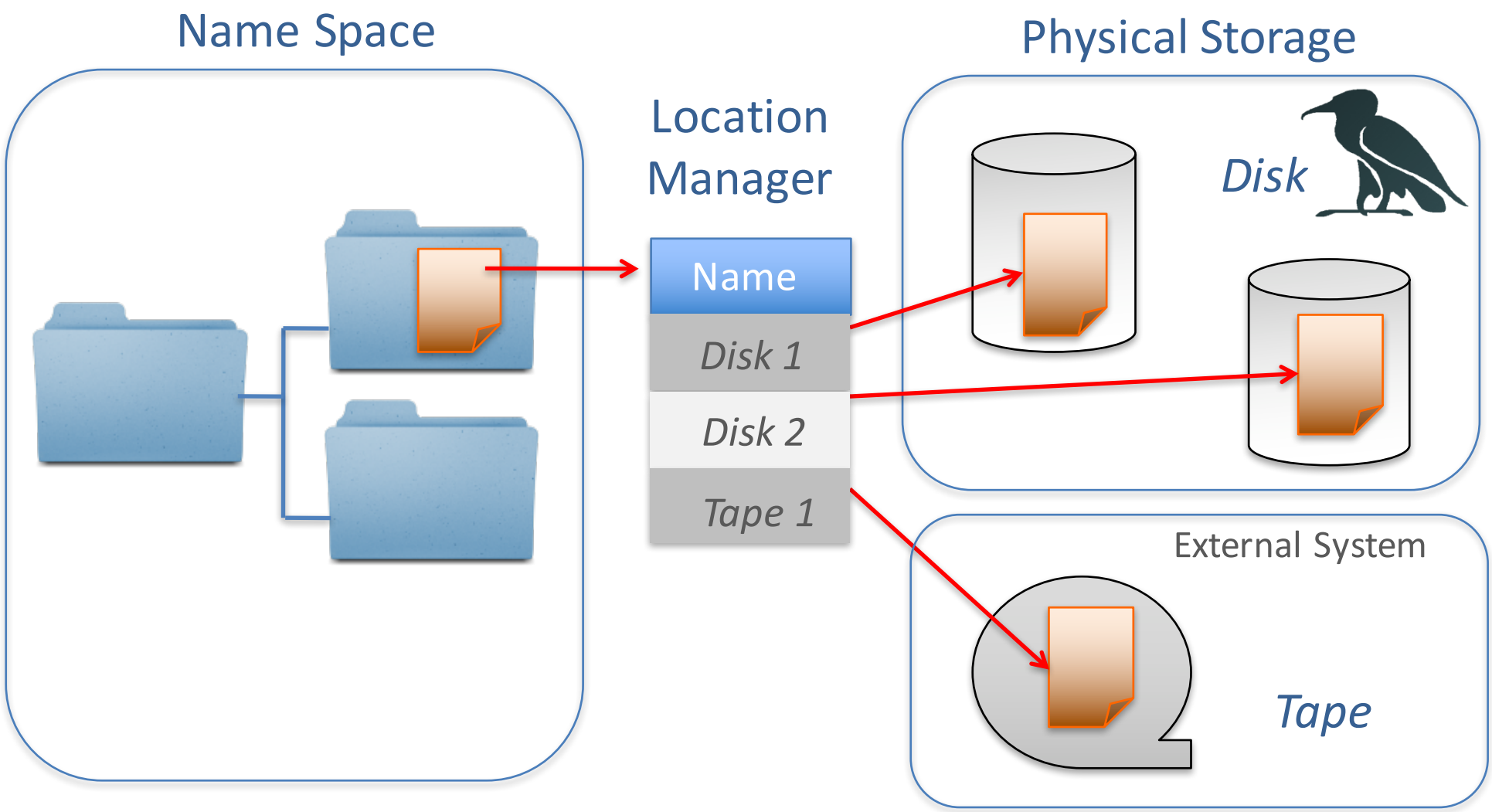


In other words

- Files are stored as objects on various data back-ends (Hardsdisk, SSD, Tape)
- Back-ends can be highly distributed, even beyond countries.
- The File namespace engine is independent of the data storage itself.
- File object location manager keeps track of copies on the various media.

Design

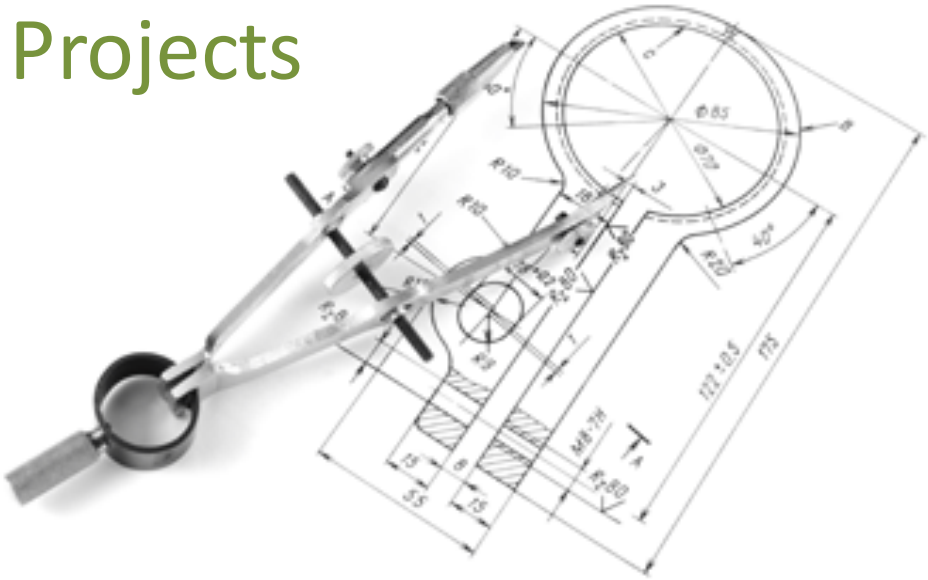
Namespace – Storage separation



Resulting Features

- Hot Spot detection
  - Files are copied from ‘hot’ to ‘cold’ pools
- Multi Media Support
  - File location is based on access profile and storage media type/properties
    - Fast streaming from spinning disks
    - Fast random I/O from SSD’s
- Migration Module(s)
  - Files can be manually/automatically moved or copied between pools.
  - Rebalancing of data after adding new (empty) pools.
  - Decommission pools.
- Resilient Manager
  - Keeps max ‘n’ min ‘m’ copies of a file on different machines.
  - System resilient against pool failures.
- Tertiary System connectivity (Tape systems)
  - Data is automatically migrating to tape.
  - Data is restored from tape if no longer on disk

Current Projects



Current Projects

- Sync'n Share
  - Currently OwnCloud dCache Hybrid System
- Getting "Open ID Connected" integrated
- **Allowing fine grained control on**
  - **Storage Quality "QoS"**
    - **Lettinging end user pick storage quality and price**
  - **Data Life Cycle**
    - **Location of data over time (policy driven)**
      - **Starting with SSD**
      - **Ending up on tape**
- **Building dCache Federations**
- Industry
  - Building highly secure Cloud System (confidential)
  - dCache In a Box (dCache Appliance) with DDN

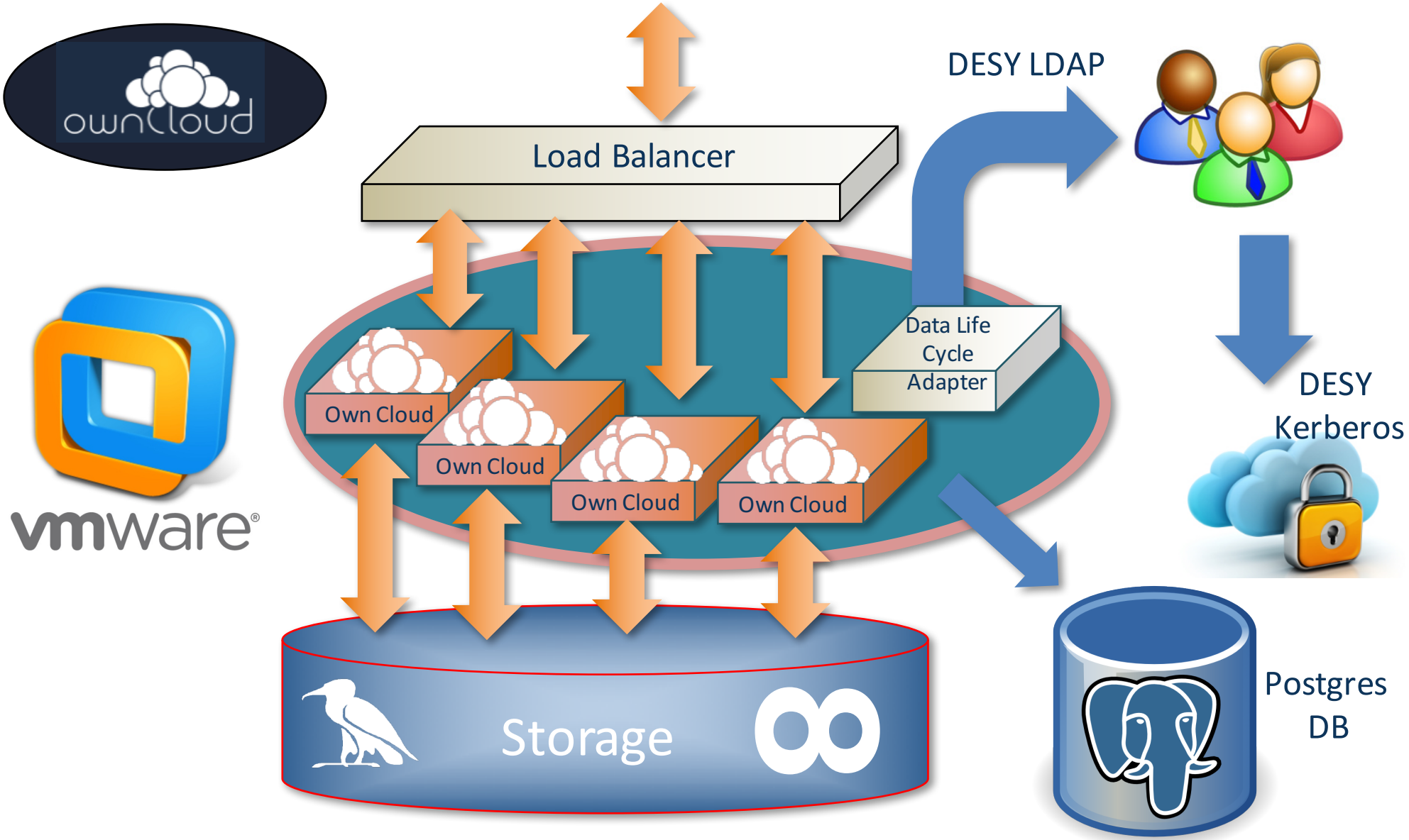
New Model in Accessing Data

dCache.org

- Anytime from everywhere
- From mobile devices
- Bidirectional sync'ing between your cloud space and your local devices
- Getting access to your "Cloud Storage" via high speed, low latency or Wide Area Transport protocols (NFS, GridFTP)
- Allow Customer to define the quality of storage (and price) w/o getting a sys-admin on the phone, eg
  - Access Latency (SSD, ONLINE, OFFLINE, …)
  - Retention Policy and Data Life Cycle Policy

The Own Cloud Part

dCache.org



‘HOME’ from user perspective

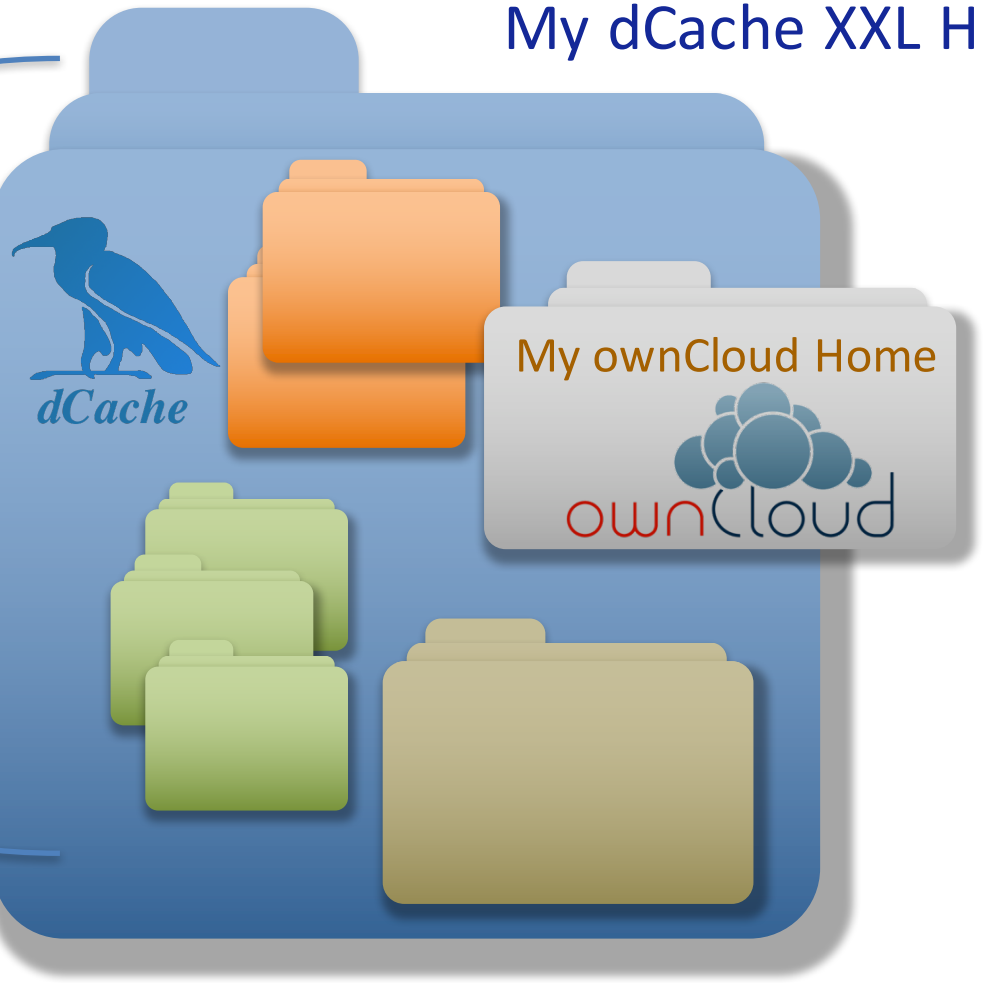
dCache.org

Quality of Service

Tape
Low Latency
Archive
Scratch

Multi Protocol

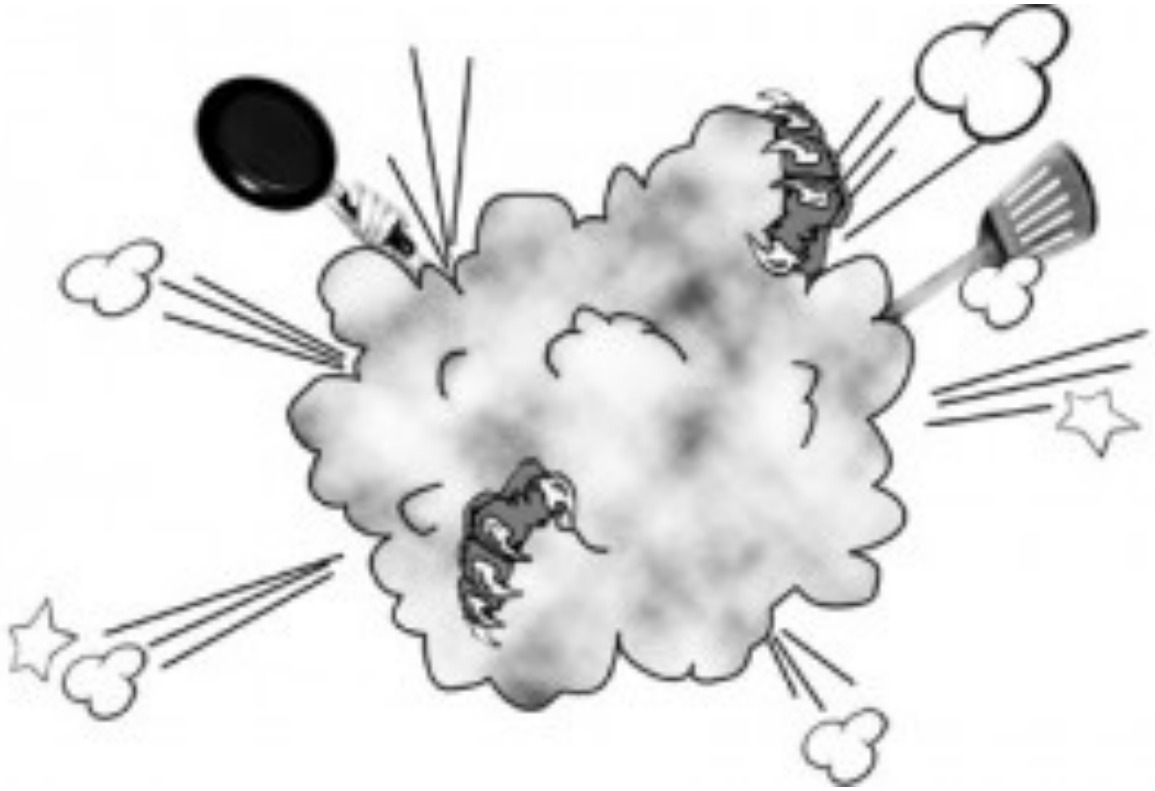
NFS 4.1/pNFS
GridFTP
WebDAV
SRM



Sync
Share
Web 2.0

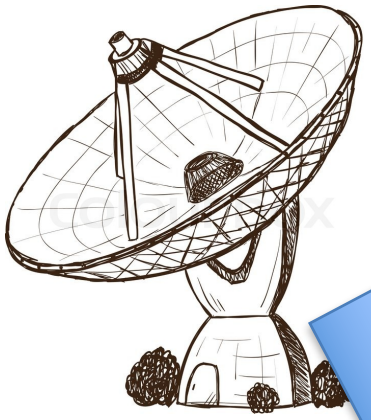
Final Goal

Scientific Storage Cloud



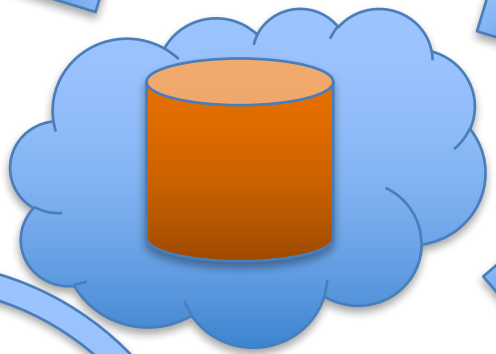
Scientific Data Flow

dCache.org

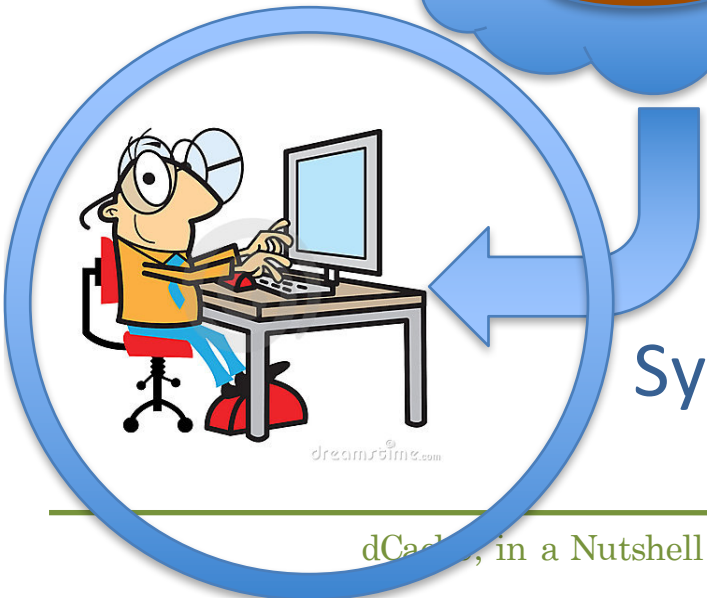


High Speed
Data Ingest

Fast Analysis
NFS 4.1/pNFS



Wide Area Transfers
(Globus Online, FTS)
by GridFTP



Sync'ing and Sharing with OwnCloud

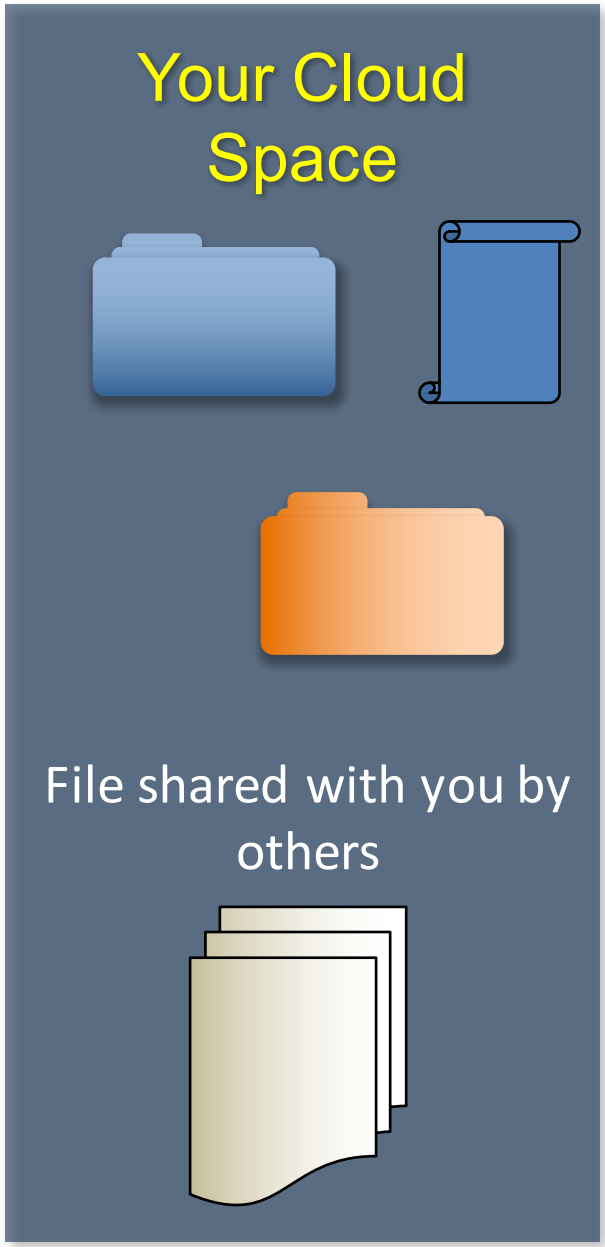
Enough for today ….

Sharing requirements from DESY users

dCache.org

- Fine grained sharing with individuals and groups.
- Sharing via intuitive Web 2.0 mechanisms (Apps or Browser)
- Sharing with ‘public’ with or w/o password protection
- Sharing of free space (upload)
- Expiration of shares

And the sharingj part



Share files/folders with individuals

Share files/folders with ‘desy groups’

Share with ‘public’ with and w/o password

(Shares can expire)

Share space(s) with others for upload

Others sharing data with you (in your home)

Why not using

?

- Because there was this gentleman who decided to leave the US towards Moscow, with a bunch of documents, changing our attitude towards foreign storage services significantly.
- The DESY directorate essentially disallowed storing DESY documents outside of DESY premises.

Evaluation of possible products

dCache.org

PowerFolder

ownCloud

CUBEPAD

- Highly secure group-ware system
- Allows sharing encrypted data

ETC.

Product evaluation (cont.)

dCache.org

We went for Own Cloud

- Open Source plus Enterprise version
- Most popular solution:
  - Reduces likelihood for ‘product disappearing’
  - Possibly building a user-community
    - TU-Berlin, FZ-Jülich, TU-Dresden ****
    - CERN, United Nations
- CERN is evaluating a similar approach and we are in contact anyway (WLCG)

Inevitable RP activities

- Collaboration with HTW Berlin (LSDMA)
- Pre-evaluation of cloud solutions by “InFa” -> Q3/2013
  - Erarbeiten und Umsetzen eines firmeninternen Online-Speicherdienstes in einer Teststellung. (Quirin Buchholz)
- Presenting the concept at HEPIX.
- Information exchange with CERN. (CHEP’13) Oct 13
- Berlin Cloud Event, (mostly OwnCloud and PowerFolder) in Mai 14 (we published first paper)
- Participating the CERN Cloud Event (Nov ‘14) including a presentation of our proposed solution.
- Various papers submitted and accepted at ISGC in Taipei in March and CHEP’15 in Japan.

However, as we do scientific computing and to just storing and sharing images, there is more to consider.

More requirements

- Request for *unlimited, indestructible storage*.
- Request for *different quality of services* (SLA), coming with different price tags and controlled by customer.
  - *Data Loss Protection* (non-user introduced), e.g.:
    - One copy.
    - Two copies on independent systems.
    - Two copies in different buildings.
    - Two copies at different sites (e.g. Hamburg and Zeuthen)
    - Some of above plus 'n' tape copies.
  - *Access latency* and max data rate, e.g.:
    - Regular sync and web access.
    - Worker-node access: High throughput
    - Low latency (e.g. on SSD) for HPC.
- User defined *Data Life Cycle*
  - Move data to tape after 'n' months.
  - Remove from random access media after 'm' months.
  - Make public after 'x' month.
  - Remove completely after 'y' months.
- Controlled by Web or API (*Software defined storage*)

And not to forget

- Access to the same data via different transport mechanisms
  - GridFTP for wide area bulk transfers
  - http/WebDAV for Web applications
  - NFS 4.1/pNFS for low latency, high speed access (e.g. HPC)
- Access with different credentials
  - Username / password
  - X509 Certificates
  - SAML (Single Sign On)
  - Kerberos
  - Macaroons

Our solution

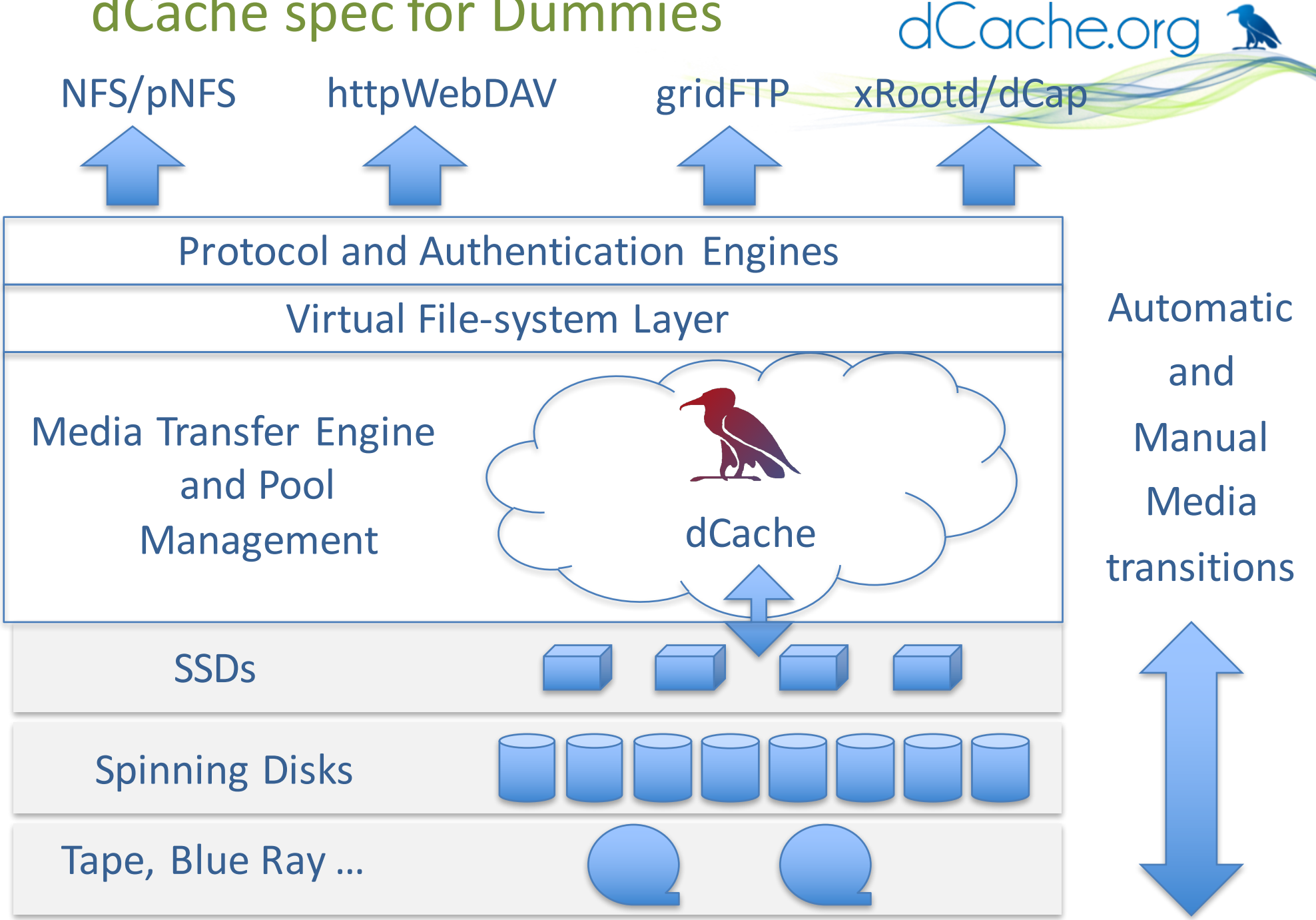
- Non of the Web 2.0 sync and share software products cover the additional requirements.
- So we went for *dCache* as the actually *storage backend*.
- Which is not really a surprise as we are part of the dCache collaboration.

3 slides on dCache.org

Back to technology

dCache spec for Dummies

dCache.org

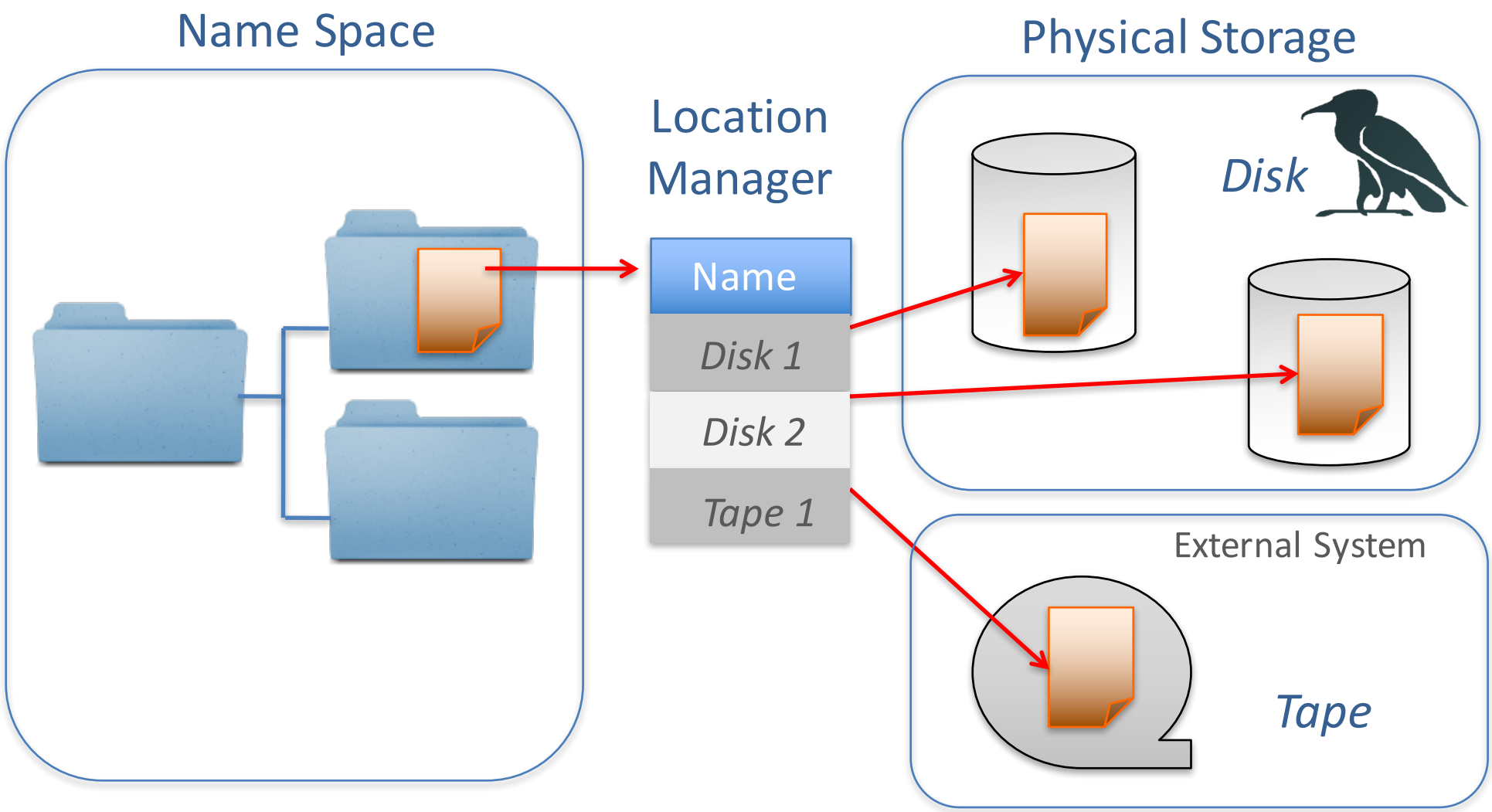


In other words

- Files are stored as objects on various data back-ends (Hardsdisk, SSD, Tape)
- Back-ends can be highly distributed, even beyond countries.
- The File namespace engine is independent of the data storage itself.
- File object location manager keeps track of copies on the various media.

Design

Namespace – Storage separation



Resulting Features

dCache.org

- Hot Spot detection
  - Files are copied from ‘hot’ to ‘cold’ pools
- Multi Media Support
  - File location is based on access profile and storage media type/properties
    - Fast streaming from spinning disks
    - Fast random I/O from SSD’s
- Migration Module(s)
  - Files can be manually/automatically moved or copied between pools.
  - Rebalancing of data after adding new (empty) pools.
  - Decommission pools.
- Resilient Manager
  - Keeps max ‘n’ min ‘m’ copies of a file on different machines.
  - System resilient against pool failures.
- Tertiary System connectivity (Tape systems)
  - Data is automatically migrating to tape.
  - Data is restored from tape if no longer on disk

And what

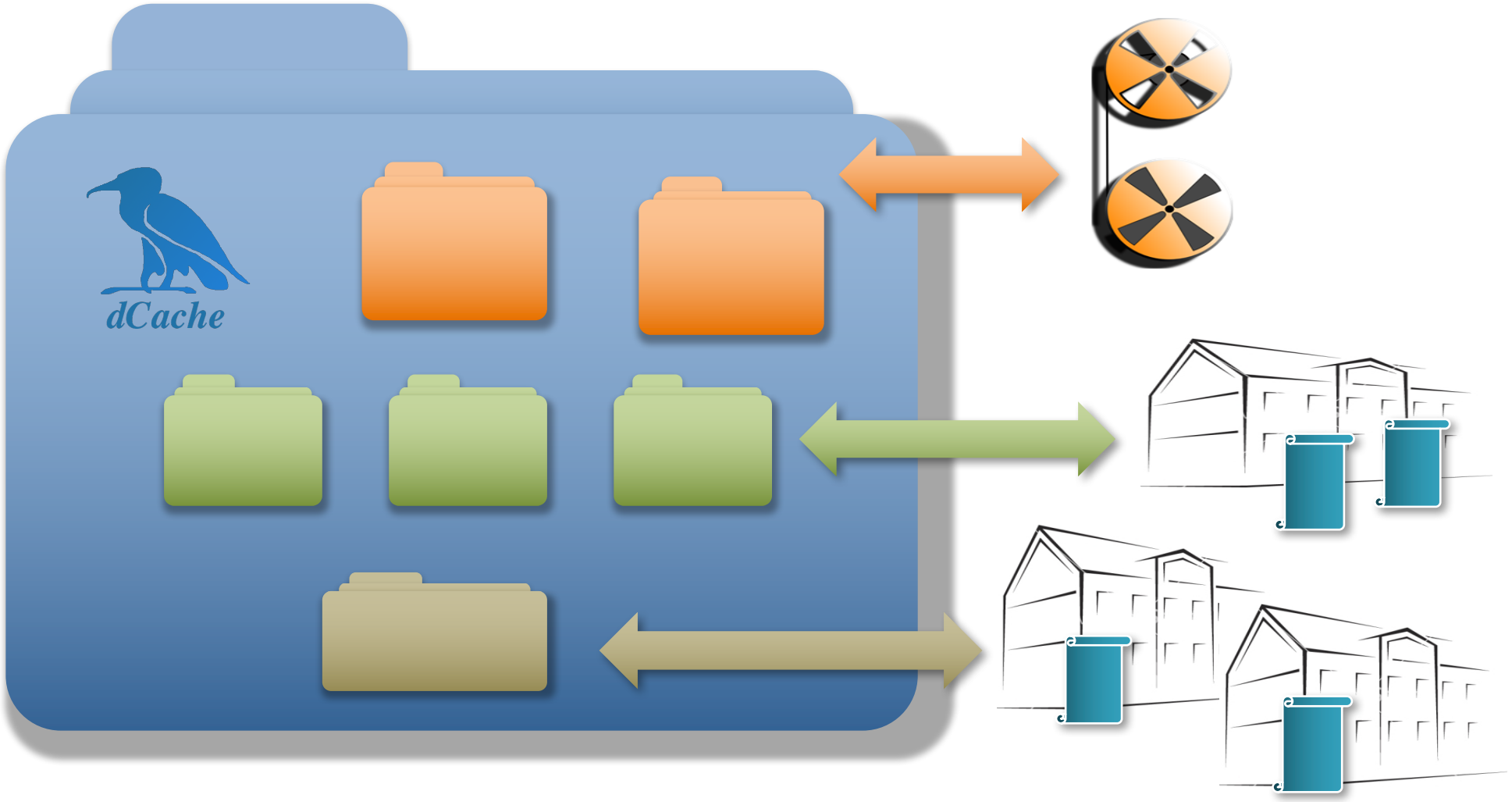
- Why do we need those features ??
- They are the basis for
  - Software defined Storage
  - Quality of Service Management
    - Defining data access latency
    - Defining data retention policies
  - Data Life Cycle support

So, what do we get

- Through Own Cloud
  - Sync'ing
  - Sharing
- Through dCache
  - Multi protocol support
  - Quality of service (Software defined storage)

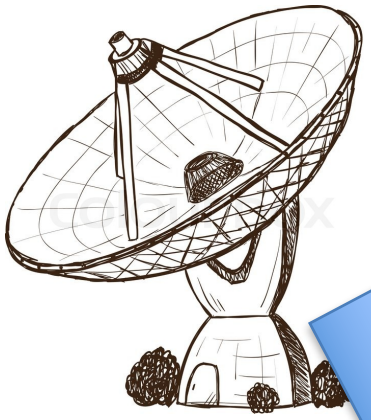
Quality of service

My dCache XXL Home



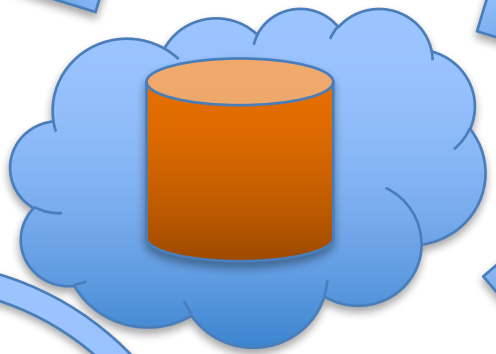
Scientific Data Flow

dCache.org

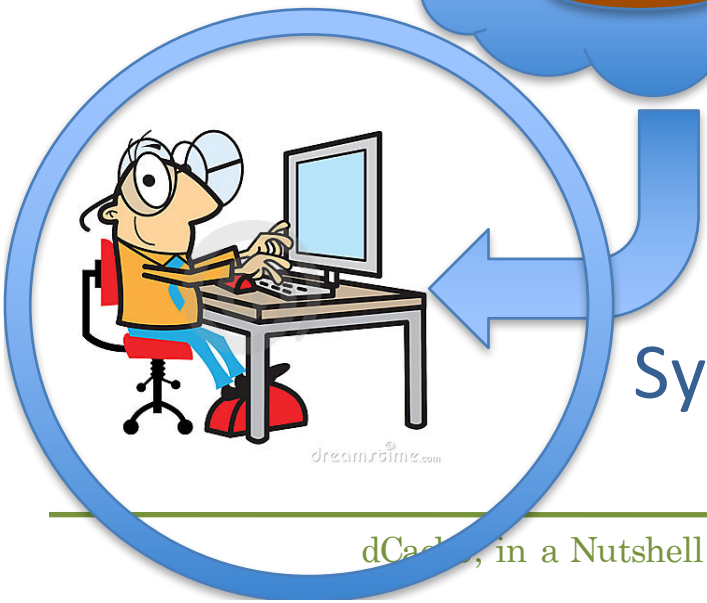


High Speed
Data Ingest

Fast Analysis
NFS 4.1/pNFS



Wide Area Transfers
(Globus Online, FTS)
by GridFTP



Sync'ing and Sharing with OwnCloud

How is that implemented at DESY ?

Integration into the DESY infrastructure

dCache.org

Monitoring

Accounting

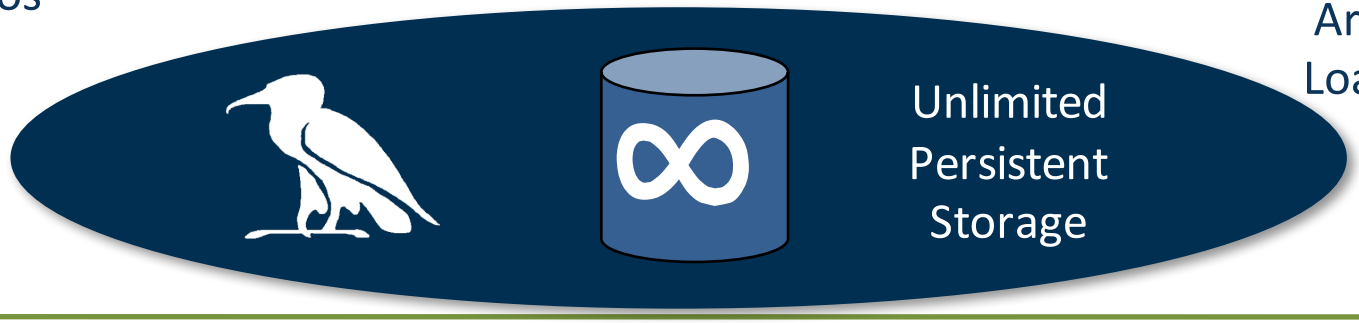
User Management
Registry
LDAP

Virtualization

vmware®

Authentication
Kerberos

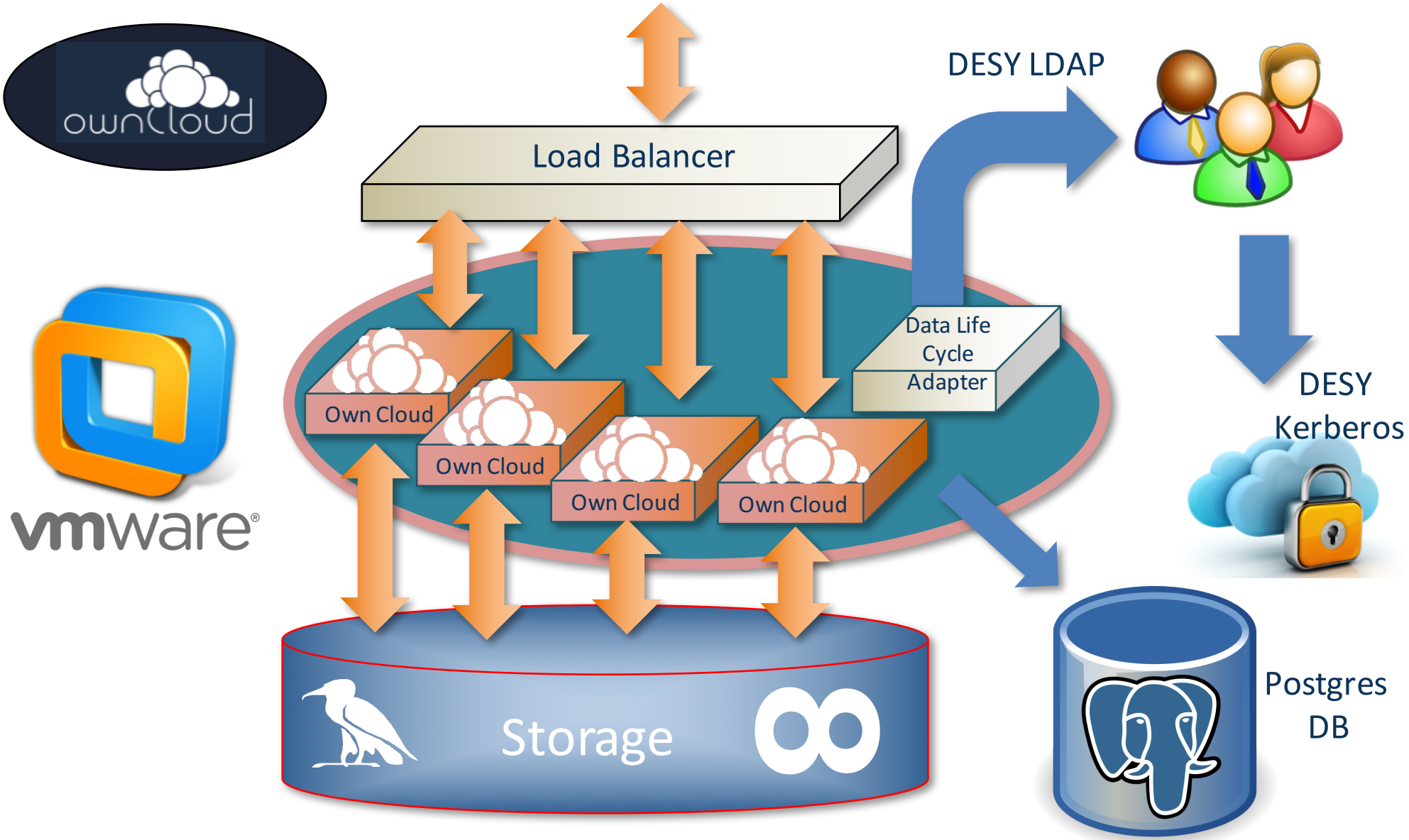
Local and Wide
Area Network
Load Balancing
Firewalls



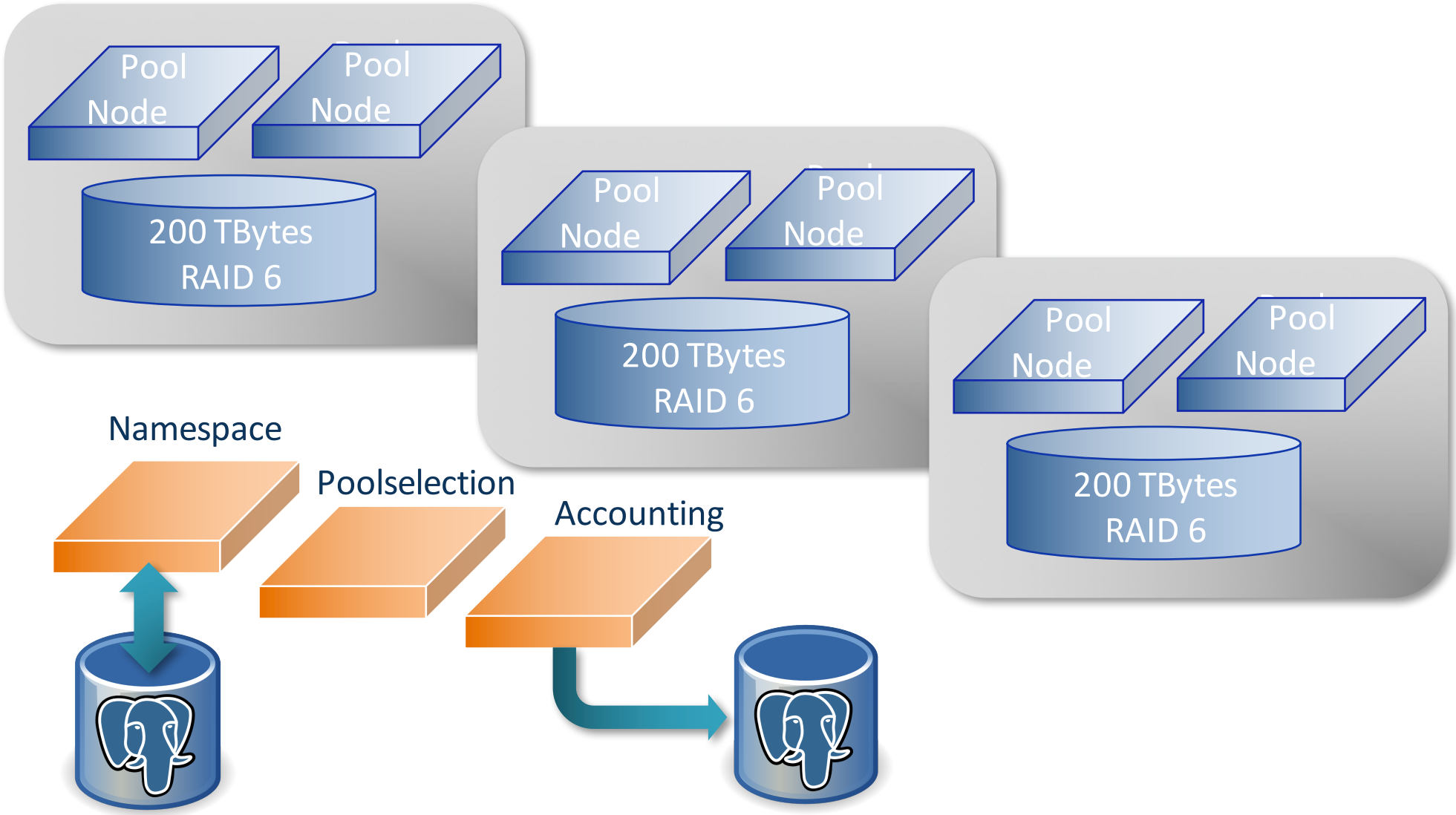
Unlimited
Persistent
Storage

The Own Cloud Part

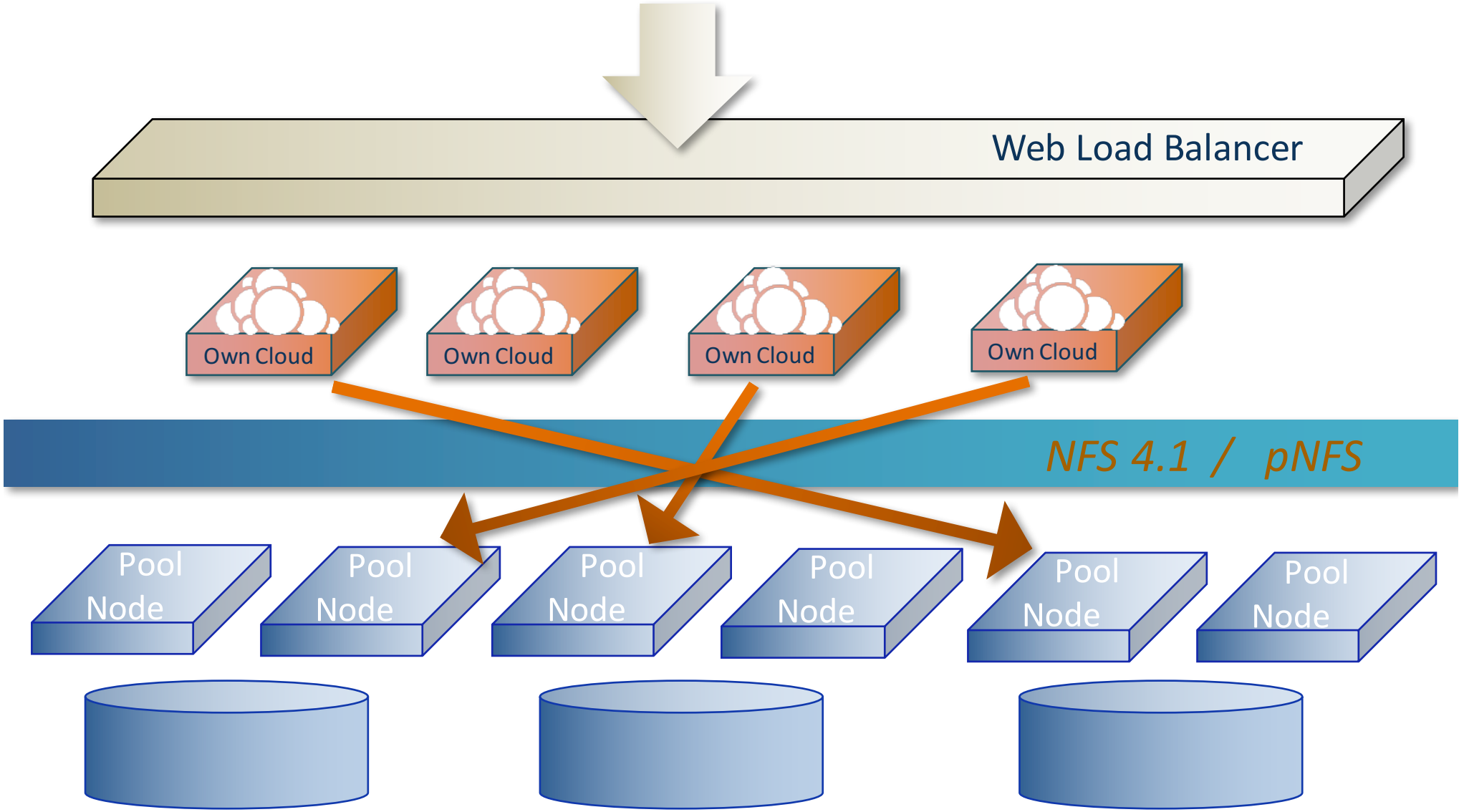
dCache.org



The dCache part



The horizontal scaling



‘HOME’ from user perspective

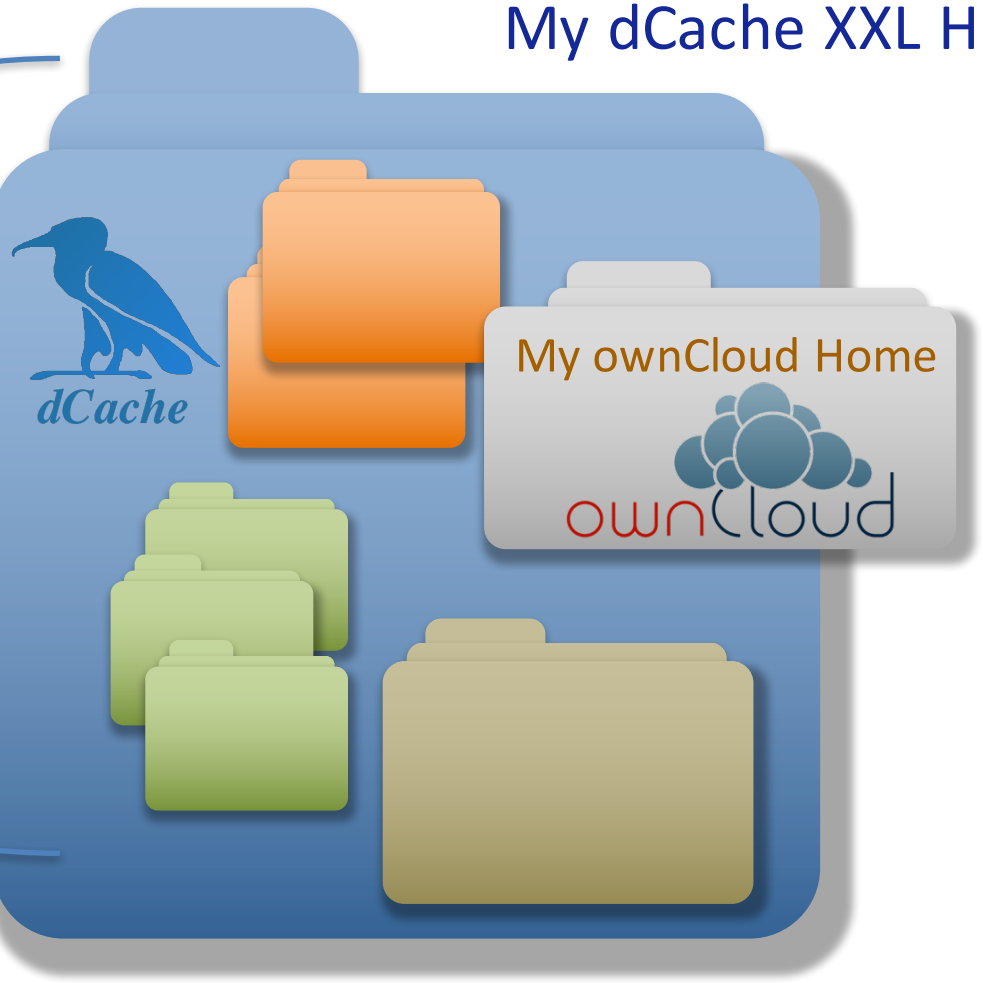
dCache.org

Quality of Service

Tape
Low Latency
Archive
Scratch

Multi Protocol

NFS 4.1/pNFS
GridFTP
WebDAV
SRM



Sync
Share
Web 2.0

Summary

- With dCache and OwnCloud, DESY offers a first prototype of a Scientific Cloud service, providing:
  - User specified Storage Properties (QoS)
    - Access Latency, Retention Policies
  - A variety of access protocols
    - Http/WebDAV, GridFTP, SRM, NFS 4.1 (CDMI)
  - Multiple Authentication mechanism
    - X509 Certificates, Kerberos, User/Password (SAML)
  - Sync and share
  - Web Browser access

The END

further reading

www.dCache.org

Response to ceph

dCache.org

- CEPH complements dCache perfectly.
  - Simplifies operating dCache disks.
  - dCache accesses data as object-store anyway already.
- dCache is evaluating a ‘two step approach’.
  - Each pools sees it own object space in CEPH
  - All pools have access to the entire space, which is a slight change of dCache pool semantics.
- Would merge CEPH and dCache advantages
  - Multi Tier (Tape, Disk, SSD)
  - Multi protocol support for a common namespace.
    - All protocols see the same namespace
  - All the dCache AAI features
    - Support for X509, Kerberos, username/password