dCache: new and exciting features

Paul Millar, on behalf of the dCache Team

LSDMA „Technical Forum“ at KIT Campus Nord
2016-10-06
https://indico.desy.de/conferenceTimeTable.py?confId=15810
dCache server releases
... along with the series support durations.

Next version will be dCache v3.0
Why v3.0?

- Have to bump the number sooner or later.
- Better reflect backwards compatibility in mixed deployment,
- Many exciting new features,
  Optional – sites don’t have to use them
- Final analysis .. just because.
New in 3.0: CEPH integration

• With dCache v3.0, dCache has CEPH integration:
  • Can deploy a dCache pool that provides access to a CEPH pool.
  • dCache files are written as RBD images. Can be accessed directly (by PNFS-ID) outside of dCache
• All dCache features are available:
  Sites with tape integration may need to tweak their scripts
• Site driven functionality
HA-dCache: benefits

No Single Point of Failure:

Rolling updates:

Horizontal scaling:

Symmetric deployment:
HA-dCache: improvements #1

Topology discovery:

Redundant topologies:

Messaging:
HA-dCache: improvements #2

Redundant services:

Horizontally scalable SRM:

HA aware doors:
HA-dCache: status

• Everything available with dCache v3.0
  It's optional – existing behaviour is the default

• Deployed in production at NDGF
  • Running recent pre-release / snapshot of 3.0.0
  • Services in HA deployment; doors using HA-Proxy and uCARP.

• Deployment at DESY is planned.
  The DESY cloud – for the rolling updates.
DESY-Cloud update

- Proved an **excellent test** for dCache NFS
  
  No longer seeing any problems.

- Folding NFS changes back into **main-line dCache**:
  
  Only a few changes remaining.

- **Current stats**: 3900 shares, 670 users, 400 TB user data, 1.2x10^7 files.

- Currently operating with **ownCloud 9**
  
  In discussion with nextCloud.
REST API for dCache

- **New interface** for interacting with dCache
  - **HTTP** request/responses:
    - GET, PUT, DELETE, POST, PATCH …
  - **JSON** requests/responses
- **Modern standard approach** – supporting easy development of clients: JavaScript, CLIs, portals, …
- Initial support is for **namespace** and **Quality of Service management**, but ultimately allow all operations.
dCache-view

- A pure JavaScript, Web-2.0 client for dCache
- Uses the RESTful interface:
  - Demonstrates the power of the RESTful interface
- Browsing and download already supported.
- Upload and rename/move/delete coming soon.
## New web interface

Here is a screenshot of the new dCache web interface:

![New web interface](image)

### Table of Files

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Creation Time</th>
<th>Locality</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The Arctic Light.mp4</td>
<td>05/10/2016, 05:03:13</td>
<td></td>
<td>66.5 MB</td>
</tr>
<tr>
<td></td>
<td>The Mountain.mp4</td>
<td>05/10/2016, 05:03:14</td>
<td></td>
<td>62.8 MB</td>
</tr>
<tr>
<td></td>
<td>The Aurora.mp4</td>
<td>05/10/2016, 05:03:15</td>
<td></td>
<td>38.3 MB</td>
</tr>
</tbody>
</table>

[https://prometheus.desy.de:3880/](https://prometheus.desy.de:3880/)
New web interface

https://prometheus.desy.de:3880/
Increased support for federation

- **Hardening** dCache inter-domain communications
  - Encrypt tunnel communication,
  - Mutual authentication (X.509),
  - Only authorised hosts can connect.
- Will also be encrypting ZooKeeper communication
- Support dCache federations over untrusted WAN.
New resilient manager

- **Replaces** replica manager.
  
  Complete rewrite by Fermi team

- **New concept:**
  
  Focus on **event based**, rather than periodic scanning

- **Better integration** with other dCache components
  
  Takes events and information gathered by other components

- Being deployed at **Fermilab**, **DESY** and elsewhere.
Future directions

• **Integration** of nextCloud into dCache

• Adding **Samba support**
  
  We have windows users, after all.

• Adding **S3 support**

  The de facto standard for cloud storage.
Next dCache workshop: Umeå, Sweden

Co-located with NeIC 2017

Last Mon/Tue in May (2017-05-29, -30)
HA dCache: SRM

- **Split** the GSI “front-end” from “SRM engine”
- **Allow multiple front-ends:**
  - horizontal scaling for encryption overhead
- **Allow multiple “SRM engines”:**
  - each scheduled request is processed by the same SRM engine, load-balancing and fault-survival.
- **Support for HAProxy protocol**
  - using TCP mode, rather than HTTP mode.
Pencil sketch of possible deployment

- Clients
- HAProxy (live)
  - CARP: “shared” IP address
- HAProxy (hot spare)
- SRM front-end
- SRM back-end
- SRM front-end
- SRM back-end
- SRM back-end
- Rest of dCache

NB: works fine with just two nodes
HA dCache: general protocol remarks

- Should work fine for TLS-based protocols (SRM, gsiftp, webdav, gsidcap)
  - Load-balancer hostname as a Subject Alternate Name (SAN) in the X.509 certificate
- Possible to configure dCache so the SRM redirects clients to individual doors, rather than HA proxy:
  - SRM already provides load-balancing.
HA dCache: FTP

• Updated to understand HAProxy protocol
• IPv4 and IPv6 supported
• Data channels connect directly to pool or door, bypassing HAProxy.
HA dCache: other protocols

- **WebDAV**: nothing major needed
- **xrootd**: updated to understand HAProxy protocol. As usual “GSI-xrootd” sucks:
  - special care needed over x.509 certificate
  - kXR_locate returns IP address; makes host name verification hard
- **dcap**: updated to understand HAProxy protocol. No other major changes.
- **NFS**: not updated to support HA.
HA-dCache: status and next steps

- Currently deployed in production at NDGF
  Catching some bugs
- Presentations for admins at dCache workshop and “dCache Presents...” live webinar.
  Considerable interest expressed.
Other thoughts/issues

- Deleting file with target free capacity:
  feedback loop: delete until enough space is free
- Multiple concurrent uploads of the same file:
  ATLAS – multiple FTS, CMS – hidden error recovery
  SRM mostly protects us from this (apart from “FTS srmRm bug”)
  What is expected behaviour when not using SRM?
- RFC 4331 WebDAV quota support:
  Work started, anticipate being in dCache v3.0.
SRM reflections

- We (dCache.org) are NOT abandoning SRM:
  - We have invested heavily in cleaning- and speeding it up.
  - New client release, including `srmfs` an interactive SRM shell.
- It works – why replace a working system?
  
  By now the spec and implementations are well understood.
- Several unique features that would need to be re-implemented (e.g., see RFC-4331) – wasting effort.
- Biggest downside of SRM is NOT the protocol but the bindings; that can be fix.
- Certainly, declaring SRM dead is a self-fulfilling prophesy.