dCache Scientific Cloud

Paul Millar

Workshop on Cloud Services for File Synchronisation and Sharing, CERN
Need a sync-n-share service at DESY

- **Hard requirements:**
  - Easy to use,
  - Store everything at DESY,
  - Integrate with existing infrastructure.

- **Anticipated future usage:**
  - change data between syncing and non-syncing storage,
  - share data without syncing,
  - 3rd party transfers between sites,
  - immediate access to sync space from compute facilities.
A scientific cloud vision

HPC & Grid Clusters
Low latency access

Cloud storage
Standard back-end for clusters and portals

Fast data ingest
Standard devices at high data rates

DropBox-like storage
Devices synchronise with storage

Bulk WAN transfer
Moving huge datasets

Remote access
Rich access via web-browser

NFS
CDMI
FTP
HTTP
WebDAV
How we solved it at DESY

• Looked around, chose two open-source projects:
  
  • **dCache**: powerful managed storage system
    Proven integration with scientific data life-cycle.
    Multiple production instances with tens-of-Petabyte capacity.
    “Hot” data can be stored on SSDs, “cold” on cheaper HDDs.
    Provides tape integration; automatic and transparent data migration.
    … but no sync and share facilities.
  
  • **ownCloud**: popular front-end
    Our collaborators adopting ownCloud makes it more attractive,
    … but assumes storage is managed.

• Combining these two gives DESY the best of both worlds:
  
  dCache is mounted on ownCloud server with **NFS v4.1/pNFS**.
  Integrated with DESY Kerberos, LDAP and Registry.
What is dCache?

LHC data stored on each storage system

- dCache (96 PB)
- DPM (34 PB)
- EOS (0 PB)
- StoRM (20 PB)
- CASTOR (14 PB)
- BeStMan (7.6 PB)
- Globus FTP (6.1 PB)
- ARC (0.01 PB)
- xrootd (22 PB)

Source: BDII (2014-11-14)
The DESY Cloud service

- Status: production, but for IT only;
  in two weeks “by invitation”
  (selected power-users);
  1st January general availability.
- Required minor patches to ownCloud & dCache.
- Changes pushed into dCache:
  - Running unpatched dCache release
  - Have a blueprint for any site to deploy ownCloud+dCache.
- Changes pushed upstream to ownCloud:
  - Not all were accepted for v7, so running patched version.
  - Will upgrade to v8 when released.
Development and future work

• Files in dCache have **user-ownership**, not ownCloud:
  Plan to expose files directly from dCache: NFS mount, 3rd party transfers, direct access from any grid worker-node, ...
  Couldn’t fix ownCloud: work-around within dCache

• **Consistency** between ACLs and shares:
  dCache ACLs to honour ownCloud shares and vice versa

• **Redirection** support for sync-client:
  ownCloud server proxying data is bottleneck; want syncing to be more efficient by taking data from where its stored.
NFS v4.1/PNFS vs ownCloud (currently)
ownCloud: currently vs with redirect
Experience: problems with ownCloud

- If underlying FS disappears, **all sync-clients delete all data**.
- If underlying FS returns **EIO** on read, sync-client creates 0-length file: **impossible to recover**.
- Bulk delete through web interface is **unreliable** (under investigation).
- Admin interface **awkward** with O(5k) users.
Not just ownCloud ...

- dCache team hosted a **two-day workshop** with project- and technical-lead of DCORE
  - Provides cloud storage with features beyond ownCloud
  - Some “big name” customers
- Initial “lite integration” by December
  (includes redirection support)
- Then providing “tight integration” with shared namespace
Thanks for listening … any questions?
Backup slides
Thinking about sync-and-share

• Like other systems, small fraction of data is “hot”
  SSDs provide better performance, but can't afford only SSDs; nice to have system that places hot data on SSDs, cold data on HDD.

• Amazon had a smart idea: allow people to choose how much to pay
  Let users choose between Normal and Glacial QoS; e.g., disable sync for Glacial-like storage but allow access via web interface
WLCG dCache instances (only WLCG sites shown)
DESY now a photon science lab
Over 10 years “Big Data” experience

<table>
<thead>
<tr>
<th>Era</th>
<th>Disk cache</th>
<th>Grid Storage</th>
<th>Generic Storage</th>
<th>Cloud Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hermes</td>
<td>ALICE</td>
<td>Fermilab</td>
<td>egi</td>
</tr>
<tr>
<td></td>
<td>ZEUS</td>
<td>ATLAS</td>
<td>Intensity Frontier</td>
<td>LSDMA</td>
</tr>
<tr>
<td></td>
<td>CDF</td>
<td>CMS</td>
<td>European XFEL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>IceCube</td>
<td>Belle II</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Industry</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CFEL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PETRA III</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SAML, OpenID, OAuth, Token, ...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Additional Communities</th>
<th>Trusted host</th>
<th>X.509, Kerberos</th>
<th>Username+PW</th>
<th>SAML, OpenID, OAuth, Token, ...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>