

What's new since 2016

Tigran Mkrtchyan for dCache Team
dCache User Workshop, Umeå, Sweden



INDIGO - DataCloud
Better Software for Better Science



HELMHOLTZ
| ASSOCIATION

What will be **NOT** covered with talks

- Changes in REST API
- New functionality in dCache-view
- CEPH
- HA-dCache
- Macarons and OpenID-connect
- New resilience manager

Version Number Model change

New major number **NOT**
compatible beyond latest
GOLDEN release

New Version Number Model

- Last release in the series is the golden one
 - 2.16 last in 2.x
 - New major version when pool compatibility brakes
 - e.g. 3.0 not compatible with < 2.16 , but with 2.16!
 - e.g. 4.0 will brake compatibility with 2.16
 - 5.0 will brake compatibility with latest 3.x
-

Random Bits and Pieces

- Sortable 'mover ls'
- No p2p limit
- Log why mover is killed
- SRM kills FTP transfer if SURL invalidated
- migration module supports -target=hsm
- 3rd party webdav copy supports openid-connect
- Per-protocol stage protection

Stage protection example

```
# all auth allowed
```

```
".+"
```

```
# all but nova can stage
```

```
"" "" "!nova.*"
```

```
# nova can't stage with NFS
```

```
"" "" "nova.*"    "!NFS4.*"
```

Support for idmap DNS record

- Automatic nfs4domain configuration
- Uses DNS record
 - `'_nfsv4idmapdomain IN TXT "foo.bar"'`
- Supported by Linux and Solaris

No level2 any more!

- historical behavior final gone!
- can be enabled with option
 - `pool.enable.hsm-flag` (h=yes | h=no)
- file size does not participate in 'isNew'
 - file size can be set prior upload

Transparent transfer migration for NFS

- `pool disable` triggers re-call IO layout from the client
- client will request a new transfer
- read requests will proceed when/if data available
- write request will fail with IO error
- other protocols may follow.

Possibility to get/set storage URI

- User side possibility to query HSM location
- Admin can append, update or remove HSM location
- Don't mess with it!



Example: GET/SET

```
# cat '(suri)(MINBIAS.root)'
```

```
osm://osm/?bf=1
```

```
# echo "osm://osm/?bf=2" > \
```

```
'(suri)(MINBIAS.root)'
```

```
# cat '(suri)(MINBIAS.root)'
```

```
osm://osm/?bf=2
```

Example: UPDATE/DELETE

```
# echo "osm://osm/?bf=3" >> \
```

```
'.(suri)(MINBIAS.root)'
```

```
# cat '.(suri)(MINBIAS.root)'
```

```
osm://osm/?bf=2
```

```
osm://osm/?bf=3
```

```
# echo "" > '.(suri)(MINBIAS.root)'
```

```
# cat '.(suri)(MINBIAS.root)'
```

```
#
```

Compatibility with GlobusOnline

- Fixed data-passive adapter keep-alive
- Dynamic checksum calculation
- BNL - Argon data transfers succeeded with verification
- Tested by BNL and PIC
- Planned to become part of 3.1 and 3.0 (2.16 if needed)

But...(Quotes from GlobusOnline):

“Globus is aggressively moving beyond GridFTP”

“...working on a major overhaul to Globus Connect Server ... to add a bunch of capabilities through HTTP and REST APIs.”

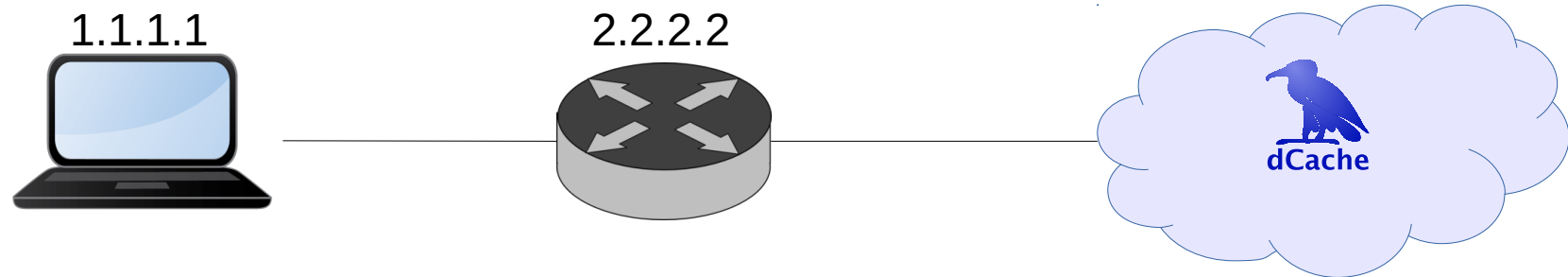
“Globus services will require more than just GridFTP”

And.....

“We currently have no funding for implementing, validating, or supporting dCache. We have no funding for the up-front work, and currently none of the dCache sites are subscribers of the Globus service.”

Self describing billing files

- Do know all fields in the billing file?
- Do know custom formats?
 - well, you know that you can specify format starting from 2.1?!
- Each file started with header defining format
- Each header line prefixed with **##**



```
05.22 00:00:44 [door:GFTP-...:request] \  
  ["/O=GermanGrid/...":21065:5296:2.2.2.2] \  
    [000094...680,16777216] [/pnfs/...] x:y@osm 163 0 {0:""}
```

HA-proxy protocol

- De-facto standard
 - HA-proxy, Jetty, stunnel, Squid
 - Amazon ELB
- SRM, FTP and xrootd, webdav understand HA-Proxy protocol
- Billing files contains client real IP and not proxy node

HSM polling

- HSM script provider can poll for the result
- Flush script will be re-invoked as long as it returns error core **72**
- Error not propagated to dCache
 - no entries in billing files
 - not entries in the log files

Continues flushing to tape

- current flush works in batches
- no new flush bunch is submitted until all files are flushed
- new option allows to add new request when there is a room available.

Dynamic pool queues

- Pool queues can be created/deleted on the flight
- no pool restart for new 'control group'
 - doors still have to be restarted
- `dcap -X io-queue=<queue>`
- `xrootd` uses application name queue.
- 'regular' is taken, if missing

Poolmanager.conf is a copy!

- configuration stored in zookeeper
- initial load from local file
- save will bump config to local disk
 - in HA-setup load from one of the nodes
 - in HA-setup saves locally on connected PoolManager host

Xrootd - FUSE



National Research Centre (NRC)
"Kurchatov Institute"



Big Data Technologies Laboratory
<http://bigdatalab.nrcki.ru/>

First experience with dCache

- Well-known and reliable software used on many T1s
- Different software platform (Java) and protocol implementations
 - dCache xrootd implementation does not support FUSE mounts
 - This is supposed to be fixed within a collaborate project between NRC "KI", JINR and DESY
- No built-in security for control channel between Manager and File servers
 - Firewall-based access control works well for a single site
 - Requires VPN or some other external security measures for Federated scenario
 - There's also an ongoing project to fix that
- No built-in Manager redundancy
 - But you can replicate the database

Xrootd - FUSE

- Ivan Kadochnikov (JINR, Dubna)
 - Join project with Russian Federated Data Storage Project
 - Multiple fixes in dCache, xrootd-fuse
 - dCache and xrootd have to release fixes!
-

Resilience Manager

- Successor of Replica manager
 - New design
 - No own database
 - New rules
 - storage group based
 - replicates ONLINE files
 - New possibilities
 - Files can be flushed to HSM!
 - (CUSTODIAL+ONLINE)
-