# ATLAS WebDAV status and plans

Pre-GDB (data access)

Cédric Serfon[1], Vincent Garonne[1], Sylvain Blunier[2]

[1]CERN, PH-ADP-CO, [2] Pontificia Univ. Catolica de Chile

May 13th 2014

## Outline

- Introduction
- Rucio redirector
- Davix, performance tests
- Metalinks
- Conclusion

# Reminders

- ATLAS is currently migrating to Rucio.
- This migration to Rucio induces number of changes, in particular :
    - Drop the use of LFC.
    - Rely on a deterministic path PFN = f(LFN).
    - Support of multiple Storage protocols (SRM, S3, XROOTD, WedDAV...)
- To move to deterministic path, a renaming campaign was run on ∼300M files using WebDAV. It finished successfully in February.

# Renaming

- Sites were required to provide a WebDAV access for the renaming. Now 67 sites (dCache, DPM, StoRM, EOS) have WebDAV configured.
- Many bugs/performance issues found during this campaign and provided feedbacks to storage providers/sites.
- It allowed us to gain experience.
- Now that WebDAV is avalaible on most of the sites, we can start using it for data access, in particular for user data access.

# Why WebDAV ?

- WebDAV provides all (but staging) functionalities that we need to interact with our data (upload, download, delete...)
- FTS also supports it for third party transfers.

→ One of the natural candidate to replace SRM.

- It's a standard protocol and a lot of clients already exist on the market (Cadaver, Cyberduck...).
- It natively supports redirection.

# WebDAV and Rucio

- The migration to Rucio is being done in 3 steps transparent for the end-users :
  - Renaming all files according to the new convention. DONE.
  - Moving the files/replicas from the LFCs to Rucio. 70% of the clouds DONE.
  - Moving the DQ2 objects (datasets, containers, subscriptions...) to Rucio. TOBEDONE.
- Once the 2nd step is done for one site, we can start using WebDAV to access the files on it. Right now :
  - 62 sites representing 329 endpoints potentially accessible via WebDAV.
  - It represents 290M files / 52 PB.
- Rucio provides 2 ways to use WebDAV :
  - Rucio redirector.
  - Metalinks

# Rucio redirector

- Secure REST API call with 302 redirection :

  ```
  GET /redirect/{file_scope}/{file_name}
  ```

- The Rucio server queries the replica table and redirects the query to a selected replica URL.
- The strategy to select the final replica is configurable:
  - random (default).
  - geoip selection based on GeoLite DB chooses the closest replica (IPv4/6 compliant).
  - selected site.

```
GET /redirect/{scope}/{name}?select=geoip|random
GET /redirect/{scope}/{name}?rse={rse_name}
```

# Rucio redirector - Example

```
# curl -LI --capath /etc/grid-security/certificates/ --cacert $X509_USER_PROXY --cert $X509_USER_PROXY
https://voatlasrucio-redirect-prod-01.cern.ch/redirect/mc12_8TeV/NTUP_SMWZ.01330897._000001.root.1

HTTP/1.1 302 Found
Date: Thu, 08 May 2014 11:04:28 GMT
Server: Apache/2.2.15 (Red Hat)
Location: https://grid05.lal.in2p3.fr:443/dpm/lal.in2p3.fr/home/atlas/atlasdatadisk/rucio/mc12_8TeV/9c/be/NTUP_SMWZ.01330897._000001
Content-Type: text/html; charset=UTF-8

HTTP/1.1 302 Found
Date: Thu, 08 May 2014 11:04:28 GMT
Server: Apache/2.2.15 (Scientific Linux)
Link: <https://grid05.lal.in2p3.fr/dpm/lal.in2p3.fr/home/atlas/atlasdatadisk/rucio/mc12_8TeV/9c/be/NTUP_SMWZ.01330897._000001.root.1
Location: https://grid40.lal.in2p3.fr/dpmpart/part1/atlas/2014-05-01/NTUP_SMWZ.01330897._000001.root.1.137750292.0?token=6Di2OwWQWE%
Vary: Accept-Encoding
Content-Type: text/html; charset=iso-8859-1

HTTP/1.1 200 OK
Date: Thu, 08 May 2014 11:04:28 GMT
Server: Apache/2.2.15 (Scientific Linux)
Content-Length: 1547866614
Content-Disposition: filename="NTUP_SMWZ.01330897._000001.root.1"
Accept-Ranges: bytes
Content-Type: application/x-troff-man
```

- First redirection by the Rucio redirector to the DPM head-node (in that case GRIF-LAL).
- Second redirection by the DPM head-node to the disk node.

# Rucio redirector - Example

- Download :

```
# curl -L -O --capath /etc/grid-security/certificates/ --cacert $X509_USER_PROXY --cert $X509_USER_PROXY
https://voatlasrucio-redirect-prod-01.cern.ch/redirect/mc12_8TeV/NTUP_SMWZ.01330897._000001.root.1?rse=GRIF-LAL_DATADISK
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 1476M  100 1476M    0     0  34.9M      0  0:00:42  0:00:42 --:--:-- 41.8M
```

# Tests with Davix

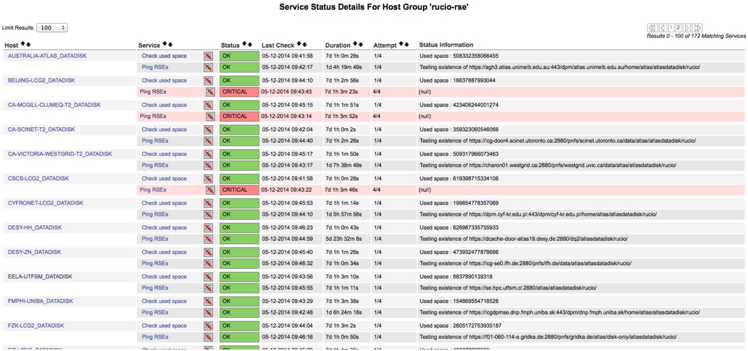
- Davix is a "lightweight toolkit for file access and file management with HTTP Based protocols" developed by CERN IT.
- It's being integrated to ROOT :

```
root [0] TFile *f = TFile::Open("https://voatlasrucio-redirect-prod-01.cern.ch/redirect/mc12_8TeV/NTUP_SMWZ.01330897._000001.root.1"
(class TFile *) 0x24ffd70
root [1] TTree *t = (TTree*)f->Get("physics")
(class TTree *) 0x294f860
root [2] t->GetEntries()
(Long64_t) 10000
```

- Tests are being conducted (Sylvain Blunier) :
  - ROOT 5.34.18 has been compiled with Davix support and deployed on CVMFS.
  - Started some prun tests that read data from the LAN/WAN via WebDAV.
  - Comparision of number of events processed per seconds.

# Validation of endpoints

- From the 62 sites with WebDAV and served by Rucio 18 are failing a simple HEAD : We put in place a nagios probe to identify these sites and they are not considered in the study.



**Service Status Details For Host Group 'rucio-rse'**

Limit Results: 100

Results 0 - 100 of 172 Matching Services

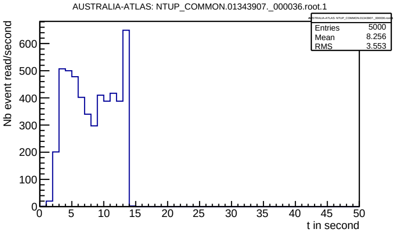
| Host | Service | Status | Last Check | Duration | Attempt | Status Information |
|---|---|---|---|---|---|---|
| AUSTRALIA-ATLAS_DATADISK | Check used space | OK | 05-12-2014 09:41:58 | 7d 1h 0m 28s | 1/4 | Used space : 508332358066455 |
|  | Ping RSEs | OK | 05-12-2014 09:42:17 | 1d 4h 19m 49s | 1/4 | Testing existence of https://agh3.atlas.unimelb.edu.au:443/dpm/atlas.unimelb.edu.au/home/atlas/atlasdatadisk/rucio/ |
| BEIJING-LCG2_DATADISK | Check used space | OK | 05-12-2014 09:44:10 | 7d 1h 2m 56s | 1/4 | Used space : 16637887993044 |
|  | Ping RSEs | CRITICAL | 05-12-2014 09:43:43 | 7d 1h 3m 23s | 4/4 | (null) |
| CA-MCGILL-CLUMEQ-T2_DATADISK | Check used space | OK | 05-12-2014 09:45:15 | 7d 1h 1m 51s | 1/4 | Used space : 423406244001274 |
|  | Ping RSEs | CRITICAL | 05-12-2014 09:43:14 | 7d 1h 3m 52s | 4/4 | (null) |
| CA-SCINET-T2_DATADISK | Check used space | OK | 05-12-2014 09:42:04 | 7d 1h 0m 2s | 1/4 | Used space : 359323060546066 |
|  | Ping RSEs | OK | 05-12-2014 09:44:40 | 7d 1h 2m 26s | 1/4 | Testing existence of https://lcg-door4.scinet.utoronto.ca:2880/pnfs/scinet.utoronto.ca/data/atlas/atlasdatadisk/rucio/ |
| CA-VICTORIA-WESTGRID-T2_DATADISK | Check used space | OK | 05-12-2014 09:45:17 | 7d 1h 1m 50s | 1/4 | Used space : 509317966073463 |
|  | Ping RSEs | OK | 05-12-2014 09:43:17 | 1d 7h 38m 49s | 1/4 | Testing existence of https://charon01.westgrid.ca:2880/pnfs/westgrid.uvic.ca/data/atlas/atlasdatadisk/rucio/ |
| CSCS-LCG2_DATADISK | Check used space | OK | 05-12-2014 09:41:58 | 7d 1h 0m 26s | 1/4 | Used space : 619398715334106 |
|  | Ping RSEs | CRITICAL | 05-12-2014 09:43:22 | 7d 1h 3m 46s | 4/4 | (null) |
| CYFRONET-LCG2_DATADISK | Check used space | OK | 05-12-2014 09:45:53 | 7d 1h 1m 14s | 1/4 | Used space : 199854778357069 |
|  | Ping RSEs | OK | 05-12-2014 09:44:10 | 1d 5h 57m 56s | 1/4 | Testing existence of https://dpm.cyf-kr.edu.pl:443/dpm/cyf-kr.edu.pl/home/atlas/atlasdatadisk/rucio/ |
| DESY-HH_DATADISK | Check used space | OK | 05-12-2014 09:46:23 | 7d 1h 0m 43s | 1/4 | Used space : 626987335755933 |
|  | Ping RSEs | OK | 05-12-2014 09:44:59 | 5d 23h 32m 8s | 1/4 | Testing existence of https://dcache-door-atlas19.desy.de:2880/dq2/atlasdatadisk/rucio/ |
| DESY-ZN_DATADISK | Check used space | OK | 05-12-2014 09:45:40 | 7d 1h 1m 26s | 1/4 | Used space : 473932477878686 |
|  | Ping RSEs | OK | 05-12-2014 09:46:32 | 7d 1h 0m 34s | 1/4 | Testing existence of https://lcg-se0.ifh.de:2880/pnfs/ifh.de/data/atlas/atlasdatadisk/rucio/ |
| EELA-UTFSM_DATADISK | Check used space | OK | 05-12-2014 09:43:56 | 7d 1h 3m 10s | 1/4 | Used space : 8837890139318 |
|  | Ping RSEs | OK | 05-12-2014 09:45:55 | 7d 1h 1m 11s | 1/4 | Testing existence of https://se.hpc.utfsm.cl:2880/atlas/atlasdatadisk/rucio/ |
| FMPHI-UNIBA_DATADISK | Check used space | OK | 05-12-2014 09:43:29 | 7d 1h 3m 38s | 1/4 | Used space : 154869554718526 |
|  | Ping RSEs | OK | 05-12-2014 09:42:48 | 1d 6h 24m 18s | 1/4 | Testing existence of https://lcgdpmse.dnp.fmph.uniba.sk:443/dpm/dnp.fmph.uniba.sk/home/atlas/atlasdatadisk/rucio/ |
| FZK-LCG2_DATADISK | Check used space | OK | 05-12-2014 09:44:04 | 7d 1h 3m 2s | 1/4 | Used space : 2605172753935187 |
|  | Ping RSEs | OK | 05-12-2014 09:46:16 | 7d 1h 0m 50s | 1/4 | Testing existence of https://f01-060-114-e.gridka.de:2880/pnfs/gridka.de/atlas/disk-only/atlasdatadisk/rucio/ |

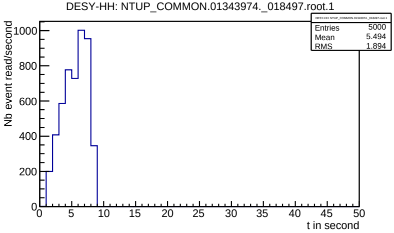
# Tests and first results

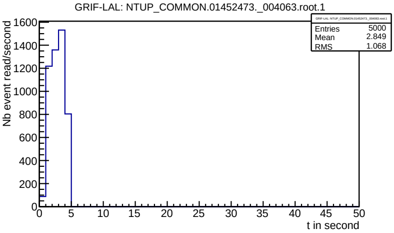
- First preliminary results of tests with prun (more expected for ATLAS Software week).
- Compilation problems on ∼50/135 ANALY queues under investigation.
- In most of the sites where the compilation succeeded, problem with the proxy used on the WNs :GGUS:105188.
- On 11 sites (RAL and some US sites), the jobs compiled and ran successfully.

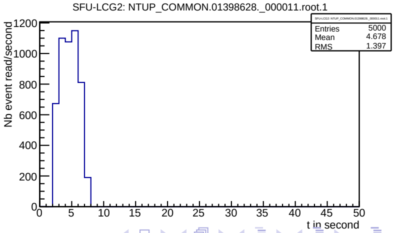
# Tests and first results

- Job running in AGLT2 and accessing file over the WAN :



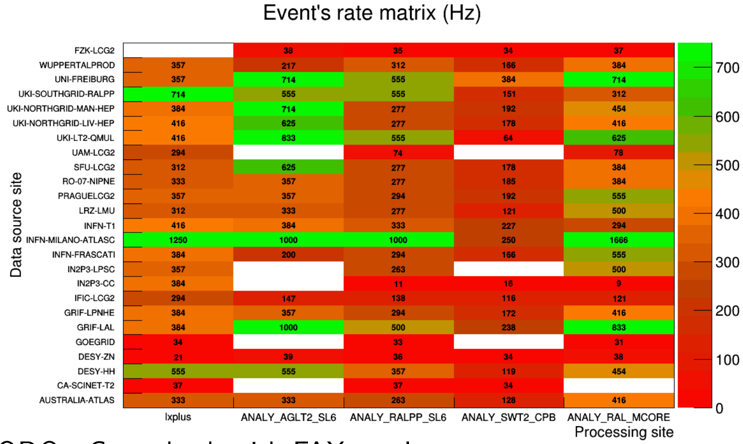






# Event rate matrix

- Disclaimer : Very preliminary with limited statistic and limited number of sites :



| | lxplus | ANALY_AGLT2_SL6 | ANALY_RALPP_SL6 | ANALY_SWT2_CPB | ANALY_RAL_MCORE |
|---|---|---|---|---|---|
| FZK-LCG2 | | 38 | 35 | 34 | 37 |
| WUPPERTALPROD | 357 | 217 | 312 | 166 | 384 |
| UNI-FREIBURG | 357 | 714 | 555 | 384 | 714 |
| UKI-SOUTHGRID-RALPP | 714 | 555 | 555 | 151 | 312 |
| UKI-NORTHGRID-MAN-HEP | 384 | 714 | 277 | 192 | 454 |
| UKI-NORTHGRID-LIV-HEP | 416 | 625 | 277 | 178 | 416 |
| UKI-LT2-QMUL | 416 | 833 | 555 | 64 | 625 |
| UAM-LCG2 | 294 | | 74 | | 78 |
| SFU-LCG2 | 312 | 625 | 277 | 178 | 384 |
| RO-07-NIPNE | 333 | 357 | 277 | 185 | 384 |
| PRAGUELCG2 | 357 | 357 | 294 | 192 | 555 |
| LRZ-LMU | 312 | 333 | 277 | 121 | 500 |
| INFN-T1 | 416 | 384 | 333 | 227 | 294 |
| INFN-MILANO-ATLASC | 1250 | 1000 | 1000 | 250 | 1666 |
| INFN-FRASCATI | 384 | 200 | 294 | 166 | 555 |
| IN2P3-LPSC | 357 | | 263 | | 500 |
| IN2P3-CC | 384 | | 11 | 16 | 9 |
| IFIC-LCG2 | 294 | 147 | 138 | 116 | 121 |
| GRIF-LPNHE | 384 | 357 | 294 | 172 | 416 |
| GRIF-LAL | 384 | 1000 | 500 | 238 | 833 |
| GOEGRID | 34 | | 33 | | 31 |
| DESY-ZN | 21 | 39 | 36 | 34 | 38 |
| DESY-HH | 555 | 555 | 357 | 119 | 454 |
| CA-SCINET-T2 | 37 | | 37 | 34 | |
| AUSTRALIA-ATLAS | 333 | 333 | 263 | 128 | 416 |

- TODO : Crosscheck with FAX results.

# Metalinks

- "Metalinks is an extensible metadata file format that describes one or more computer files available for download."
- It allows failovers, multisources.
- In Rucio all the file replicas for file(s) are listed through a Rucio RESTful API and the client can request an answer in metalink(3/4) format

  ```
  GET /replicas/{file_scope}/{file_name}
  application/metalink+xml
  application/metalink4+xml
  ```

- Bulk method is available

# Metalinks example

```
# export token=‘curl -i -H "X-Rucio-Account: ddmusr01" --cacert $X509_USER_PROXY --cert $X509_USER_PROXY
--capath /etc/grid-security/certificates/ -X GET https://voatlasrucio-auth-prod.cern.ch/auth/x509_proxy | grep X-Rucio-Auth-Token‘

# curl -s -H "$token" -H ’Accept: application/metalink4+xml’  --cacert /etc/pki/tls/certs/CERN-bundle.pem
https://voatlasrucio-server-prod-04.cern.ch/replicas/mc12_8TeV/NTUP_SMWZ.01330897._000001.root.1?select=geoip

<?xml version="1.0" encoding="UTF-8"?>
<metalink xmlns="urn:ietf:params:xml:ns:metalink">
<files>
 <file name="NTUP_SMWZ.01330897._000001.root.1">
  <identity>mc12_8TeV:NTUP_SMWZ.01330897._000001.root.1</identity>
  <hash type="adler32">8b93d074</hash>
  <size>1547866614</size>
   <url location="GRIF-LAL_DATADISK" priority="0">https://grid05.lal.in2p3.fr:443/dpm/lal.in2p3.fr/home/atlas/atlasdatadisk/rucio/
                 mc12_8TeV/9c/be/NTUP_SMWZ.01330897._000001.root.1</url>
   <url location="SARA-MATRIX_DATADISK" priority="1">https://bee34.grid.sara.nl:2882/pnfs/grid.sara.nl/data/atlas/atlasdatadisk/
                 rucio/mc12_8TeV/9c/be/NTUP_SMWZ.01330897._000001.root.1</url>
   <url location="UKI-SOUTHGRID-CAM-HEP_DATADISK" priority="2">https://serv02.hep.phy.cam.ac.uk:443/dpm/hep.phy.cam.ac.uk/home/
                 atlas/atlasdatadisk/rucio/mc12_8TeV/9c/be/NTUP_SMWZ.01330897._000001.root.1</url>
 </file>
</files>
</metalink>
```

- The metalinks can be used by clients like aria2c.

# Advantages of using Rucio+WebDAV

- The only thing that the site needs to configure is WebDAV.
- Rucio is aware of the sites' downtimes and can exclude the replicas on sites down.
- Rucio knows exactly where to find the replicas and doesn't need to check on N different locations/space tokens.
- We can use common HTTP tools and technics (e.g. to prevent DoS, HTTP caching). Effort needed for HTTP**S** proxy.
- New selection strategy can be easily implemented (e.g. based on cost matrix) and deployed (just need to update the server in one place).
- The metalink approach looks better than the redirector since it allows failover if one replica is unavailable.

# Conclusion

- WebDAV is a interesting protocol that can replace SRM. It is already used for Rucio functional tests to injects/delete files.
- It's already available on more than half of the ATLAS sites, but not the same QoS than SRM yet (no SAM tests). As we are starting using it for user data download, we request a production quality service service similar to SRM, gridFTP on the sites where it's deployed.
- Rucio provide 2 ways to use WebDAV : Rucio redirector and Metalink server.
- Tests are being conducted for LAN/WAN accesses :
  - Plan to do some Hammercloud tests with Rucio redirector and extract some performances plots as for FAX.
  - Want to try root + Davix + Metalinks (should be available in a future release).
- Work needed on monitoring.