

dCache, managed Cloud Storage

For CHEP'16 in San Francisco

Patrick Fuhrmann

On behalf of the project team



That this presentation about ?

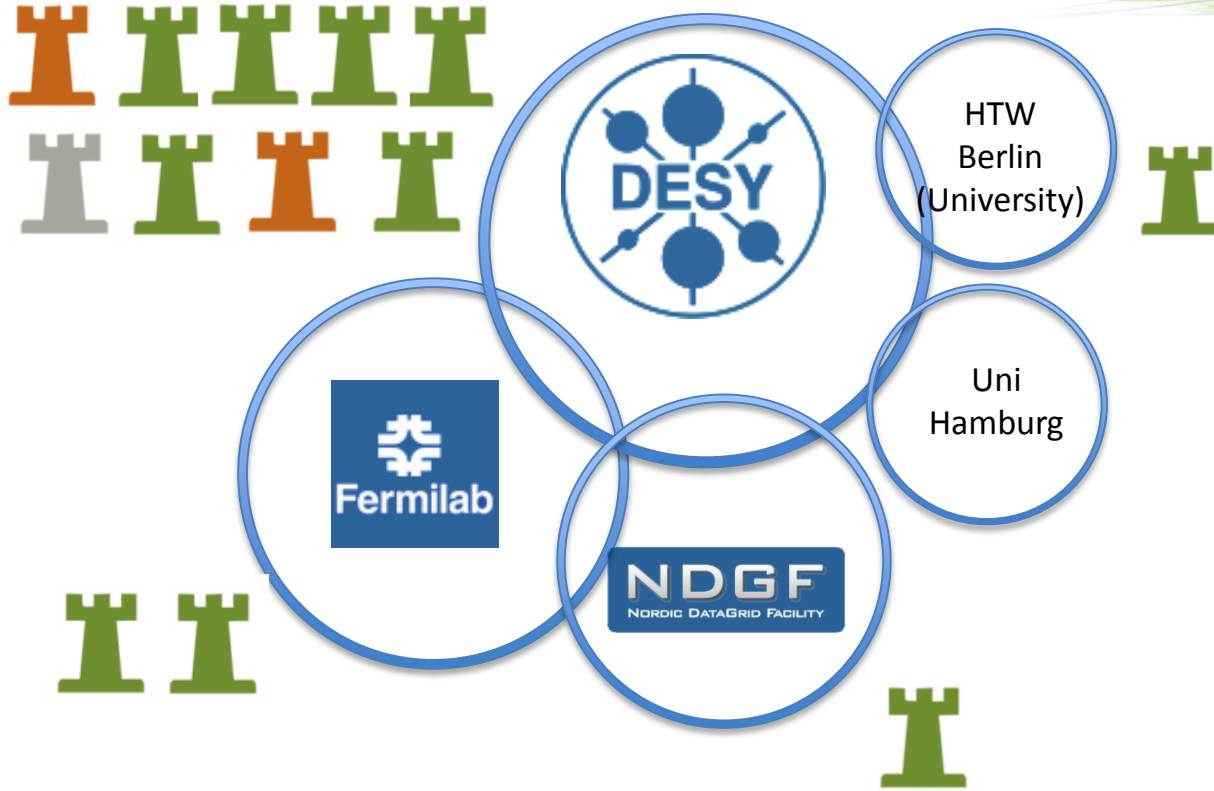
- Some remarks on the team, funding and high level objectives.
- Managing the system
 - Full High Availability Functionality
 - Managing underlying storage; delegate storage (CEPH)
 - Managing system configuration and monitoring (REST)
- Managing your storage
 - Quality of storage management
 - New Resilient Manager
 - New APIs (REST API for QoS translated to CDMI)
- Cloud access
 - ownCloud © interface
 - New authentication and authorization mechanisms

The Technology Cheat Sheet

- Combines heterogeneous storage nodes under a common virtual file system tree and scales into 100PB region.
- Provides access to data via a variety of protocols, e.g. NFS4.1, WebDAV, GridFTP, etc.
- Provides a variety of authentication mechanisms, like User/Pass, X509 Certificates, Kerberos, in preparation SAML and OpenID Connect, Macarons.
- Multi Tier support: moves data around between different media types, like Tape, Spinning Disks and SSDs.
 - By user request.
 - Automatically based on the access profile, hot spot.
- Provides resiliency, e.g. through multiple automatically managed copies.

Some strategic considerations

The dCache.org collaboration



Funding and roadmap



EOSC

???

INFRA-12-2017

Standardization

Deploying new technologies into Production and exploring new communities



INDIGO DataCloud

Data Life Cycle

Multi Tier Storage

Quality of Service

Migration Archiving

Deployment

Interoperability

Reducing Costs

AAI

2010

NFS 4.1 / pNFS

HTTP / WebDAV

Contributing to the Dynamic Federation

2013

2015

2018

Improved System Management I

Full High Availability

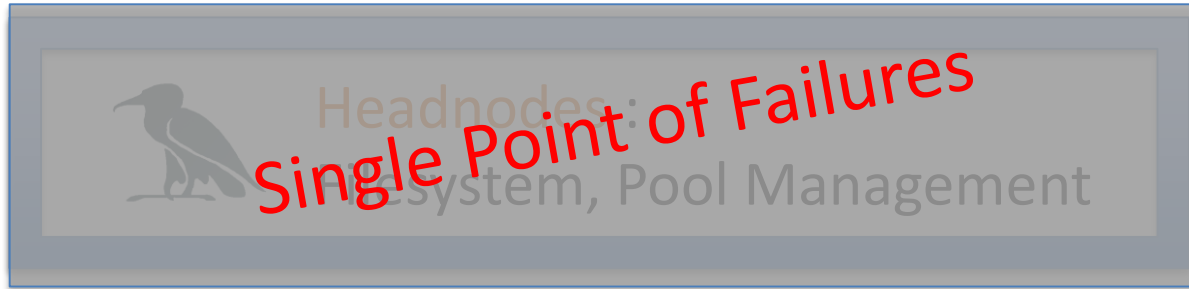
With the

- increasing size of dCache installations,
- the usage beyond WLCG, with more challenging SLAs and with
- the pressure to run 'operator less' for days
 - Weekends
 - Dark Data Centers
 - Reduce Costs

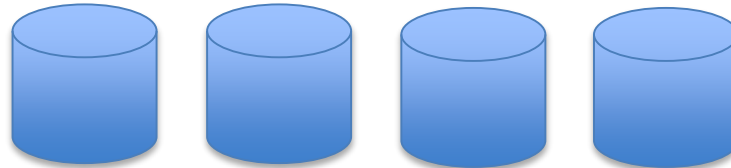
High Available functionality became inevitable.

High Availability

Redundant **Doors**
(Protocol Engines)



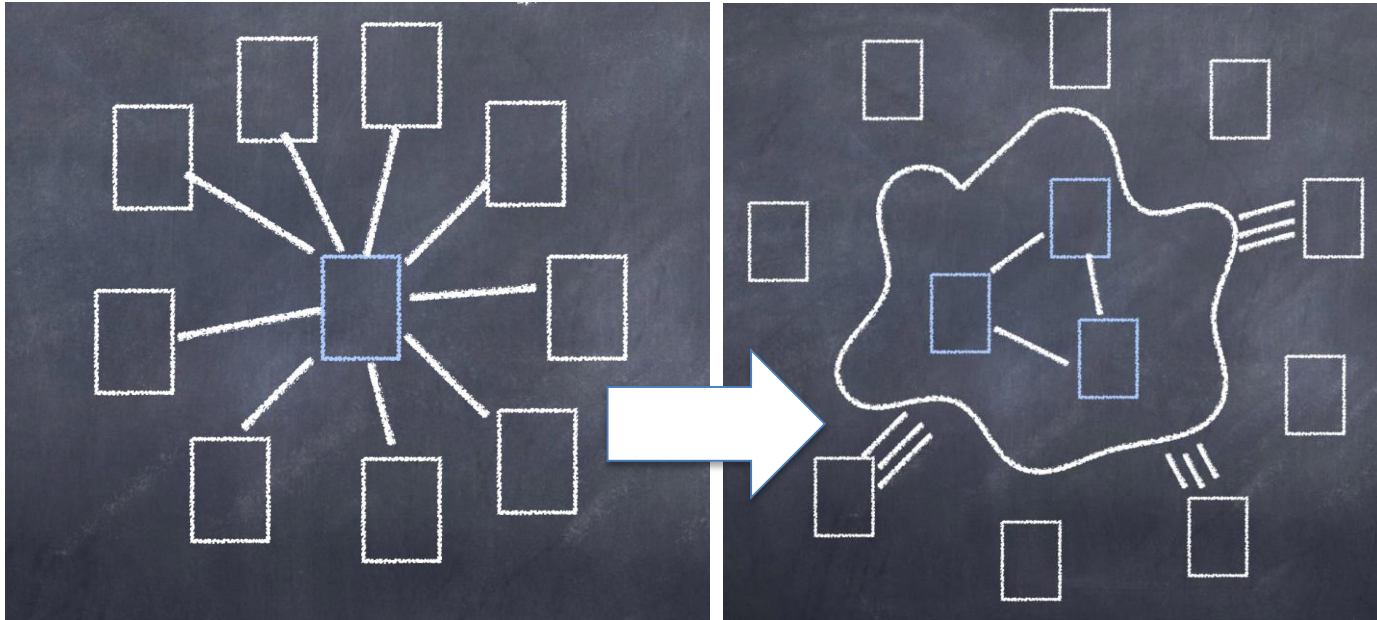
Redundant **Pools**
(Replica Manager)
Multiple File Copies



Mainly the following issues need to be fixed for dCache HA

- The dCache core “message passing system” had to be fixed to overcome failures on single path segments: ‘rerouting’
- dCache sub-services needed to become redundant
 - Single point of failure services
 - state-less services

High Availability: Message Passing

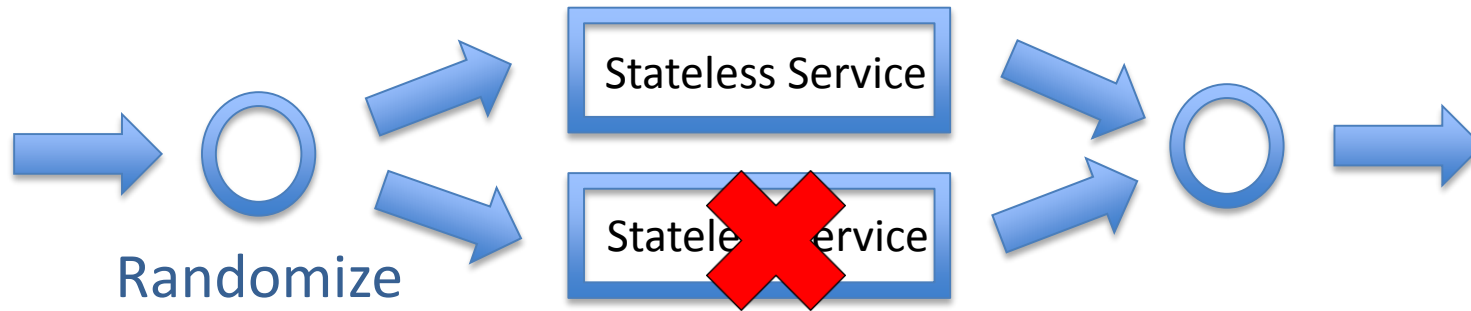


Any single component can fail, w/o breaking the service

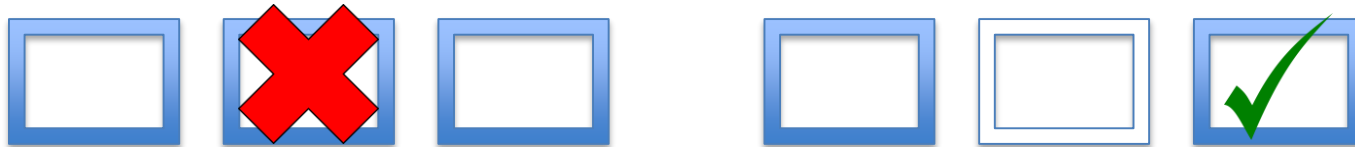
Stolen from Gerd Behrman, NDGF

High Availability

- Stateless services : use publish subscribe



- Singletons (build quorums, e.g. using Zookeeper)



Result (With 3.0)

At any point in time, one internal service (node) can fail without consequences for the overall service.

Essential for a huge 24/7 installations.

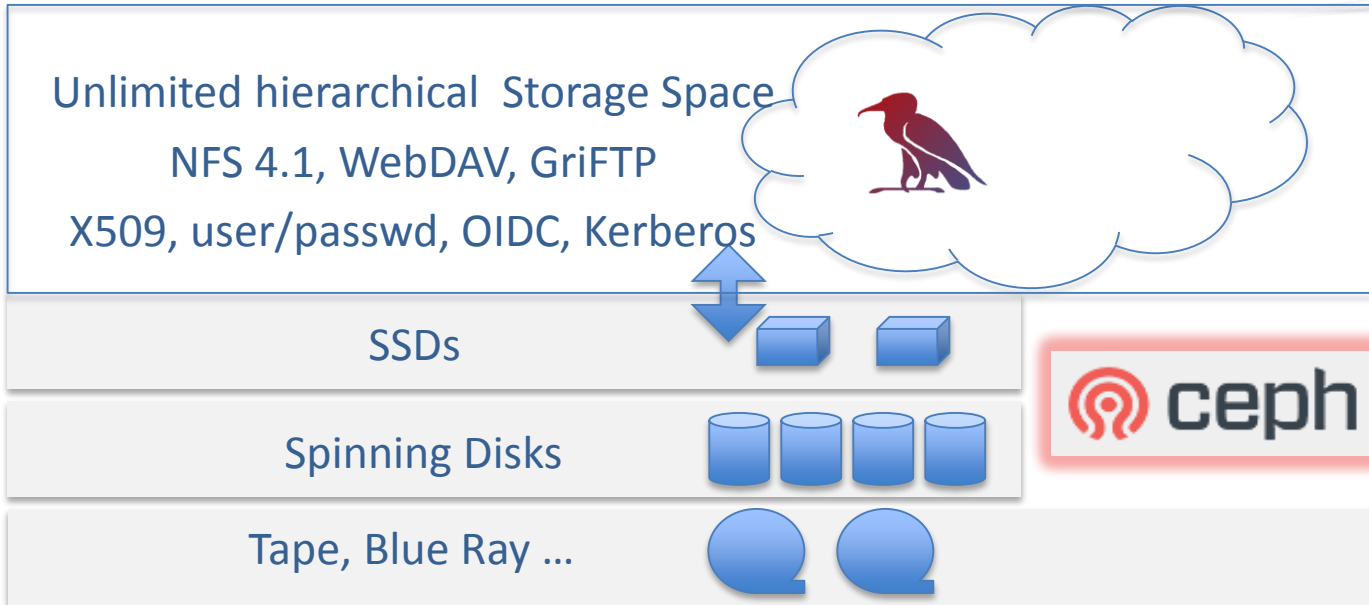
Improved System Management II

Delegating low level storage functionality

Delegating work to external Storage Layers

- Provides a single-rooted namespace.
- Metadata (namespace) and data locations are independent.
- Uniquely handles different Authentication mechanisms like x509, Kerberos, login+password,
- Provides access to the data via variety of access protocols (WebDAV, NFSv4.1/pNFS, xxxFTP, DCAP, Xrootd, DCAP)
- Provides data migration between multiple tiers of storage (DISK, SSD, TAPE).
- Aggregates multiple storage nodes into a single storage system.
- Manages data movement, replication, integrity.

Can be delegated



See Tigrans Talk
Today 14:00 at
GG C3

Benefits: Best of two worlds

- Low maintenance CEPH functionality
- High level dCache functionality : access protocols and authentication mechanisms

Moving to managed storage or Quality of Service in Storage

Why ?

Storage has different qualities and
different prices.

Pre-requisite for "Support for Quality of Service in Storage"

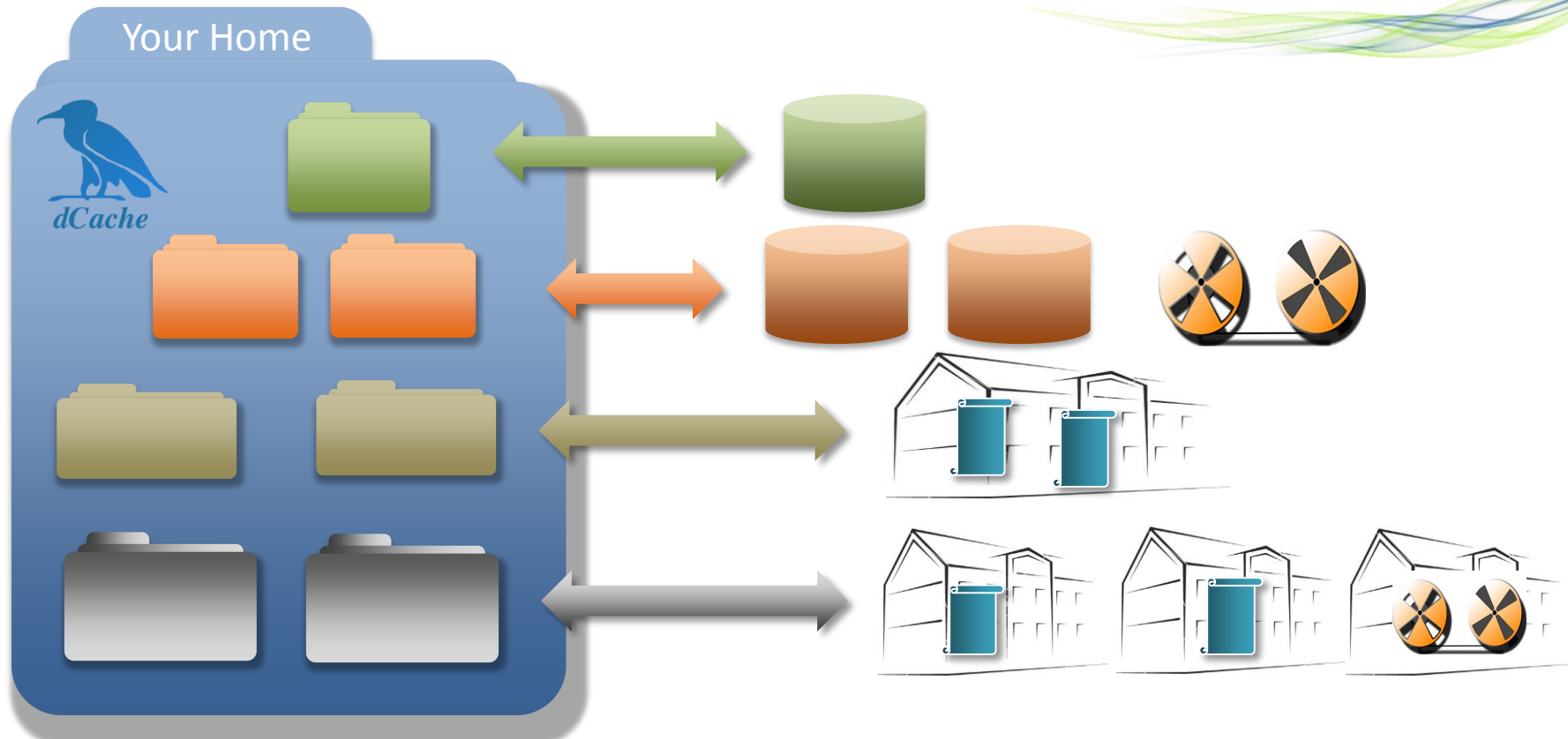
- Integrated support for different storage media
- Transparent transition between media types based on
 - Internal rules
 - APIs for remote "Platform as a Service" systems
 - Manual intervention (Operator , User)
- Consistent and transparent management of replica
 - System must be aware of media attributes (Speed, Latency, Price)
 - System must be aware of locations (Rack, Power supply group, Buildings)
- Management API's (SRM, REST, CDMI)
- Graphical Interfaces

Integrated Media Technology Support



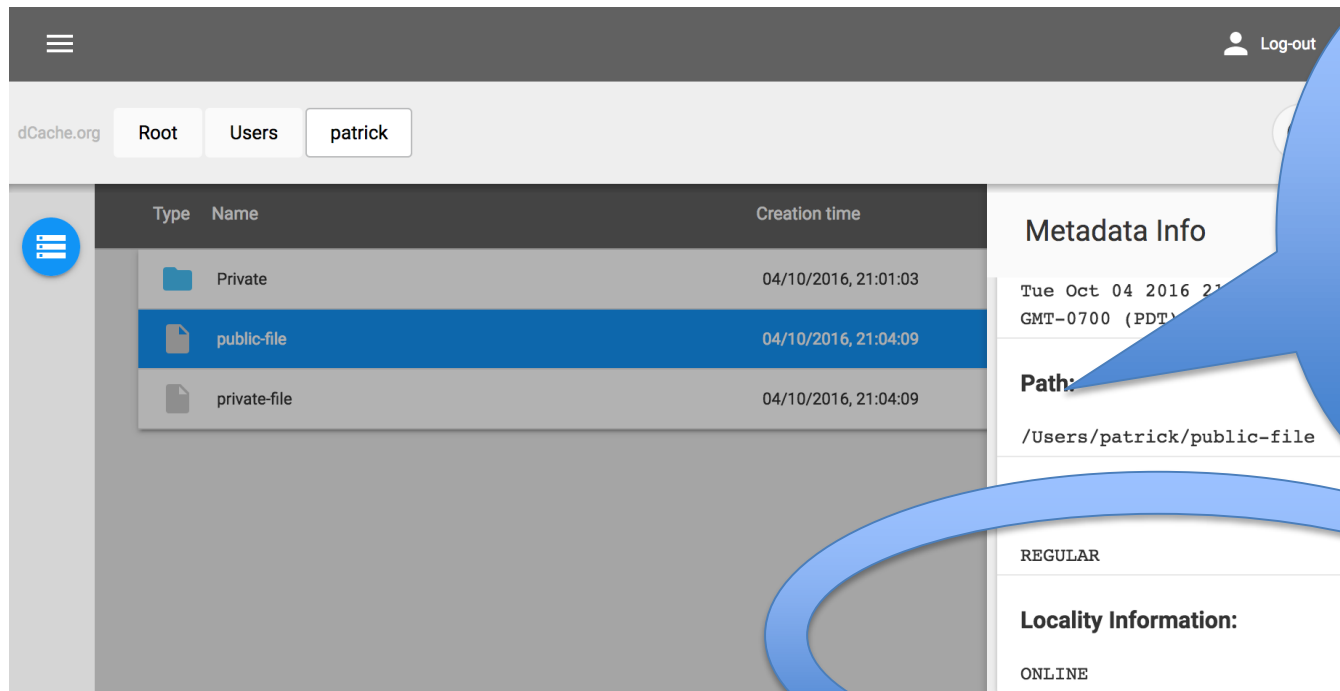
Technology	Driver
Spinning Disk	POSIX Driver
SSD	POSIX Driver
Shingled (SMR)	POSIX (disable remove)
	TAPE Driver
Removable devices	TAPE Driver
CEPH	New CEPH Block Driver

Directory Based Storage Quality selection



Disk replicas managed by "Replica Manager" see Poster

Storage Management GUI



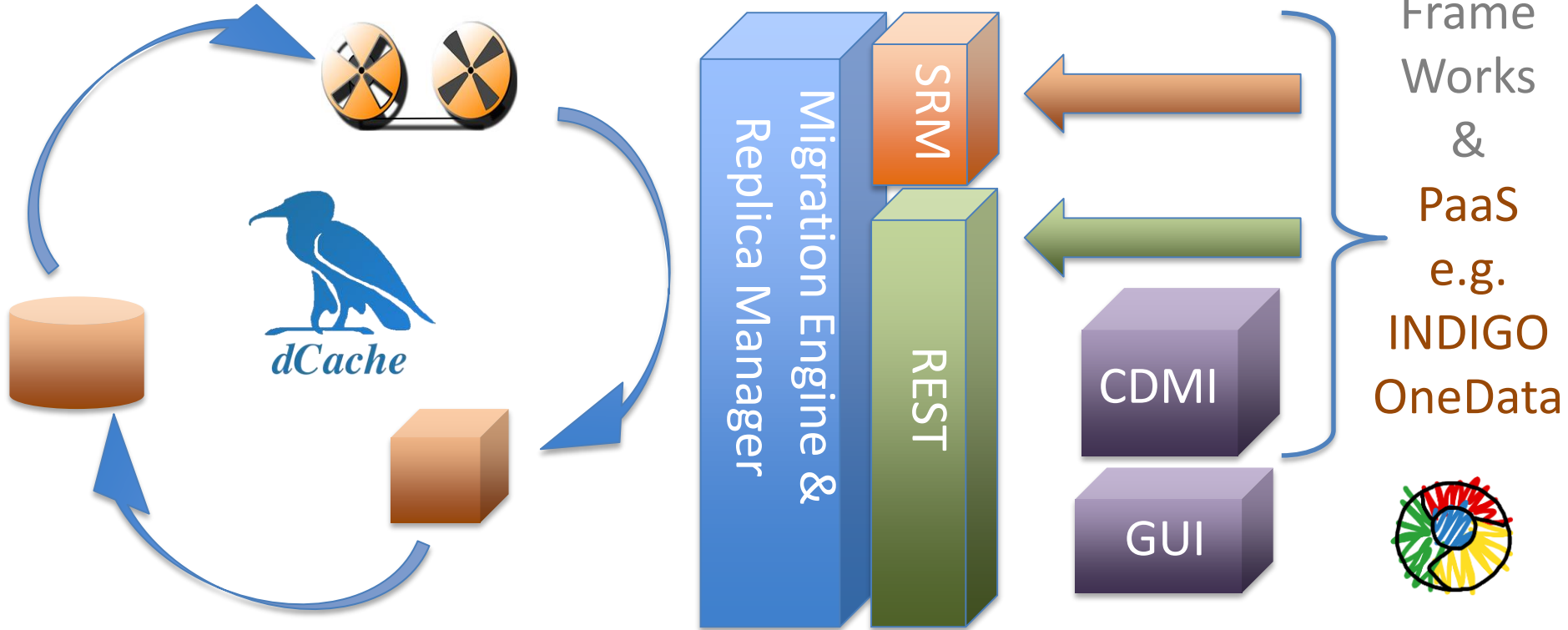
The screenshot shows the dCache Storage Management GUI. At the top, there is a navigation bar with a hamburger menu icon on the left and a 'Log-out' button on the right. Below the navigation bar, the breadcrumb path is 'dCache.org > Root > Users > patrick'. The main content area is divided into two sections. The left section is a table listing files and folders:

Type	Name	Creation time
Folder	Private	04/10/2016, 21:01:03
File	public-file	04/10/2016, 21:04:09
File	private-file	04/10/2016, 21:04:09

The right section displays 'Metadata Info' for the selected file 'public-file'. It shows the date and time: 'Tue Oct 04 2016 21:04:09 GMT-0700 (PDT)'. Below this, the 'Path:' is shown as '/Users/patrick/public-file'. There are two dropdown menus: the first is set to 'REGULAR' and the second is set to 'ONLINE'. The 'Locality Information:' label is circled in blue.

Locality
QoS
DISK
TAPE

QoS Management interfaces



For details see Pauls presentation tomorrow 11:15.

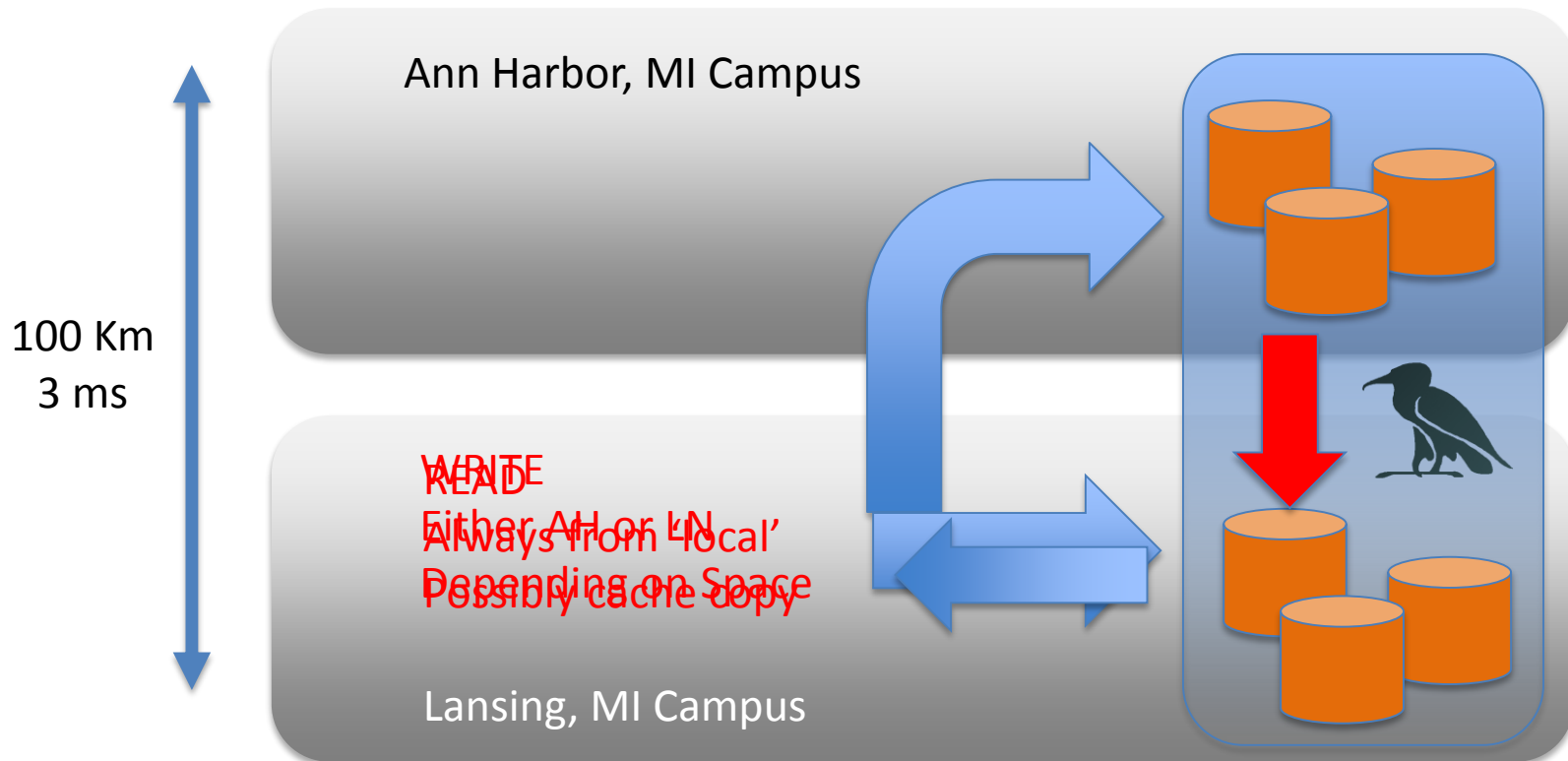


Managed Storage II

Managing data flows in federated systems.

Federating Storage

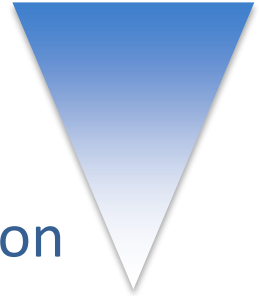
The Michigan Setup



Provided by Shawn McKee

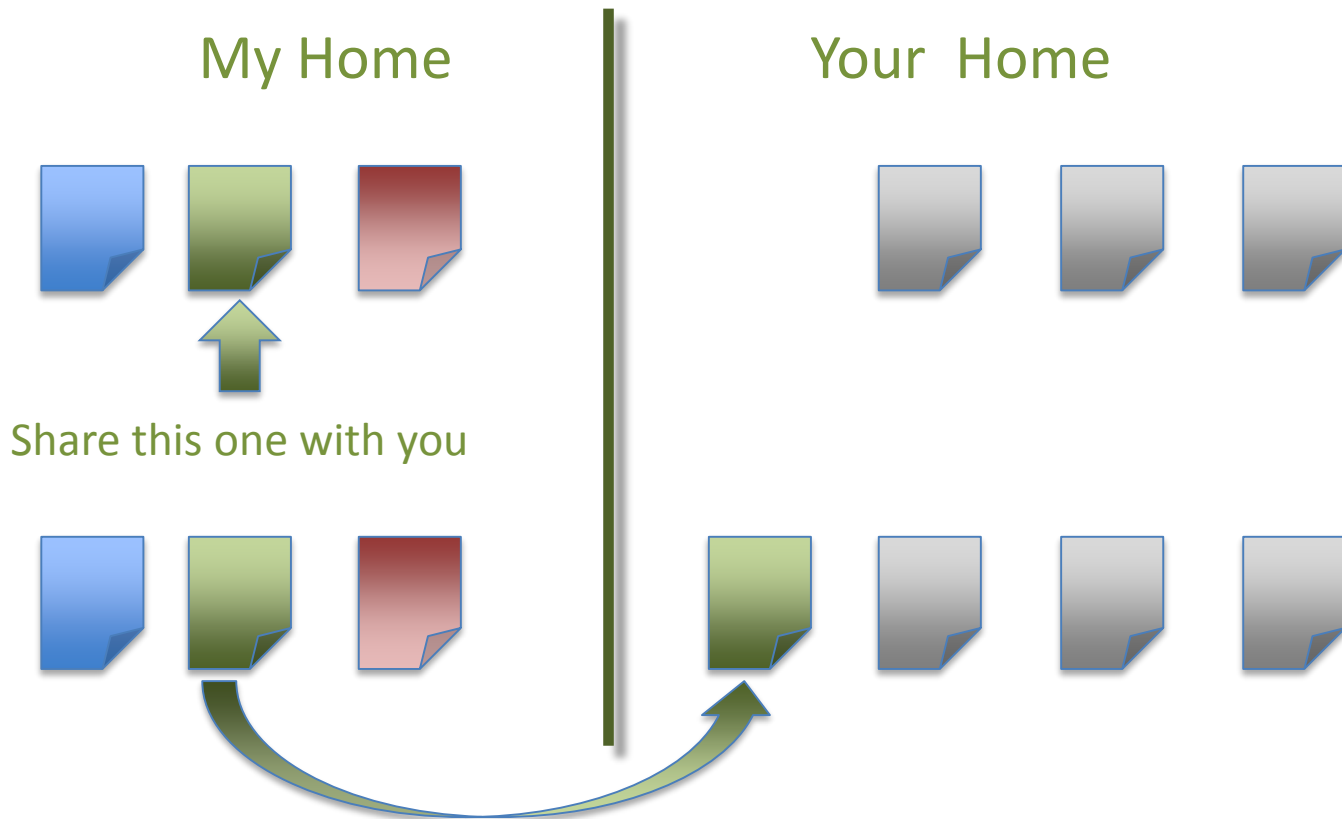
Progress of supporting x-site federation.

- Federation Mechanisms in place for long time.
- Remote component secure connection with 3.0
- Remote component authentication and authorization following soon.



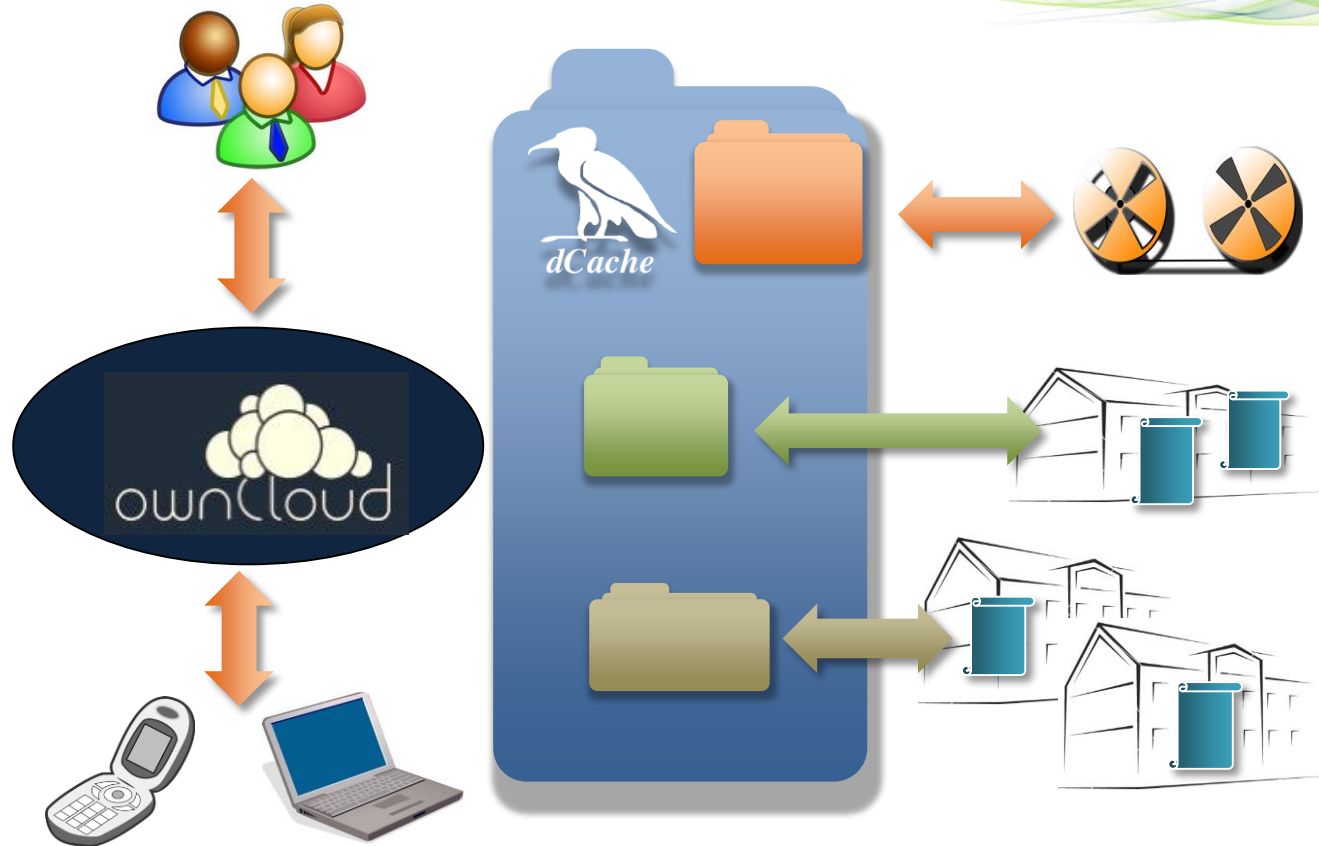
The Cloud Part

The cloud feeling, The **Sharing**



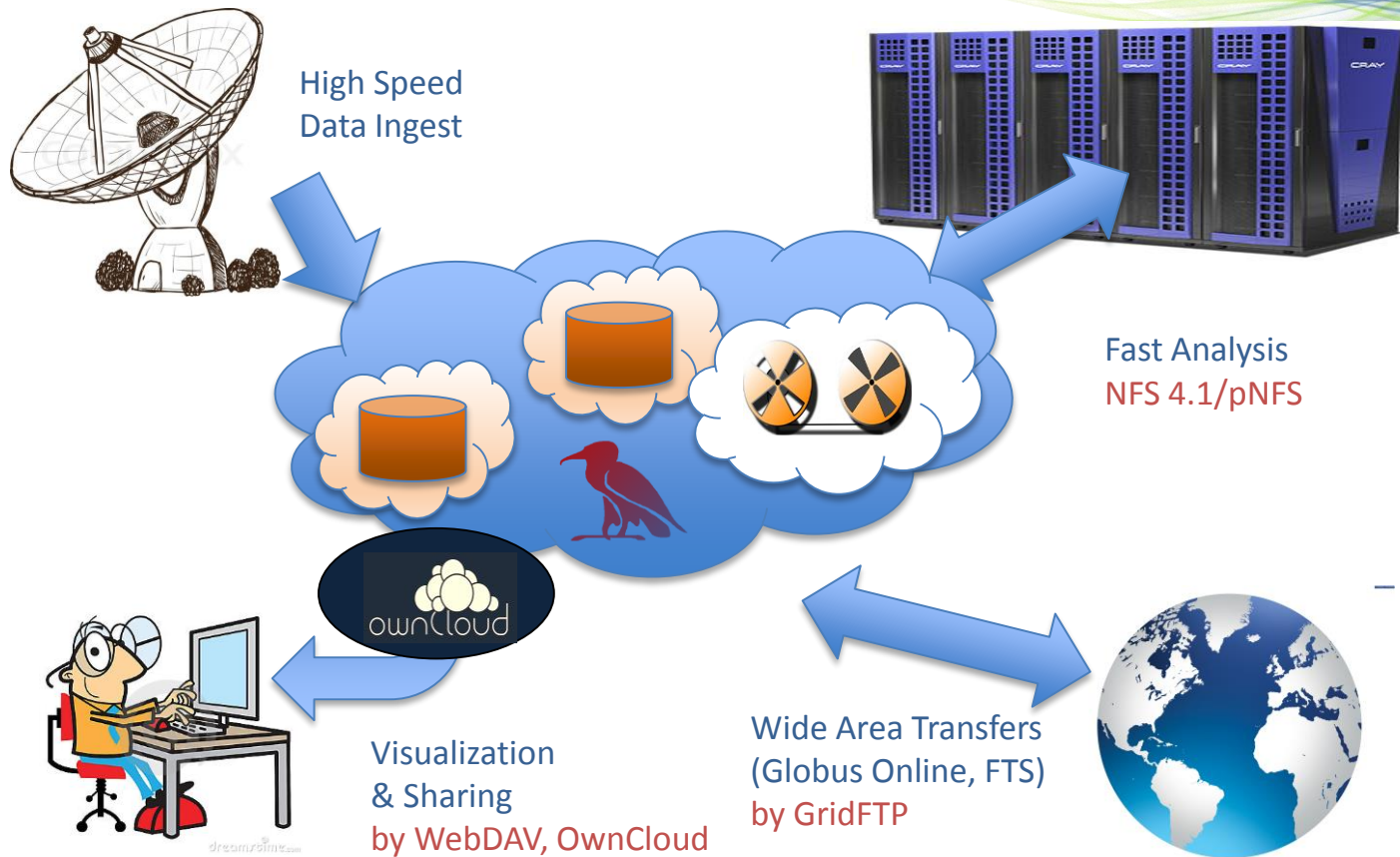
We decided to go for ownCloud © to get this done.

The Hybrid System



Putting it all together

The scientific Data Cloud



Well on the way for new strategic direction on storage:

- Bringing costs down for operations
 - Using ‘low maintenance’ storage backends (e.g. CEPH)
 - Introducing High Availability mechanisms to support ‘dark data centers’
- Allowing Quality of Service in storage to be defined by sysadmins and users.
 - With GUI
 - Remotely : SRM, REST, CDMI
- Introducing Cloud mechanisms (e.g. sharing) not only for home directory type data but as well for experiment data.
- Implementing the “Scientific Storage Cloud”
- Stick with industry standards
 - Allow sharing resources at sites
 - Benefit from development and fast turnaround in security components and mechanisms.

The END

further reading
www.dCache.org