Project Status and Plans, new and further developments.

Tigran Mkrtchyan for dCache
28.05.2018, Hamburg
Thank you, Patrick

- Patrick involved into broader tasks.
- Tigran will take over the project.
- Be prepared for big changes! :-D
Meet the team

- Olufemi Adeyemi
- Paul Millar
- Tigran Mkrtchyan
- Marina Sahakyan
- Jürgen Starek
- Sibel Yasar
- Patrick Fuhrmann
- Dmitry Litvintsev
- Albert Rossi
- Vincent Garonne
The big picture

Data management & workflow control (Rucio, Kafka, SSE)

High Speed Data Ingest

Visualization & Sharing

Fast Analysis NFS 4.1/pNFS

Wide Area Transfers (Globus Online, FTS) by GridFTP, HTTP

dCache project status and plans | Tigran Mkrtchyan | 5
Strategic Direction

- WLCG
- DESY
  - XFEL
  - CTA
- SKA
- Long tail science
  - BioMed, Astro, Neutrino experiments
- Scalable storage
Main developments

- HA-enhancements
- Namespace scale-out
  - SQL strikes back
- Storage events
- Data protection
- Web admin and REST-api consolidation
- QoS
- Mutable files
- Integration into existing infrastructure
WLCG – the biggest customer

- Majority of sites are Tier1 and Tier2
- Grid protocols are high priority
- Support for new Storage Usage record
- Strong involvement into Data Lakes
  - More after Coffee break
FTP

- Pipelining
  - re-use of existing TCP connection for multiple transfers
- Better directory listing output
- Removed support for FTPTransaction.log
  - use .access file
- More compatibility with Globus-Online
XrootD

- Support for FUSE driver
- Support for persist-on-close
  - remove incomplete files
- Third-party-copy
  - work in progress
HA demand

- non WLCG instances require other SLA
- Minimize unscheduled downtimes
- Typical deployment must survive (multi)node failure
  - NDGF and DESY run HA-dcache in production
Resilience manager changes

• Automatic re-stage if all disk copies are offline.
• Replica distribution verification according to pool tags
• Finer control of available pools
  • read-only vs. allowed for replica
• Extra log file for failed activities
Beyond HEP

• All sites have non-HEP activities
  • including FNAL, NDGF and DESY
• On-site development
  • See Birgit’s and Onno’s presentations
• Share your experience
  • Workshop
  • User forum
  • Do we need an other channel?
NFS

- Better READDIR performance on big directories
- New export option -lt=
  - Control pNFS layout type
    - protocol used by clients to talk to pools
  - Use flexfile layout for new clients
    - RHEL 7.5 (3.10.0-862.el7)
    - kernel: 4.15
- populate access log

```plaintext
level=INFO ts=2018-04-24T17:36:03.643+0200 event=org.dcache.nfs.remove
user.mapped=500:500,491,499,500 socket.remote=131.169.185.213:829
obj.path=/exports/data/mdbench.ani.1221963
obj.id=00007A3DD95913BC49DCA7827864A6C656FF nfs.status=NFS4_OK
```
NFS (cont.)

• Added support for exports.d directory
  • aux exports file for dedicated group of hosts
  • better export file composition
  • can be added/removed dynamically

```bash
# ls -l exports exports.d/
-rw-r--r-- 1 root root 265 May 9 15:00 exports

exports.d/:
total 8
-rw-r--r-- 1 root root 65 May 9 15:00 flexfiles.exports
-rw-r--r-- 1 root root 48 May 24 18:08 lab.exports
```


WebDAV/HTTP

- “507 Insufficient Storage” on out of space
- Redirect to HTTPS
  
  `webdav.redirect.allow-https=false`

- requires host certificates on the pools
HTTP: Want-Digest

$ curl -I -H 'Want-Digest: adler32' http://dcache-lab000:8080/data/file1

HTTP/1.1 200 OK
Date: Thu, 24 May 2018 15:58:23 GMT
Server: dCache/4.2.0-SNAPSHOT
Accept-Ranges: bytes
ETag: "0000C21DD084254F49C6B7C9A5ED0CD11348_102126320"
Last-Modified: Wed, 29 Apr 2015 16:50:36 GMT
Digest: adler32=9c8b2d72
Content-Length: 1015

$
**gplazma-ldap**

- One-stop shop
  - No extra JAAS
  - No mutator
- Requires OpenJDK/OracleJDK (no IBM or Azul-Zing JMVs)

```bash
# gplazma.conf
auth   optional   ldap
map    optional   ldap
session optional   ldap
identity optional   ldap
```
Macaroons: the other cookies

- Contextual Caveats
- Decentralized Authorization
- Recipe by Paul tomorrow.
Internal communication

- Added TLS support for core/satellite domain communication
  
dcache.broker.core.client.channel.security
dcache.broker.satellite.channel.security

- Satellite domain client restriction
  
dcache.broker.core.allowed-subnets
Connection to external projects

- eXtreme Data Clouds (More details from Patrick)
  - WP2 - XFEL use case
  - WP4.1 – data life cycle
  - WP4.2 – smart caching (data lakes)
- EOSC Pilot
  - Inter-operability demonstrations
- RDA (Research Data Alliance)
  - QoS vocabulary standardization
- IETF – NFSv4 working group
- dCache H2020 cluster projects starting 2019
- HTW-Berlin
• Close work with HTW-Berlin
• Block course on Big Data and Open Source
  • Introduction to open-source project
  • Data management and challenges
• Students contribute code with small tasks
  • (1) and (2) from effort diagram (comes later)
SSH

• SSH key restriction
  • key can be used by desired user only
    ssh-rsa AAAAB...fPQ== tigran@localhost
  • can limit host name
    from="*.dcache.org,*desy.de" ssh-rsa AAAAB...fPQ== tigran@

• Write access log

  level=INFO ts=2018-05-11T13:16:30.888+0200
  event=org.dcache.services.ssh2.login username=admin
  (RSA SHA256:al9RPdXETJBEfFQsVqsIWXSFCofUCxzVjlkfXexBQ8)"
  successful=true
Monitoring and deployment
Storage Events

- A mechanism to trigger external actions
- Expose internal event to non dCache world
- First steps with SSE and Apache-Kafka
  - More details from Paul, Marina and Michael
REST-API

- Monitoring and Admin
  - Wait for Al's presentation!
- API documentation
  - try out in the browser
    - https://<frontend>:3880/api/v1/
  - generate clients
Packaging

- systemd for DEB packages
  - still fuzzy, distro specific
  - SL6 vs SL7
- Out-of-box docker container
- What is your preference?!
This-and-that

- Lot of cleanup with help of HTW-Berlin students
- Lot of internal re-factorings to enable various changes
- Update of external dependencies
- Lot of diagnostics into ZooKeeper-based components
Pool

- Support TLS for HTTP movers
- Report IO statistics to billing
  - read/write speed
  - number of bytes
- Calculate multiple checksums on transfer, if configured so.
- Added mongodb based backend to metadata
- Enhanced CEPH support
Effort guideline

<table>
<thead>
<tr>
<th>Impact</th>
<th>Difficult to do</th>
<th>Easy to do</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Minor</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Our domain

1. Implement immediately
2. Implement immediately
3. Long term plans and development
4. Discarded
Not with us any more

• Webadmin (‘The new one’)
  • admin pages in frontend service
• Replica manager
  • Resilience manager
• WebDAV’s html pages counting their days….
  • frontend service
GDPR
GDPR ready dCache

• Where to look
  • .access
  • billing files
  • billing DB
  • log files

• Rule of thumb
  • 7 days, if there are no operational needs

• Let us know if we can help
GDPR and dCache.org

- No public searchable mail archive
- No Google map, G+, Twitter and co.
- Request Tracker e-mail change
- No GDPR complaint way to delete accounts
  - But can be disabled (Art. 18)
- Possible implication – no email gateway.
  - All will need RT accounts
Did we miss your favorite feature?
Be active and enjoy the workshop!
Flexfiles pNFS layout

- Client-side data mirroring
  - client can write multiple copies if required
- Error reporting
  - Client tells door when communication with pool failed
- Client can be instructed never to fall-back to proxy IO
- Proposed IETF standard
  - NO RFC number yet