



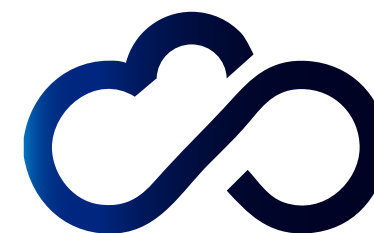
dCache.org 

DESY CLOUD, THE SCIENTIFIC DATA CLOUD

Managed Shared Storage

At the “ownCloud Connects Business” workshop

Dr. Patrick Fuhrmann
Quirin Buchholz
Tigran Mkrtchyan
Peter van der Reest
Lusine Yakovleva



INDIGO DataCloud





Content

- Storage @ DESY ?
- Sync'n Share at DESY
 - Motivation
 - Requirements
 - Implementation
 - Setup
- Requirements from Science Communities.
- dCache for Dummies.
- The ownCloud – dCache Hybrid system
- Summary and outlook.



Storage @ DESY



- HERA [Tier 0] (1992 – 2007)
 - Particle accelerator (Proton – Electron)
 - 6.3 Km (Ring)
 - Some hundred scientists
 - **5 PB in total**
- LCG [WLCG Tier 2] (2008,2009 ...)
 - Particle accelerator (Proton – Proton)
 - 26.7 (Ring)
 - About 10.000 scientist
 - **15 PB / year**

- Petra III [Tier 0] (2012 ...)

- Synchrotron Radiation
- 14 Beamlines
- Beamline Guest Scientists
- **1 PB / year – 5 PB / year**

- European [Tier 0] XFEL (2017 ...)

- 3.4 Km (Linear)
- 2017 (First beamline)
- Beamline Guest Scientists
- **10 – 100 PB / year**

2020
100 PBytes

1992





More storage at DES Y

dCache.org 

- The DES Y data management team has quite some experience in managing huge amounts of data.
- In collaboration with other 'big data' sites, we are providing a data management system 'dCache', deployed at 70 sites around the world.
 - See later.
- **So, why are we running ownCloud ?**



Motivation

dCache.org 

- DESY has no experience in sophisticated data sharing.
 - Data sharing was done in the traditional way with ACL's and 'group' directories
- However : Young scientists start their careers at Universities and Lab's with *Sync'n Share* in their blood. (Drop Box Generation).
- Public IT departments, for a very long time, didn't regard *Sync'n Share* as being their problem as many commercial solutions were around.
- It essentially became an issue after Snowden.
 - Legal Requirement : Data had to be stored 'on site' or at least in Germany
- Consequence : CC needed to provide *Sync'n Share like* mechanisms.



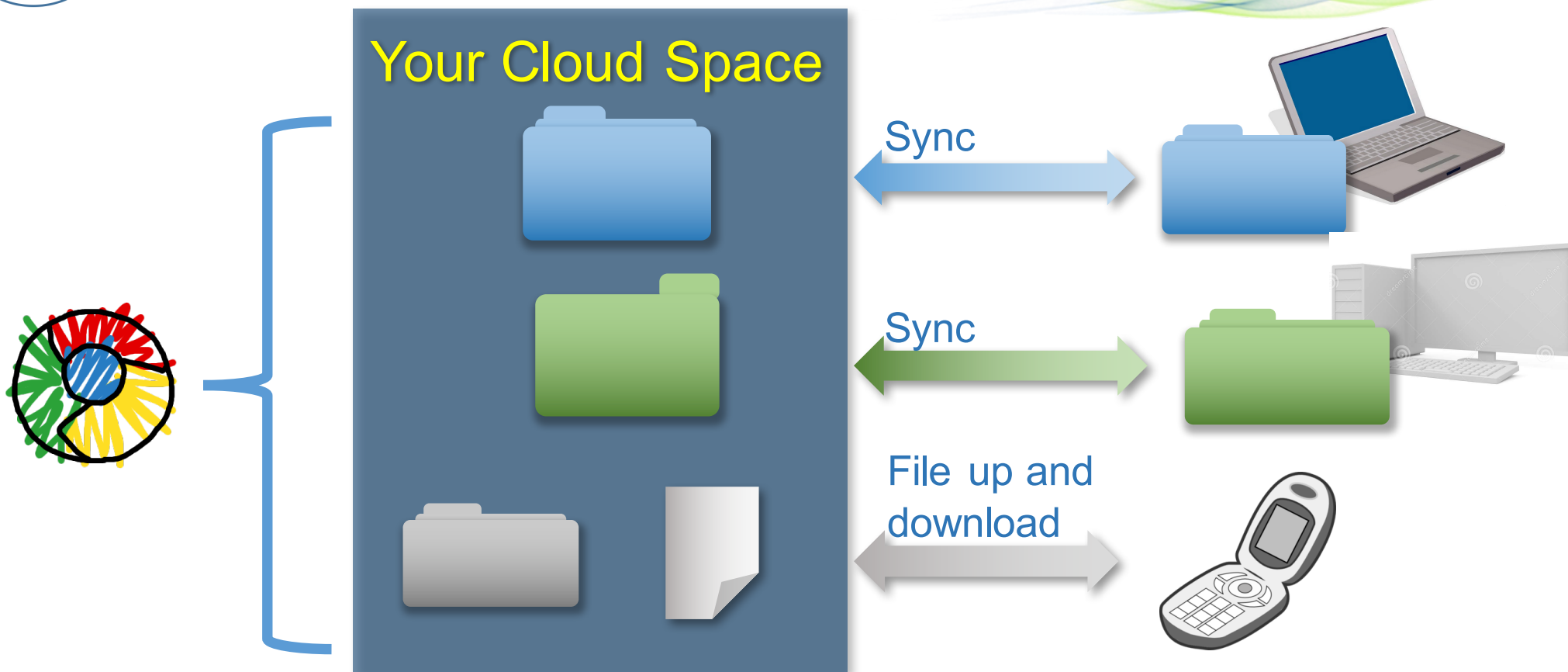
Requirements

dCache.org 

- Fine grained sharing of files and directories with individuals and groups.
- Sharing via intuitive Web 2.0 mechanisms (Apps or Browser)
- Sharing with 'the public' with or without password protection
- Sharing of space to upload data. (protected)
- Expiration of shares
- Automatic bidirectional synchronization of data between mobile devices and central repository.



Typical Application





Steps taken by DESY

dCache.org 

- Evaluated possible solutions in 2013.
- Decided to go for ownCloud
 - Provides most of the features needed.
 - Open Source
 - Was in use by many institutes and Universities in Germany
 - Used by colleagues at SURFSara (Amsterdam) and CERN
- Evaluation showed :
 - Very good Sync'n Share feature set
 - Very good in planning ahead (roadmap)
 - Plans for cross site federated access (now in place).
 - A bit weak in data management
- Started prototype installation at DESY beginning of 2014



What should the DESY Setup look like ?

(Actually will look like in July)



The Infrastructure

dCache.org 



Virtualization

User Management
Registry
LDAP



Authentication
Kerberos

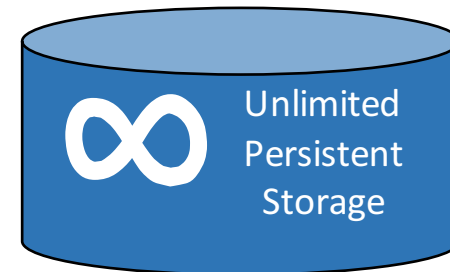
Monitoring



Accounting



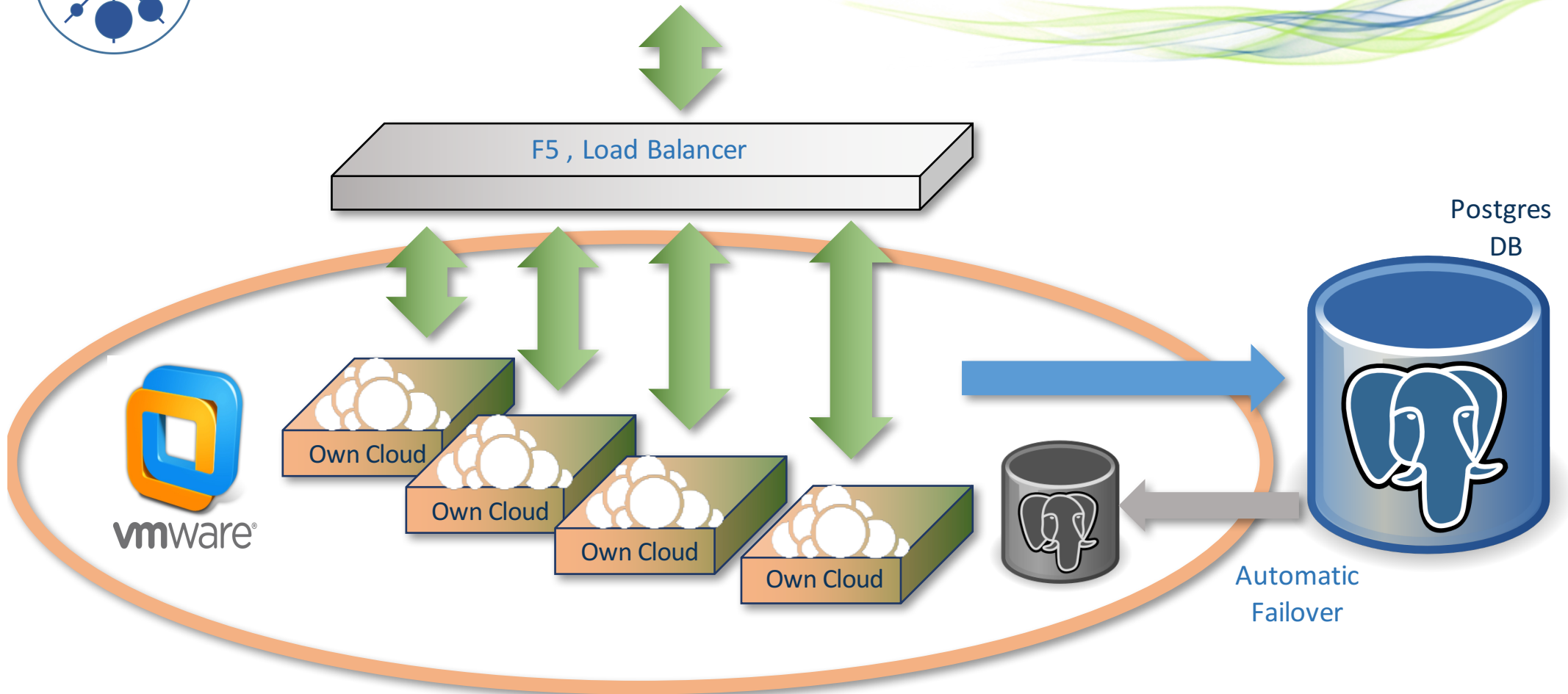
Local and Wide Area Network
Load Balancing Firewalls





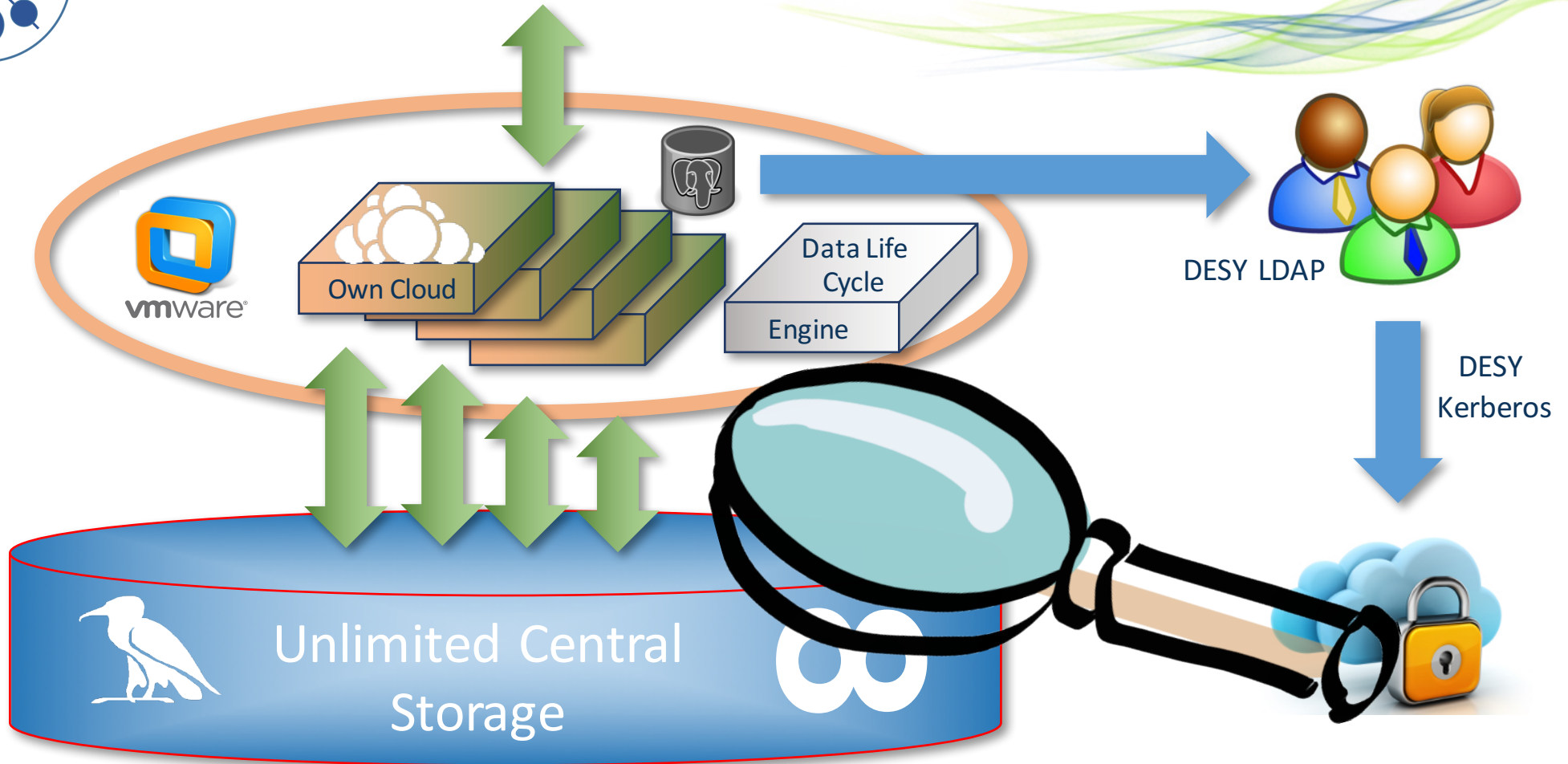
Infrastructure Integration

dCache.org 



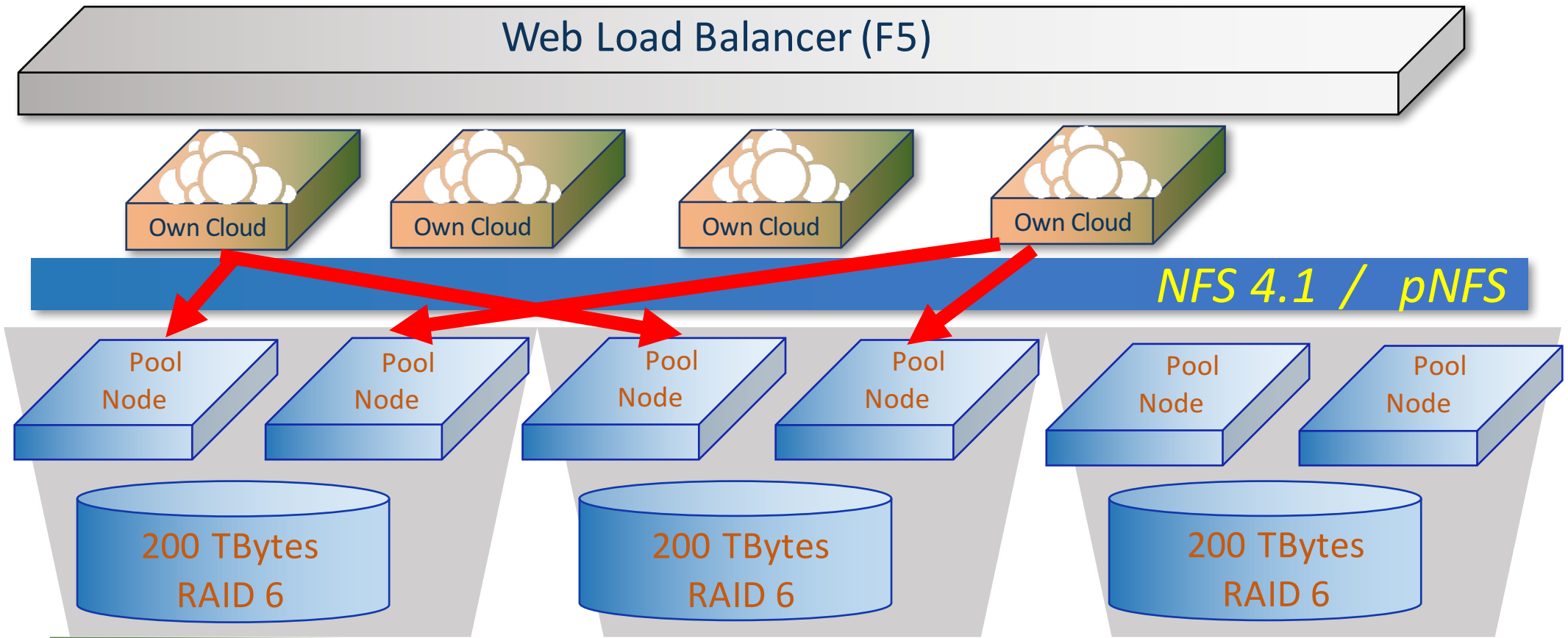


More Integration





Horizontally Scaling Backend dCache.org



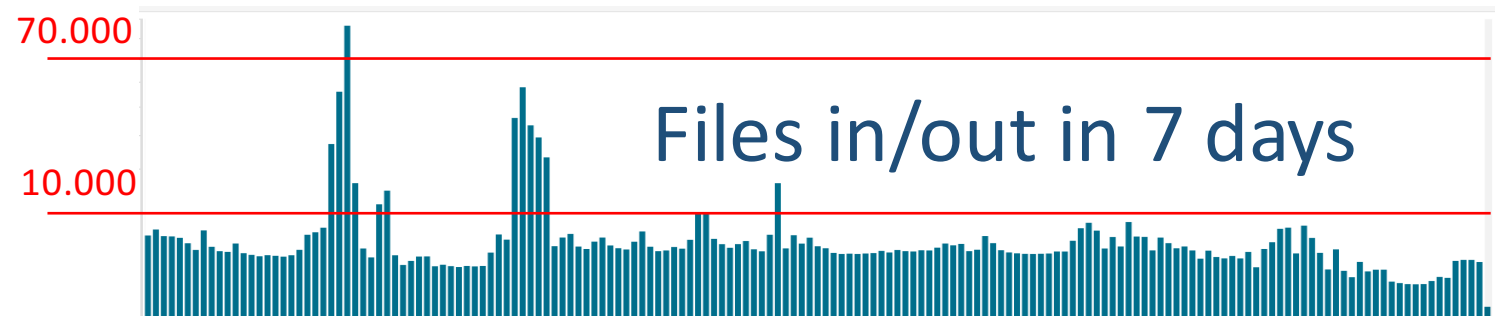


Some Statistics

Users Total	490
Users Active	277
Space Available	567 TBytes
Space Used	2 * 30 TBytes
Files	10 Millions

Current Default Quality
Two Replicas on
different storage nodes.

Files in/out per hour

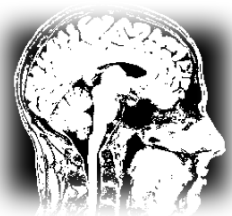
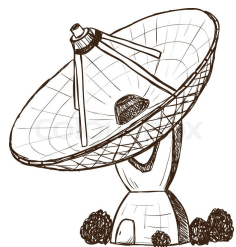




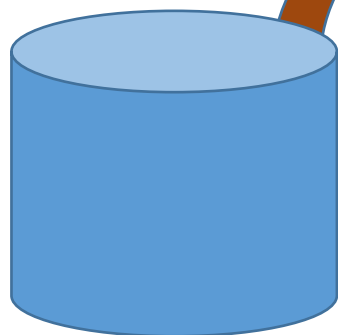
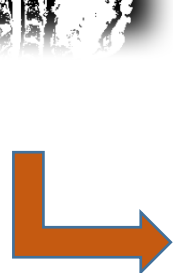
Is that sufficient for scientists ?



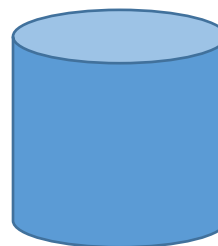
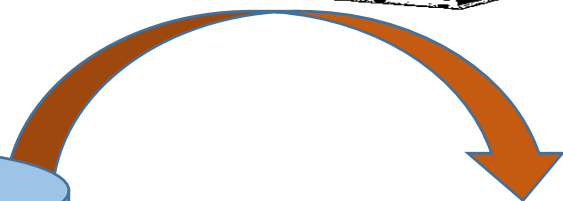
Typical Workflow



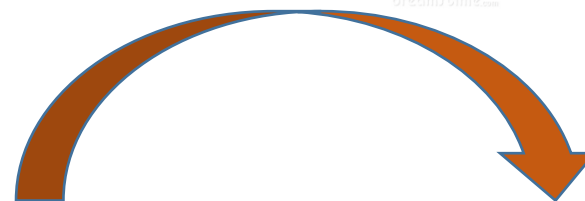
Sharing



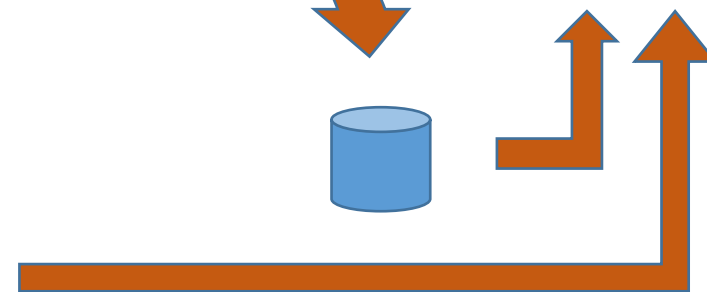
Raw



Derived



Publication





Data Categories

dCache.org 

Amount

Category

Typical Application

1 – 100 PB

Raw

LHC Detector data
Raw X-Ray Images
Brain Scans

10 - 100 TB

Derived

Reconstructed (Ntuples)
Purified Images
Brain Maps

1 TB

Publication

Papers, Presentations, Histograms



What do we need to support 'science workflows' ?



More Requirements

dCache.org 

- **Storage must be manageable : Defined QoS and Data Lifecycle**
 - Different type of data must have different QoS attached, regarding access latency (performance) and data durability (how safe is my data?)
 - Spinning Disk for streaming
 - SSD for fast random access
 - Tape for archive
 - Multiple copies in different locations on different media for long term data preservation
 - Moving data between different QoS types has to be performed
 - w/o service interruption
 - transparently to the user
 - w/o changes in the namespace



Quality of Service

Raw



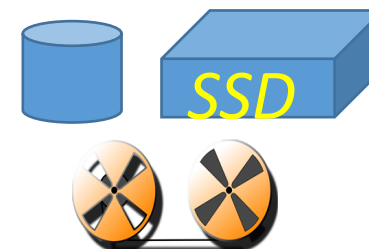
Long Term Preservation
(Legal Requirement)

Derived



Low Latency
(HPC, Analysis)

Publication



Fast, Multi Stream
Access



Even more Requirements

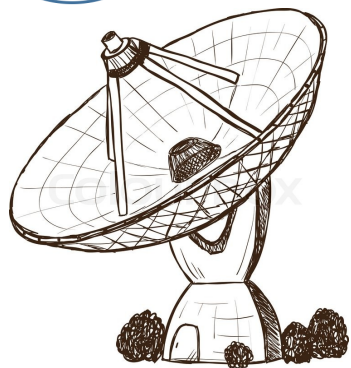
dCache.org 

- **Different access protocols for different applications**
 - POSIX Mounted FS (nfs4.1/pNFS) for fast analysis
 - FTP dialects (gridFTP) for wide area transfers with GLOBUS, WLCG-FTS
 - http/WebDAV mostly for browser based applications, visualization, ..
- **Different authentication mechanism must be available.**
 - Username/password for web applications
 - SAML to support traditional IdP's
 - Open ID Connect for google/facebook like IdP's
 - Certificates for https or GRID applications
- **Different credentials must be map-able to the same identity.**



Scientific Data Cloud

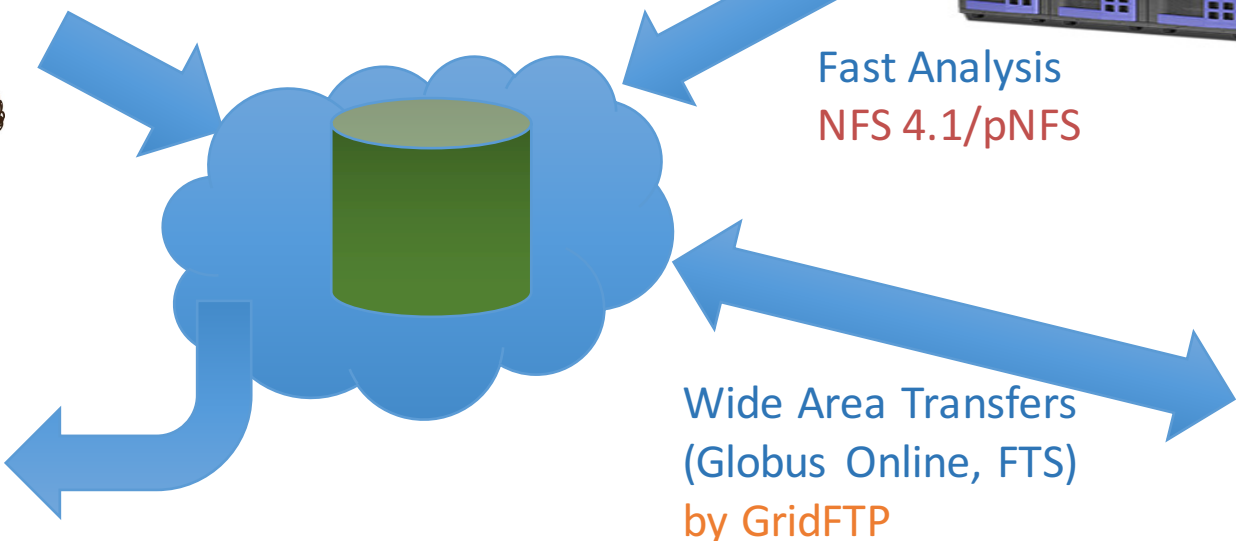
dCache.org 



High Speed
Data Ingest



Sync'ing and Sharing with OwnCloud



Fast Analysis
NFS 4.1/pNFS



Wide Area Transfers
(Globus Online, FTS)
by GridFTP

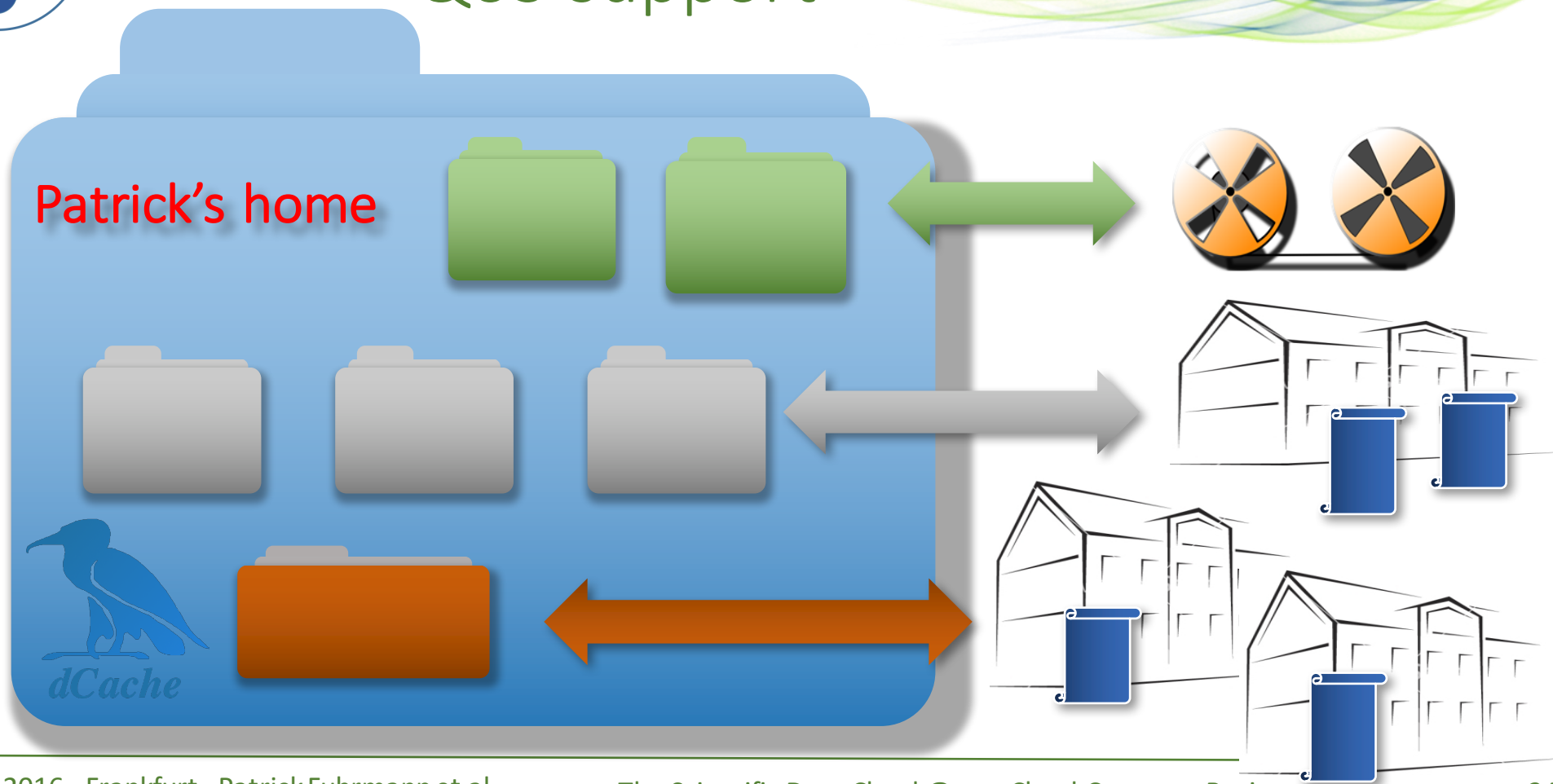


What would that look like
from the user's perspective ?



My DESY XXL Home QoS support

dCache.org 



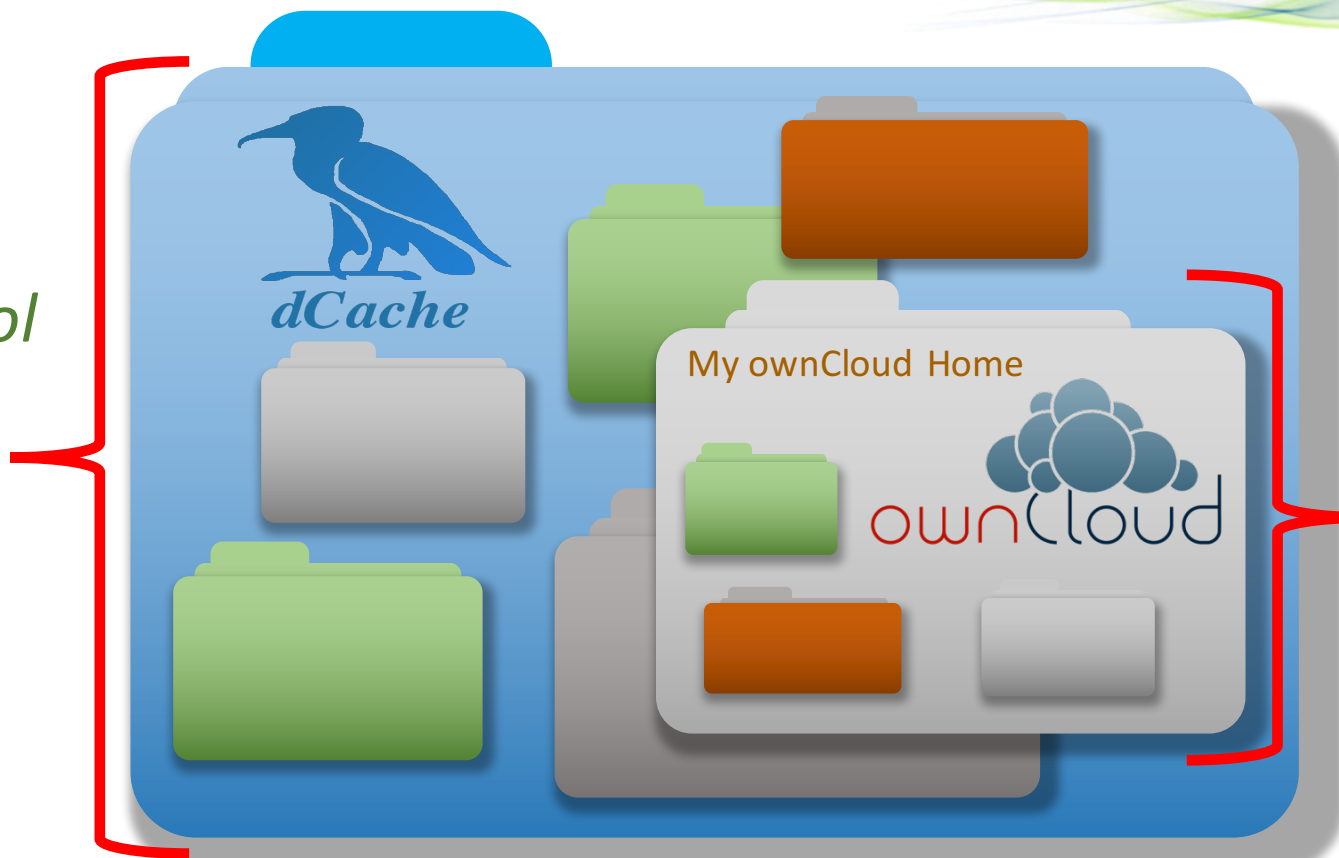


My DESY XXL Home Protocol Support



Multi Protocol

NFS 4.1/pNFS
GridFTP
WebDAV
SRM



Sync
Share
Web 2.0
ownCloud



How do we achieve those goals ?

dCache.org 

The dCache logo features a stylized blue bird perched on a wave of green and blue lines that flows across the top of the slide.

The scientific data cloud

OR

Choosing dCache as the storage backend for ownCloud !



Side Track

What's dCache ?



dCache in a nutshell (cont.)

dCache.org 

- Started 2000'
- International collaboration (DESY, FERMILab, NDGF)
- About 10 members : developers, deployment, support, management
- Software deployed at about 70 sites Europe, US, Asia, Russia
- Largest deployments in the order of 20 PBytes on tape and disk.
- Total storage close to 200 PBytes.
- Geographically largest installation spans 4 countries.
- Largely funded by INDIGO-DataCloud, DESY, FERMILab and NDGF



INDIGO DataCloud



dCache Design



NFS/pNFS

httpWebDAV

gridFTP

Protocol and Authentication Engines

Virtual file-system name space Layer

Media Transfer Engine and Pool Management



Automatic and Manual Media transition

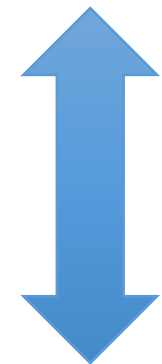
SSDs



Spinning Disks



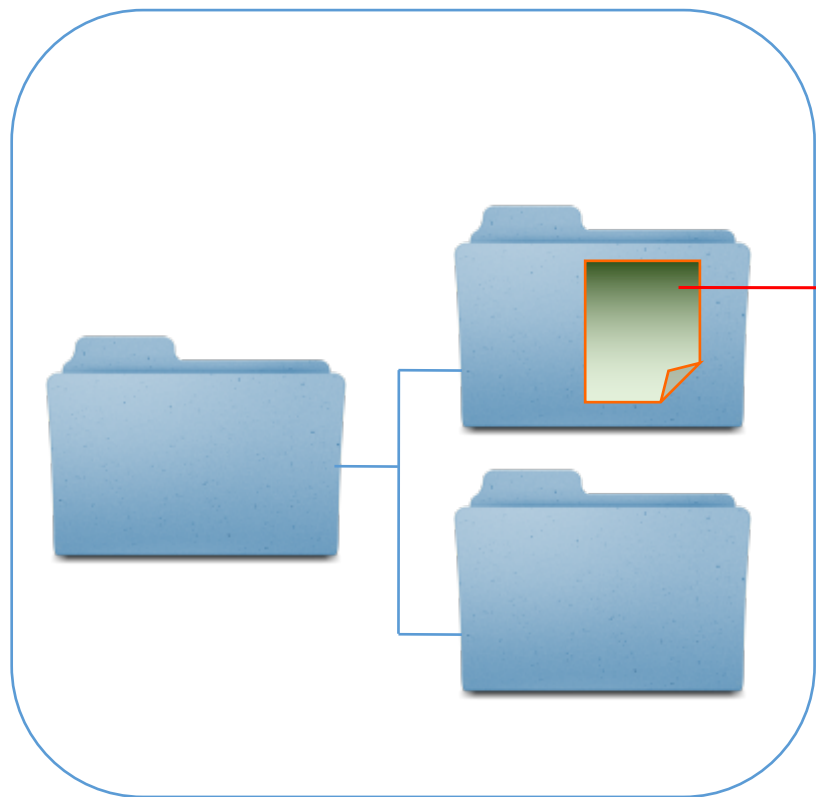
Tape, Blue Ray ...





Namespace Design

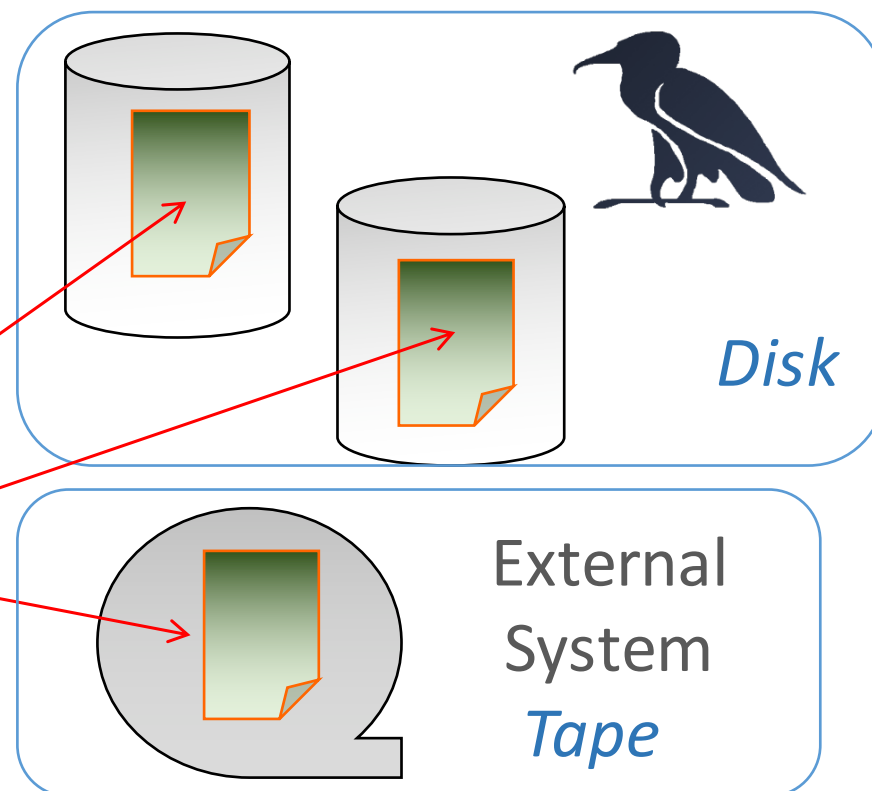
Name Space



Location Manager

Name
Disk 1
Disk 2
Tape 1

Physical Storage





Design Consequence

dCache.org 

- Files are stored as objects on various data back-ends
 - Random Devices : Harddisk, SSD
 - Removable Media : Tape
 - Object stores : CEPH
- Back-ends can be highly distributed (even beyond countries).
- The File namespace engine is independent of the data storage itself.
- Internal and external services can move data around w/o service interruption.



dCache Features

supporting our idea of a **scientific data cloud**

dCache.org 

- **Multi Protocol Support (Transfer and Authentication)**
 - Transfer protocols : NFS/pNFS, http, WebDAV
 - Multi Authentication Credential support (OpenID Connect, Kerberos, passwd)
- **Sophisticated Data Management**
 - Multi Media support (Tape, Spinning Disk, SSD, ...)
 - Automatic and manual media transitions
 - Adding and removing data nodes w/o service interruption
 - Automatic replica management
 - Enforces $n < x < m$ copies of data files.
 - External storage support (e.g. Tape systems: TSM, HPSS, OSM, DMF)

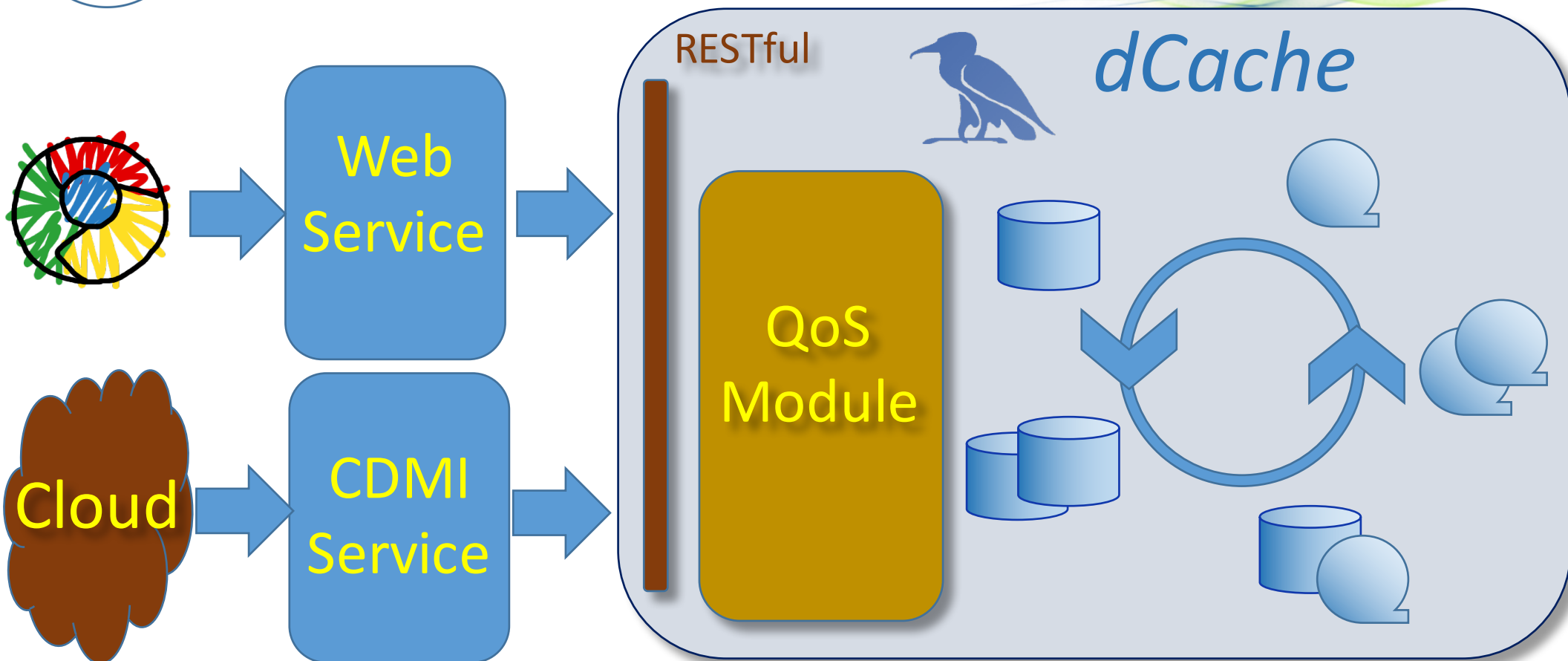


In particular : The QoS Interface



dCache QoS Interfaces

dCache.org 





The QoS Web Interface



dCache View

dCache.org Root patrick

DISK **TAPE**

eurogate.banner	5/2/2016, 6:15:16 PM			513 Bytes
indigo.banner	5/2/2016, 6:15:16 PM			331 Bytes
dcache.banner	5/2/2016, 6:15:15 PM			387 Bytes

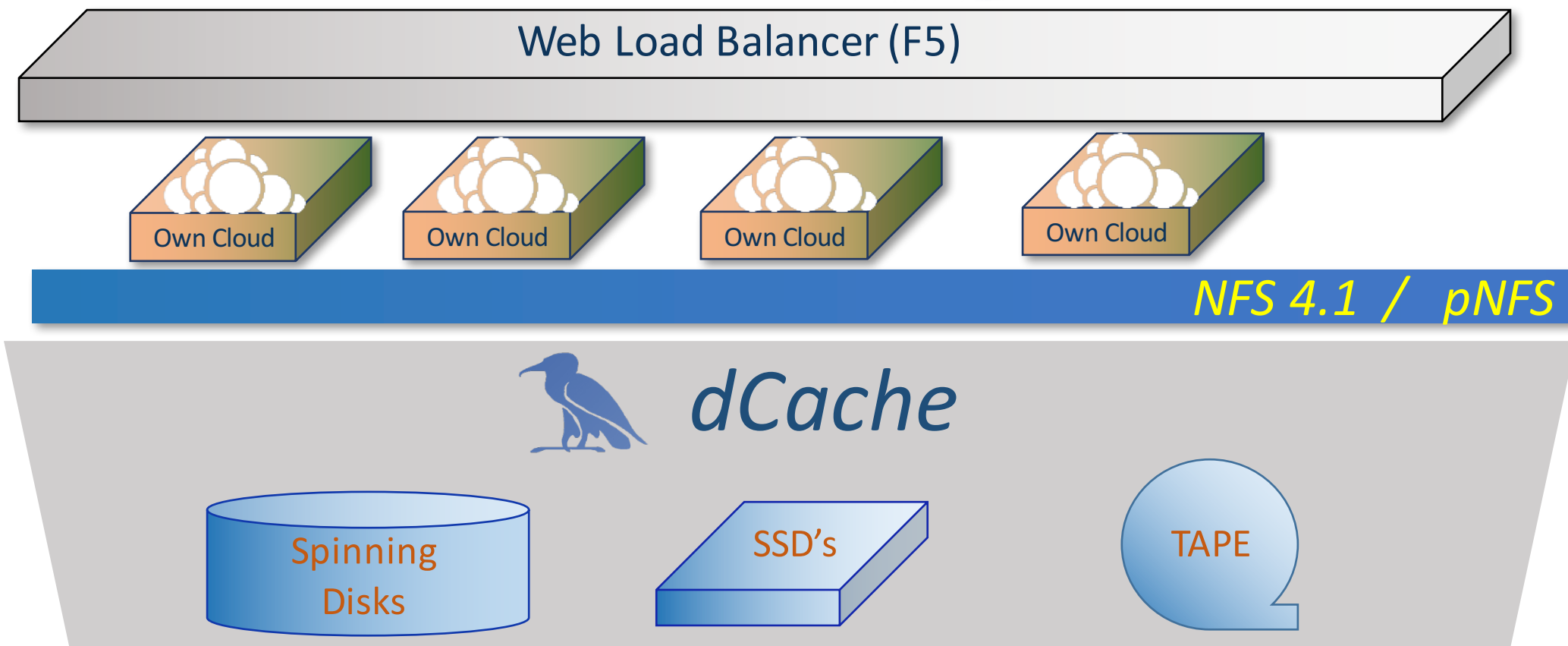
Click, to get File back from Tape.



Putting pieces together



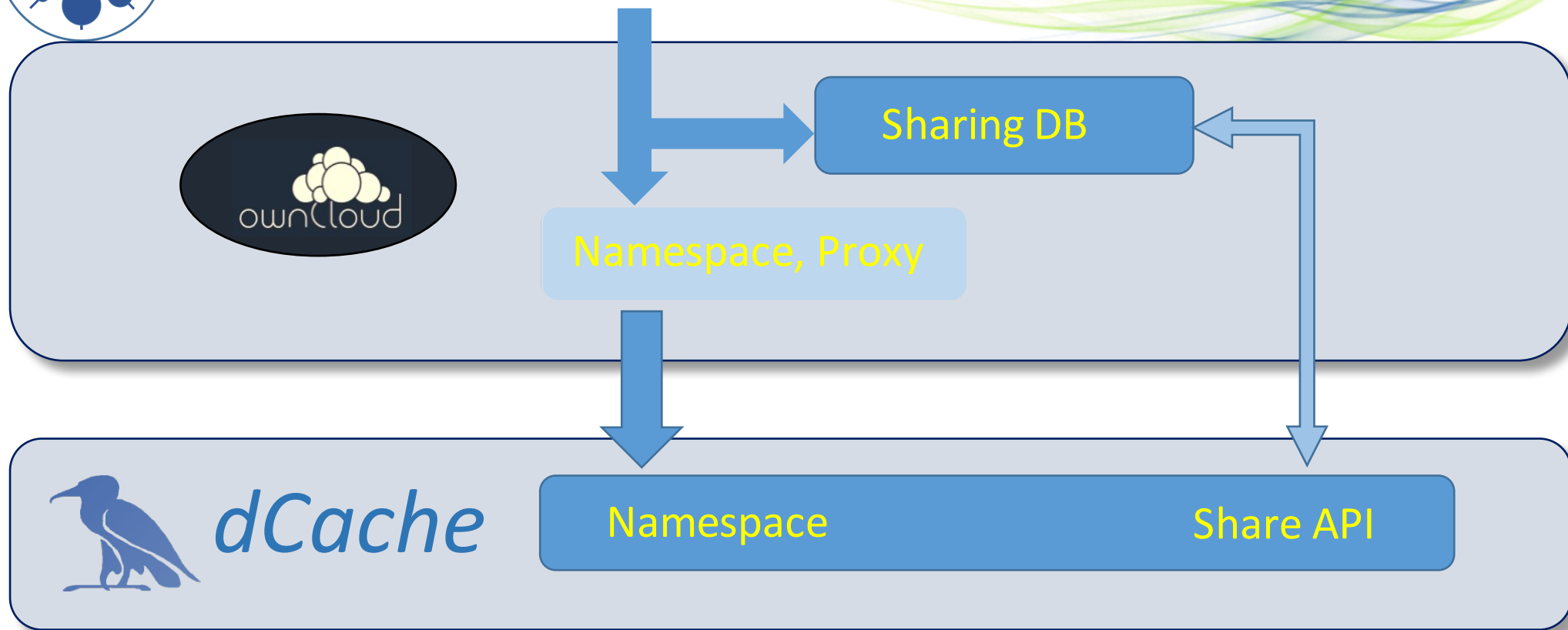
The Data Path





Future Work The Namespace Path

dCache.org 





dCache – OwnCloud hybrid

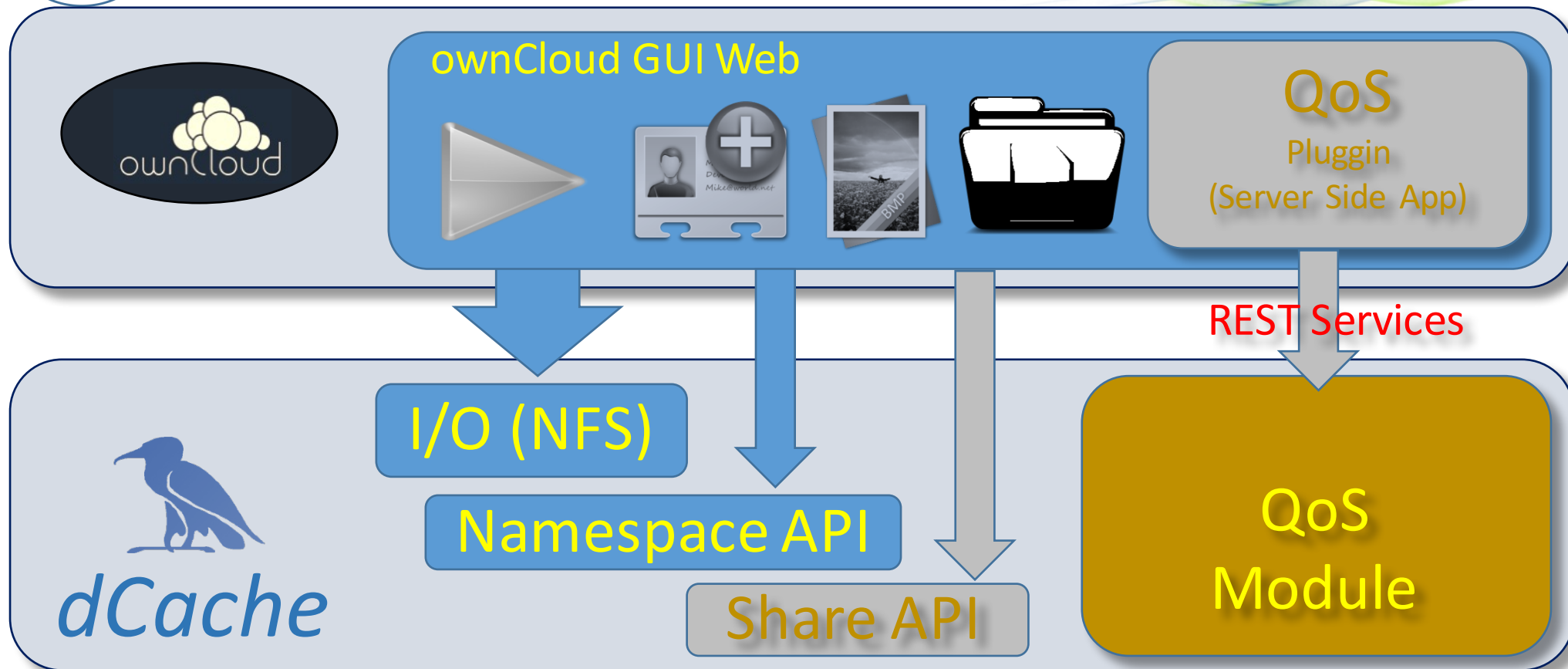
dCache.org 

- Data path is the easiest part. Works nicely.
- Namespace synchronization is/was very difficult
 - Important to let all protocols see synchronized namespace.
 - ownCloud didn't expect the underlying storage system to change name space tree.
 - Manually triggered synchronization took too long.
 - OwnCloud 9 provides first attempt for an API for external namespace.
- Exposing 'shares' to external component not yet in ownCloud.
 - Important to allow all protocols to use ownCloud-defined shares.
 - Prerequisites :
 - ownCloud : needs API to expose 'shares'
 - dCache : needs to have a 'share' object implemented.



ownCloud and QoS

dCache.org 





Summary

dCache.org 

- An OwnCloud - dCache Hybrid is a perfect system for providing managed shared storage to scientists.
- Sync'n Share is provided by ownCloud.
- Access protocols and Authentication Mechanisms used in science are provided by dCache.
- Unlimited storage spaces (via removable media, e.g. tape)
- Quality of Service support
 - automatic and manual media transitions
 - Automatic replica management resulting in high availability and data durability.
- Reduced downtimes due to transparent data migration.



Outlook



- The current version of the ownCloud-dCache Hybrid satisfies the need for
 - Sync'n Share
 - Highly scalable and manageable back-end storage
- For a full integration
 - The name-spaces of the two systems need to be synchronized (OC9)
 - The ownCloud 'shares' need to be exposed to have them visible in all protocols (nfs, gridFTP, ...)
 - We need to provide an ownCloud plugin (server side app) to make the dCache QoS storage types visible in ownCloud.



The END

further reading
www.dCache.org