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
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

<table>
<thead>
<tr>
<th>1992</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>HERA [Tier 0] (1992 – 2007)</td>
<td>Petra III [Tier 0] (2012 ... )</td>
</tr>
<tr>
<td>Particle accelerator (Proton – Electron)</td>
<td>Synchrotron Radiation</td>
</tr>
<tr>
<td>6.3 Km (Ring)</td>
<td>14 Beamlines</td>
</tr>
<tr>
<td>Some hundred scientists</td>
<td>Beamline Guest Scientists</td>
</tr>
<tr>
<td>5 PB in total</td>
<td>1 PB / year – 5 PB / year</td>
</tr>
<tr>
<td>LCG [WLCG Tier 2] (2008,2009 ... )</td>
<td>European [Tier 0] XFEL (2017 ... )</td>
</tr>
<tr>
<td>Particle accelerator (Proton – Proton)</td>
<td>3.4 Km (Linear)</td>
</tr>
<tr>
<td>26.7 (Ring)</td>
<td>2017 (First beamline)</td>
</tr>
<tr>
<td>About 10,000 scientist</td>
<td>Beamline Guest Scientists</td>
</tr>
<tr>
<td>15 PB / year</td>
<td>10 – 100 PB / year</td>
</tr>
</tbody>
</table>

June 1, 2016, Frankfurt, Patrick Fuhrmann et al.
More storage at DESY

- The DESY 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

- 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

- 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**

Your Cloud Space

- Sync
- File up and download

The Scientific Data Cloud @ ownCloud Connects Business
Steps taken by DESY

• 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

Virtualization

User Management
Registry
LDAP

Authentication
Kerberos

Monitoring

Accounting

Local and Wide Area Network
Load Balancing
Firewalls

Unlimited Persistent Storage
Infrastructure Integration

F5, Load Balancer

Own Cloud

Own Cloud

Own Cloud

Own Cloud

Postgres DB

Automatic Failover
More Integration

Unlimited Central Storage

DESY LDAP

DESY Kerberos

vmware

Own Cloud

Data Life Cycle Engine

dCache.org
Horizontally Scaling Backend

Own Cloud
Own Cloud
Own Cloud
Own Cloud

Web Load Balancer (F5)

NFS 4.1 / pNFS

Pool Node
Pool Node
Pool Node
Pool Node

200 TBytes RAID 6
200 TBytes RAID 6
200 TBytes RAID 6

June 1, 2016, Frankfurt, Patrick Fuhrmann et al.
The Scientific Data Cloud @ ownCloud Connects Business
Some Statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Users Total</td>
<td>490</td>
</tr>
<tr>
<td>Users Active</td>
<td>277</td>
</tr>
<tr>
<td>Space Available</td>
<td>567 TBytes</td>
</tr>
<tr>
<td>Space Used</td>
<td>2 *30 TBytes</td>
</tr>
<tr>
<td>Files</td>
<td>10 Millions</td>
</tr>
</tbody>
</table>

Current Default Quality
Two Replicas on different storage nodes.

Files in/out per hour

Files in/out in 7 days

June 1, 2016, Frankfurt, Patrick Fuhrmann et al.

The Scientific Data Cloud @ ownCloud Connects Business
Is that sufficient for scientists?
Typical Workflow

Raw → Derived → Publication

Sharing
Data Categories

<table>
<thead>
<tr>
<th>Amount</th>
<th>Category</th>
<th>Typical Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 100 PB</td>
<td>Raw</td>
<td>LHC Detector data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Raw X-Ray Images</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Brain Scans</td>
</tr>
<tr>
<td>10 - 100 TB</td>
<td>Derived</td>
<td>Reconstructed (Ntuples)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Purified Images</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Brain Maps</td>
</tr>
<tr>
<td>1 TB</td>
<td>Publication</td>
<td>Papers, Presentations, Histograms</td>
</tr>
</tbody>
</table>
What do we need to support ‘science workflows’?
More Requirements

- **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

Derived

Publication

Long Term Preservation (Legal Requirement)

Low Latency (HPC, Analysis)

Fast, Multi Stream Access

SSD

SSD

SSD
Even more Requirements

• 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

High Speed Data Ingest

Fast Analysis
NFS 4.1/pNFS

Wide Area Transfers
(Globus Online, FTS)
by GridFTP

Sync’ing and Sharing with OwnCloud

June 1, 2016, Frankfurt, Patrick Fuhrmann et al.
The Scientific Data Cloud @ ownCloud Connects Business
What would that look like from the user’s perspective?
My DESY XXL Home
QoS support
My DESY XXL Home Protocol Support

Multi Protocol
NFS 4.1/pNFS
GridFTP
WebDAV
SRM

Sync
Share
Web 2.0
ownCloud

My ownCloud Home
How do we achieve those goals?

The scientific data cloud

OR

Choosing dCache as the storage backend for ownCloud!
Side Track

What’s dCache ?
dCache in a nutshell (cont.)

- 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
The Scientific Data Cloud @ ownCloud Connects Business
Namespace Design

Name Space

Physical Storage

Location Manager

Name

Disk 1
Disk 2
Tape 1

Disk

External System

Tape

June 1, 2016, Frankfurt, Patrick Fuhrmann et al.

The Scientific Data Cloud @ ownCloud Connects Business
Design Consequence

• 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

• 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

Web Service

CDMI Service

QoS Module

RESTful

dCache

Cloud

The Scientific Data Cloud @ ownCloud Connects Business
The QoS Web Interface

Click, to get File back from Tape.
Putting pieces together
The Data Path

Web Load Balancer (F5)

Own Cloud

Own Cloud

Own Cloud

Own Cloud

NFS 4.1 / pNFS

dCache

Spinning Disks

SSD’s

TAPE
Future Work
The Namespace Path

namespace, proxy

sharing db

dCache

namespace

ownCloud

namespace, proxy

sharing db
• 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

ownCloud GUI Web

QoS
Pluggin
(Server Side App)

I/O (NFS)

Namespace API

Share API

REST Services

QoS Module

June 1, 2016, Frankfurt, Patrick Fuhrmann et al.
Summary

• 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 pluggin (server side app) to make the dCache QoS storage types visible in ownCloud.
The END

further reading

www.dCache.org