# WEBDAV EXPERIENCE FOR WAN AND LAN DATA ACCESS

Johannes Elmsheuser
on behalf of several

Ludwig-Maximilians-Universität München

4 December 2014/ATLAS Facilities Jamboree

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

# INTRODUCTION

Fast and reliable access in reasonably long jobs essential for analysis jobs

## Panda configuration:

- Input data access mode is configured on a site-by-site basis in Panda
- DPM usually copy-to-scratch or recently some sites use xrootd
- dCache uses usually dcap

→ Explore webdav input access:

- Industry standard
- Same client and access protocol everywhere
- Use aria2c for stage-in
- Use Davix plugin for ROOT I/O

# PANDA DATA ACCESS

2 access modes in analysis jobs:

- First local stage-in of analysis source code libds dataset from previous build jobs → aria2c
- Then ROOT/Athena event loop with local (or remote) input ROOT file access → Davix

Setup:

- aria2c stage-in has been used at a couple of DE sites for production for some time
- Vincenzo Lavorini has improved the Panda pilot aria2c site mover in the last months
- Davix is a ROOT plugin/library developed in CERN IT/SDC - Adrien Devresse lead developer
- Sylvain Blunier extensively tested and built early version of Davix/ROOT and tested sites
- Rucio provides a redirector with very easy syntax
- Rod, Günter, Cedric, Paul technical consultant/support

# WEBDAV CONFIG AND ARIA2C STATUS

Rucio/Site webdav:

- Redirection, webdav configured in AGIS at:
  `http://atlas-agis.cern.ch/agis/service/table_view/`
  `?type=SE/HTTP&state=ACTIVE`
- Is this list up-to-date and every listed site actually works as expected ?
- Functional test against webdav door somewhere ?

aria2c status:

- Extensive HammerCloud testing at 40-50 sites, reusing dedicated FAX datasets - using Panda pilot overwitequeuedata option to switch from defaults
- Panda stage-in works at most webdav/AGIS configured sites
- Not sure if authentication problem is fixed everywhere ?
- Tedious remaining site by site checking
- Future: aria2c stage-out (?)

# DAVIX/HTTP - STATUS I

- ROOT plugin for Davix (TDavixFile) allows to use Davix for HTTP, WebDav or S3 I/O
- Use it with: `TFile::Open("https://myse/scope/myfile.root")`
- Rucio redirector allows together with Davix to redirect to any webdav configured ATLAS SE
- Example:

  `https://rucio-lb-prod.cern.ch/redirect/mc14_8TeV/AOD.01507240._010001.pool.root.2?site=LRZ-LMU`

- Options to URL: `?site=LRZ-LMU`, `?rse=LRZ-LMU_LOCALGROUPDISK`, `?geoip` to select best replica
- Currently in use in Panda pilot: `?site=LRZ-LMU`
- Future:
  - option for metalink usage - this is used in the Panda aria2c sitemover - allows multiple streams from different sites.

# DAVIX/HTTP - STATUS II

Builds:

- Since a few weeks davix 0.3.6 available together with ROOT 5.34.22-x86_64-slc6-gcc48-opt and later in CVMFS, enabled by default in system.rootrc
- Now possible to use full ASG AnalysisBase releases with davix
- Before only older customized builds - found/fixed several bugs in several release iterations with davix developer over longer period

# HAMMERCLOUD TEST I

`http://hammercloud.cern.ch/hc/app/atlas/test/20044842/`
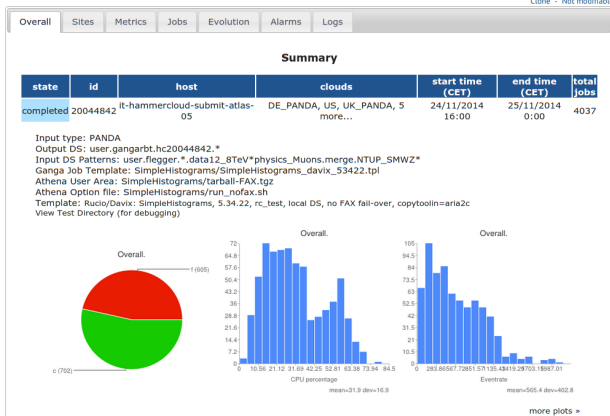


Only functional test and no read speed since rc_test pilot was used here

# HAMMERCLOUD TEST II - FUNCTIONAL SITES

http://hammercloud.cern.ch/hc/app/atlas/test/20044842/

| Site | S | R | C | F | Eff | T | Datasets | Queue | Max R | Resubmit | R. Force | Link |
|------|---|---|---|---|-----|---|----------|-------|-------|----------|----------|------|
| ANALY_GLASGOW_SL6 | 0 | 4 | 130 | 7 | 0.95 | 143 | 1 | 0 | 1 | yes | no | » |
| ANALY_LIV_SL6 | 389 | 0 | 102 | 38 | 0.73 | 529 | 1 | 0 | 1 | yes | no | » |
| ANALY_PIC_SL6 | 243 | 0 | 69 | 0 | 1.00 | 314 | 1 | 0 | 1 | yes | no | » |
| ANALY_AGLT2_SL6 | 7 | 8 | 63 | 2 | 0.97 | 80 | 1 | 0 | 1 | yes | no | » |
| ANALY_DESY-HH | 9 | 4 | 59 | 7 | 0.89 | 82 | 1 | 0 | 1 | yes | no | » |
| ANALY_FZU | 289 | 4 | 55 | 51 | 0.52 | 403 | 1 | 0 | 1 | yes | no | » |
| ANALY_IFAE | 0 | 0 | 41 | 59 | 0.41 | 100 | 1 | 0 | 1 | no | no | » |
| ANALY_FZK | 29 | 3 | 33 | 15 | 0.69 | 80 | 1 | 0 | 1 | yes | no | » |
| ANALY_JINR | 14 | 0 | 31 | 0 | 1.00 | 45 | 1 | 0 | 1 | yes | no | » |
| ANALY_MANC_SL6 | 5 | 0 | 25 | 28 | 0.47 | 58 | 1 | 0 | 1 | yes | no | » |
| ANALY_LRZ | 48 | 3 | 20 | 0 | 1.00 | 71 | 1 | 0 | 1 | yes | no | » |
| ANALY_CAM_SL6 | 181 | 1 | 20 | 1 | 0.95 | 204 | 1 | 0 | 1 | yes | no | » |
| ANALY_wuppertalprod | 22 | 0 | 13 | 13 | 0.50 | 48 | 1 | 0 | 1 | yes | no | » |
| ANALY_CPPM | 113 | 4 | 13 | 3 | 0.81 | 133 | 1 | 0 | 1 | yes | no | » |
| ANALY_LANCS_SL6 | 142 | 1 | 11 | 16 | 0.41 | 171 | 1 | 0 | 1 | yes | no | » |
| ANALY_LAPP | 65 | 4 | 8 | 11 | 0.42 | 88 | 1 | 0 | 1 | yes | no | » |
| ANALY_RALPP_SL6 | 0 | 31 | 6 | 7 | 0.46 | 45 | 1 | 0 | 1 | yes | no | » |
| ANALY_DESY-ZN | 0 | 3 | 2 | 24 | 0.08 | 29 | 1 | 0 | 1 | yes | no | » |
| ANALY_BNL_SHORT | 0 | 11 | 1 | 0 | 1.00 | 12 | 1 | 0 | 1 | yes | no | » |
| ANALY_SLAC | 24 | 8 | 0 | 62 | 0.00 | 95 | 1 | 0 | 1 | no | no | » |
| ANALY_MWT2_SL6 | 0 | 0 | 0 | 65 | 0.00 | 65 | 1 | 0 | 1 | no | no | » |
| ANALY_MPPMU | 0 | 0 | 0 | 0 | 0.00 | 0 | 1 | 0 | 1 | no | no | » |
| ANALY_OU_OCHEP_SWT2 | 78 | 5 | 0 | 0 | 0.00 | 85 | 1 | 0 | 1 | yes | no | » |
| ANALY OX SL6 | 139 | 0 | 0 | 6 | 0.00 | 145 | 1 | 0 | 1 | yes | no | » |

Not all sites actually ran jobs in this too short rc_test pilot test

# HAMMERCLOUD TEST III - NON FUNCTIONAL SITES

http://hammercloud.cern.ch/hc/app/atlas/test/20044842/

| Site | S | R | C | F | Eff | T | Datasets | Queue | Max R | Resubmit | R. Force | Link |
|------|---|---|---|---|-----|---|----------|-------|-------|----------|----------|------|
| ANALY_MWT2_SL6 | 0 | 0 | 0 | 65 | 0.00 | 65 | 1 | 0 | 1 | no | no | » |
| ANALY_SLAC | 24 | 8 | 0 | 62 | 0.00 | 95 | 1 | 0 | 1 | no | no | » |
| ANALY_IFAE | 0 | 0 | 41 | 59 | 0.41 | 100 | 1 | 0 | 1 | no | no | » |
| ANALY_GOEGRID | 0 | 3 | 0 | 51 | 0.00 | 54 | 1 | 0 | 1 | yes | no | » |
| ANALY_FZU | 289 | 4 | 55 | 51 | 0.52 | 403 | 1 | 0 | 1 | yes | no | » |
| ANALY_LIV_SL6 | 389 | 0 | 102 | 38 | 0.73 | 529 | 1 | 0 | 1 | yes | no | » |
| ANALY_CERN_SLC6 | 0 | 0 | 0 | 38 | 0.00 | 38 | 1 | 0 | 1 | no | no | » |
| ANALY_QMUL_SL6 | 0 | 0 | 0 | 33 | 0.00 | 33 | 1 | 0 | 1 | no | no | » |
| ANALY_MANC_SL6 | 5 | 0 | 25 | 28 | 0.47 | 58 | 1 | 0 | 1 | yes | no | » |
| ANALY_DESY-ZN | 0 | 3 | 2 | 24 | 0.08 | 29 | 1 | 0 | 1 | yes | no | » |
| ANALY_INFN-FRASCATI | 12 | 0 | 0 | 23 | 0.00 | 35 | 1 | 0 | 1 | yes | no | » |
| ANALY_SWT2_CPB | 13 | 2 | 0 | 18 | 0.00 | 33 | 1 | 0 | 1 | yes | no | » |
| ANALY_LANCS_SL6 | 142 | 1 | 11 | 16 | 0.41 | 171 | 1 | 0 | 1 | yes | no | » |
| ANALY_FZK | 29 | 3 | 33 | 15 | 0.69 | 80 | 1 | 0 | 1 | yes | no | » |
| ANALY_wuppertalprod | 22 | 0 | 13 | 13 | 0.50 | 48 | 1 | 0 | 1 | yes | no | » |
| ANALY_LAPP | 65 | 4 | 8 | 11 | 0.42 | 88 | 1 | 0 | 1 | yes | no | » |
| ANALY_RAL_SL6 | 59 | 1 | 0 | 11 | 0.00 | 71 | 1 | 0 | 1 | yes | no | » |
| ANALY_DESY-HH | 9 | 4 | 59 | 7 | 0.89 | 82 | 1 | 0 | 1 | yes | no | » |
| ANALY_RALPP_SL6 | 0 | 31 | 6 | 7 | 0.46 | 45 | 1 | 0 | 1 | yes | no | » |
| ANALY_GLASGOW_SL6 | 0 | 4 | 130 | 7 | 0.95 | 143 | 1 | 0 | 1 | yes | no | » |
| ANALY_OX_SL6 | 139 | 0 | 0 | 6 | 0.00 | 145 | 1 | 0 | 1 | yes | no | » |
| ANALY_FREIBURG | 27 | 3 | 0 | 5 | 0.00 | 36 | 1 | 0 | 1 | yes | no | » |
| ANALY_CYF | 81 | 0 | 0 | 4 | 0.00 | 85 | 1 | 0 | 1 | yes | no | » |
| ANALY_BU_ATLAS_Tier2_SL6 | 50 | 4 | 0 | 4 | 0.00 | 58 | 1 | 0 | 1 | yes | no | » |
| ANALY_CPPM | 113 | 4 | 13 | 3 | 0.81 | 133 | 1 | 0 | 1 | yes | no | » |
| ANALY_INFN-NAPOLI | 177 | 2 | 0 | 3 | 0.00 | 182 | 1 | 0 | 1 | yes | no | » |
| ANALY_AGLT2_SL6 | 7 | 8 | 63 | 2 | 0.97 | 80 | 1 | 0 | 1 | yes | no | » |
| ANALY_CAM_SL6 | 181 | 1 | 20 | 1 | 0.95 | 204 | 1 | 0 | 1 | yes | no | » |
| ANALY_PSNC | 0 | 0 | 0 | 0 | 0.00 | 0 | 1 | 0 | 1 | no | no | » |

not all sites like e.g. in US, are setup for webdav/http

# Davix/http - Status III

Several issues in latest davix release:

- Error:DAVIX: NEONRequest::endRequest -> error 256 Error
  closing request -> Unknow Error from libneon
  at a 1-10 % level at almost all sites
    - DPM: seems to be site firewall issues, not all connections open
    - dCache:
        - opens HTTP door at random port dynamically and redirect the client
          to it with a token
        - After a given time, the door is closed in an arbitrary way and trigger
          "Connection timeout"
        - In principle expected behavior since no close all call, but reopening
          should work

- Bad event rate reading performance observed at start of file reading
  when small chunks are read $\rightarrow$ was promised that an advanced
  window buffering algorithm will be implemented

Davix client 0.4.0 should be released in a couple of days which (hopefully)
addresses the first issue

# XAOD READ SPEED TESTS I

Comparison of different protocols:

- Use "real-life" tutorial analysis:
  https://twiki.cern.ch/twiki/bin/view/AtlasComputing/
  SoftwareTutorialxAODAnalysisInROOT
- EventLoop based xAOD analysis, reads/processes muons, jets, stores
  some output
- Discussed ATLAS distributed ROOT I/O meetings

# xAOD read speed tests II

- Using (older) r5720 xAOD input sample
- local nfs, dcap, xrootd, davix access to LRZ-LMU_LOCALGROUPDISK
- AnalysisRelease, Base,2.0.19 with ROOT 5.34.22-x86_64-slc6-gcc48-opt
- Access with `optXaodAccessMode_class` or `optXaodAccessMode_branch`
- TTreeCache is enabled with 10 events/10 MB learning
- Using TTreeCache in ~/.rootrc with
  TTreeCache.Prefill:   0
  TTreeCache.Size:   10485760.
  enables TTreeCache for FAX and Davix access
- Feature in xAOD EventLoop code: patch xAODRootAccess/Root/TEvent.cxx improves TTreeCache learning

# Tests - local access and LRZ-LMU

Randomly repeated 3 times, class access mode

- local: 40/88/77 events/s
- dcap: 51/74/64 events/s
- FAX: 71/69/59 events/s
- Davix: 27/26/26 events/s (no TEvent.cxx patch)
- Davix: 46/56/56 events/s (with TEvent.cxx patch)

Randomly repeated 3 times, branch access mode

- local: 66/100/101 events/s
- dcap: 108/93/106 events/s
- FAX: 90/78/92 events/s
- Davix: 48/47/50 events/s (no TEvent.cxx patch)
- Davix: 65/81/91 events/s (with TEvent.cxx patch):

Important:

- Event rate only for particular analysis - seen factor 5-10 faster event rates (without systematics and heavy output mode)
- Davix suffers from missing buffering in start-up - event rate in later event loop otherwise fine

# SUMMARY AND CONCLUSIONS

- Full Panda setup working to access input via webdav using aria2c/Davix
- Many iterations to debug issues
- Not all tested sites verified to work with aria2c
- Still remaining issues in latest current version of davix 0.3.6

  $\rightarrow$ Still work-in-progress - Davix 0.4.0 should fix some (all?) of the discussed issues.

- When everything correctly setup http/davix is expected and has shown to be similar in performance like other direct I/O protocols