

Operations Guide of AAA

Major goals

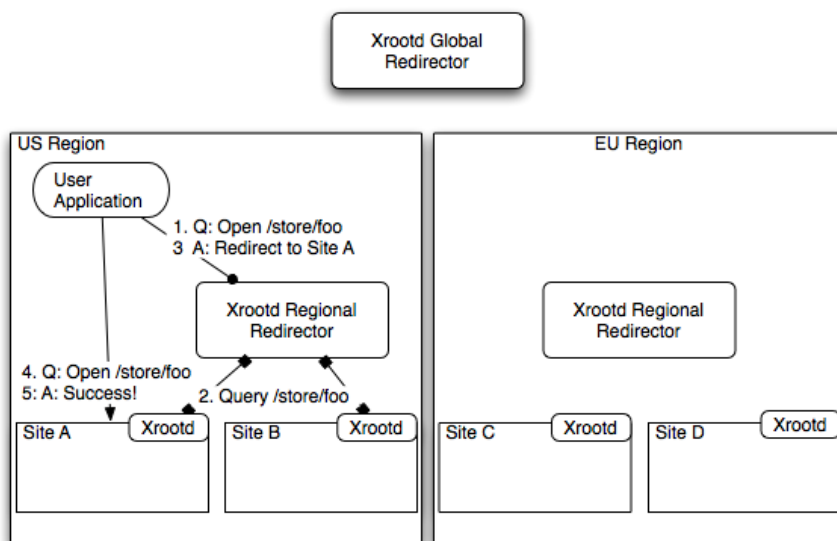
- project focuses on development and deployment of storage resources that are accessible for any data, anywhere at anytime.
- for CMS specifically, the goal is to enable easier, faster and more efficient processing and analysis of data recorded at the LHC through more flexible use of computing resources.

Terminology

- **XRootD protocol:** core SW technology for federating storage
- **Federation:** participating WLCG sites joining AAA concept
- **Topology:** hierarchical structure of the federated data access across participating sites
- **Redirectors:** topology check-points subscribing upstream in the hierarchy of redirectors; redirecting access to local data storage

Federated Storage Infrastructure

- XRootD infrastructure that spans all of the Tier-1 and Tier-2 sites in EU and US CMS;
- Each site's xrootd server is interfaced with the local storage system, allowing it to export the CMS namespace;
 - ◆ current storage systems: dCache and proxy dCache, HDFS, Lustre and DPM, EOS
- site servers subscribe to a local redirector (recommended)
 - ◆ the local redirectors from each site then subscribe to a redundant regional redirector:
 - ◇ US: hosted by FNAL (DNS round-robin alias `cmsxrootd.fnal.gov`)
 - ◇ EU: Bari, Pisa, Paris (DNS round-robin alias `xrootd-cms.infn.it`)
 - ◇ Transitional : hosted by CERN (DNS round-robin alias `cms-xrd-transit.cern.ch`)
- this forms a large tree-structured federated storage system;
 - ◆ a user can request a file from the regional redirector, which will then query the child nodes in the tree and redirect the user to a server that can serve the file.
 - ◆ entire AAA infrastructure overlays on top of existing storage systems, allowing users access to any on-disk data without knowing its location. Simplified version of this process is shown in the attached figure:



Monitoring

See details at [XrootdMonitoring](#)

Monitoring of Xrootd servers and data transfers

What is currently being monitored:

1. **Xrootd summary monitoring - tracking server status**
 - ◆ MonALISA-based monitoring webpages for Xrootd are located here: <http://xrootd.t2.ucsd.edu/>
 - ◆ More details / time-series can be accessed via the ML Java GUI, select group `xrootd_cms` after the window comes up.
2. **Xrootd detailed monitoring - tracking user sessions and file transfers**

Detailed file-access monitoring is implemented as a custom C++ application, !XrdMon. This also provides a HTML view of currently open files (instructions for URL parameters). When a file is closed a report is sent to CERN Dashboard and, for US sites, to OSG Gratia for further processing.
3. **Dashboard-based monitoring**

The Dashboard receives file access report records from XrdMon collectors and integrates them as a part of its WLCG-wide transfer monitoring. The development version is available here: [http://dashb-wdt-xrootd.cern.ch/ui/#vo=\(cms\)#](http://dashb-wdt-xrootd.cern.ch/ui/#vo=(cms)#)

Monitoring Components (System requirements)

Summary Monitoring

- **xrd-rep-snatcher** -- perl script that receives summary UDP packets (quasi xml), normalizes the names, times, etc, calculates rates and forwards the data to ML service. This uses about 10% CPU now, with close to 2000 machines reporting every 30s or so.
- **ML service** -- receives data from xrd-rep-snatcher and other sources, keeps memory cache for last 4 hours (or so) and allows one to do detailed real-time plots of what is going on. This is java, set to use 4GB of ram.
- **ML repository** -- stores data from the service into Postgres, provides long-term storage and web-interface for plots. This uses 4GB ram for java + 8GB ram for postgres, to make it reasonable fast. This requires fast disks, we have 4 SAS disks in raid-5 configuration, currently using about 300+ GB for 2 years worth of data. SSD disks don't work, ours died after 4 months, same was reported by ALICE, they burned three before going to a SAS raid-5.

Detailed Monitoring

- Xrootd servers report user logins, file opens and read ops via binary UDP -- total for cms is between 200 - 600 kB/s.
- UDP-sucker-TCP-server: listens for UDP packets, tags them with receive time + source server, writes them into a root tree (to be able to play them back and debug the stream / collector (up to 10GB / day) and serves them to connected collectors over TCP. Typically, one can run the production collector (next service) + testing / live-monitoring one on desktop (like Matevz does).
- TCP GLED collector: aggregates the messages and builds in-memory representation of everything that is going on in a federation. Reports the following:
 - ◆ serves HTML page of currently opened files (has to be cert protected via apache);
 - ◆ file-access-report on file close to:
 - ◇ AMQ broker@CERN
 - ◇ Gratia collector at UNL (more later)
 - ◇ write report into ROOT trees for detailed analysis (~couple 100 MBs / month).

- UDP collector of "very detailed I/O information" (fully unpacked vector reads, for detailed analysis of read offsets and access patterns). All UCSD servers (~20) and one server from UNL are also reporting in this format. For this, I only write out ROOT trees (about GB per month).
- violation policy concerns by IGTF, we run two collectors:
 - ◆ US collector@UCSD: xrootd.t2.ucsd.edu
 - ◆ EU collector@CERN: CMS-AAA-EU-COLLECTOR.cern.ch:9330
- different config implementation depending on the storage, e.g.: dCache and DPM sites

Summary and Detailed Monitoring Links

Filtered queries

- e.g. how to check xrootd version across every server which reports to collector, (one can found a bit more about collector below)
 - ◆ http://xrootd.t2.ucsd.edu/dump_cache.jsp?pred=%25%2FCMS%3A%3A%25%3A%3AXrdReport%25%2F%25%2F%25.vanderbilt.edu%2F%25&submit
- similarly can monitor CmsdReport streams:
 - ◆ http://xrootd.t2.ucsd.edu/dump_cache.jsp?pred=%25%2FCMS%3A%3A%25%3A%3ACmsdReport%25%2F%25%2F%25.vanderbilt.edu%2F%25&submit
 - ◆ http://xrootd.t2.ucsd.edu/dump_cache.jsp?pred=%25%2F%25%2F%25.vanderbilt.edu%2F%25&submit
- You can query any value like that, if you want to see what is available for given host you can do:
 - ◆ http://xrootd.t2.ucsd.edu/dump_cache.jsp?pred=%25%2F%25%2F%25.vanderbilt.edu%2F%25&submit
- filter by site name
 - ◆ http://xrootd.t2.ucsd.edu/dump_cache.jsp?pred=%25%2FCMS%3A%3A%25%3A%3AXrdReport%25%2F%25%2F%25.vanderbilt.edu%2F%25&submit
- filter by unknown site (collector)
 - ◆ http://xrootd.t2.ucsd.edu/dump_cache.jsp?pred=%25%2FCMS%3A%3Aunknown%25%3A%3AXrdReport%25%2F%25%2F%25.vanderbilt.edu%2F%25&submit
- This is passed as a "like" query to postgres, so % is a wildcard.

Aggregated Reports (Plots)

- http://xrootd.t2.ucsd.edu/display?page=xrd_report/aggregated/xrootd_rdr_all
- http://xrootd.t2.ucsd.edu/display?page=xrd_report/xrootd_redir_by_site

Detailed Monitoring Links

- Tracking user sessions and file transfer
 - ◆ https://sentry.t2.ucsd.edu:4242/xuser/?no_same_site

Dashboard

CMS xrootd dashboard can be found at <http://dashb-cms-xrootd-transfers.cern.ch/>

Kibana a.k.a. shifters instructions

- testing 4 redirector DNS alias instances and 9 redirector host instances :
 - ◆ GLOBAL: cms-xrd-global.cern.ch
 - ◆ EU: xrootd-cms.infn.it
 - ◆ US: cmsxrootd.fnal.gov
 - ◆ Transitional Federation : cms-xrd-transit.cern.ch
- two hosts behind DNS alias, sort of HA service, to produce XMLs which will be pulled by Kibana
 - ◆ xrdfedmonitor-cms.cern.ch (vocms037 and vocms038)
 - ◆ script: /var/www/XRDFED-kibana-probe.py
 - ◇ it executes two commands as root
 - ◇ xrd <REGIONAL_REDIRECTOR | DNS_ALIASED_REDIRECTOR> query 1 a

```
◇ xrdcp -d 2 -f -DIRReadCacheSize 0 -DIRedirCntTimeout 180 root://  
<REGIONAL_REDIRECTOR |  
DNS_ALIASED_REDIRECTOR>/<SAM_test_file> /dev/null  
· SAM tests file :  
  /store/mc/SAM/GenericTTbar/AODSIM/CMSSW_9_2_6_91X_mcRun1_realistic_v2
```

◇ implementation: cms service certificate; service proxy (needed depending on sites policy and cms user mapping)

- xml based information visualized via Kibana
 - ◆ https://meter.cern.ch/public/_plugin/kibana/#/dashboard/temp/CMS::XrootD
 - ◆ If you want to add notifications :
 - ◇ You need to subscribe to ai-admins e-group
 - ◇ Once you have approval to ai-admins group, you will have to add notifications to metric manager. (You have to be in CERN's network or tunnel to access metric manager)
 - ◇ When you are in metric manager, first click on "login" on the right
 - ◇ Then click on "manage", "add notifications"
 - ◇ Fill the areas based on this
- Shifter instructions: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/CMSCriticalServiceXrootd>

Scale tests

- Scale testing measures ability of CMS T2 sites to handle predicted peak loads for AAA
- Tests emulate CMS jobs running at CMS sites
- Two measurements performed:
 - ◆ Rate to open files
 - ◆ Rate of reading data from files
- <https://twiki.cern.ch/twiki/bin/view/Main/CmsXrootdOpenFileTests>
 - ◆ use this twiki to see latest scale test performance per site
- How to run opening and reading scale test from the dedicated cern condor pool
<https://twiki.cern.ch/twiki/bin/view/Main/AAAHowToRun>

Operations and troubleshooting

Troubleshooting guide can be found at

<https://twiki.cern.ch/twiki/bin/view/CMSPublic/CompOpsAAATroubleshootingGuide>

New environment setup

- Basic requirements:
 - ◆ documentation: <http://configdocs.web.cern.ch/configdocs/tutorial/permissions.html>
 - ◆ CERN account:
 - ◇ A connection onto the CERN network.
 - ◇ Belong to a computing group.
 - ◆ Permissions:
 - ◇ Join the ai-admins e-group to have access to the configuration management system. You can join there. The membership will need to be approved.
 - ◇ Log into the Foreman(<https://judy.cern.ch/>) service. If you're already on an e-group that is responsible for some hosts, then you should see them. Otherwise, you should be able to see the "Playground" hosts. If not, please open a SNOW ticket.
 - ◇ When your membership will be approved at ai-admins you should be able to access: <https://gitlab.cern.ch/ai/it-puppet-environments>.
- New environment creation:
 - ◆ documentation: <http://configdocs.web.cern.ch/configdocs/details/environments.html>

- ◆ open terminal:
 - ◇ log into aiadm.cern.ch (ssh username@aiadmNOSPAMPLEASE.cern.ch). Use full hostname and CERN's user name from outside CERN network.
 - ◇ git clone https://:@gitlab.cern.ch:8443/ai/it-puppet-environments.git
 - ◇ cd it-puppet-environments
 - ◇ ai-create-environment-metadata -e new_environment_name -n username@cernNOSPAMPLEASE.ch -d new_environment_branch_name
 - if ai-create-environment-metadata command doesn't work please check `which python` you are using
- ◆ after successful environment creation go to judy.cern.ch
 - ◇ in actions column click on edit
 - ◇ change the environment to which you just created
 - ◇ submit a form
- ◆ last step change environment in puppet.conf
 - ◇ open terminal
 - ◇ log into vocmsXXX
 - ◇ change environment in puppet config (/etc/puppetlabs/puppet/puppet.conf)

AAA support groups

- **Hypernews:** hn-cms-wanaccess@cernNOSPAMPLEASE.ch
 - ◆ interface between site admins and CMS xrootd experts tackle unexpected behaviors in their local infrastructure of redirector(s)
 - ◆ xrootd client/servers and storage setup and other WAN related issues
- **GGUS:** https://ggus.eu/?mode=ticket_cms
 - ◆ Type of issue: "CMS_Datatransfers"
 - ◆ Support Unit "CMS AAA - WAN Access"
 - ◆ usually targeted to site problems affecting federated access and helping review sites' configuration within AAA 'standards' (not only within AAA, also if they have just standalone xrootd cluster not joining federation), storage.xml configuration and TFC configuration
 - ◆ in all cases, first line of defense for operation issues is cms-comp-ops-transfer-team@cernNOSPAMPLEASE.ch
- **Redirectors Admin contacts:**
 - ◆ Global: cms-service-xrootd-global@cernNOSPAMPLEASE.ch
 - ◆ US: cms-service-xrootd-us@cernNOSPAMPLEASE.ch
 - ◆ EU: cms-service-xrootd-eu@cernNOSPAMPLEASE.ch
- **E-group:** cms-service-xrootd@cernNOSPAMPLEASE.ch
 - ◆ contact to report critical incidents, -GLOBAL, -EU, -US and CRC will receive the message.
 - ◆ scheduled maintenance any of the redirectors which may affect AAA file access in general (always CC hn-wan-access@cernNOSPAMPLEASE.ch)

Managing xrootd machines

For machine access, puppet configs, crontabs related info, see CompOpsCentralServicesXrootd

How to examine badly behaving sites in AAA

- The following link shows the **production federation** metric :
<https://dashb-ssb.cern.ch/dashboard/request.py/siteviewhistory?columnid=224>
 - ◆ The sites which are labeled as "red" are failing metric conditions and hence they are candidate for the **transitional federation**. (Transitional Federation)
 - ◆ The metric conditions for being a **bad site** are following :
 - ◇ AAA-related ticket in GGUS open for longer than two weeks.

- ◊ SAM xrootd access test < 50% for two weeks.
- ◊ Hammer Cloud (HC) xrootd test success rate < 70% for two weeks
- If site is failing or a T3:
 - ◆ Open ticket (cc CMS WAN Access and TT) to site asking them to change their redirector
 - ◊ Instructions:
https://twiki.cern.ch/twiki/bin/view/Main/RedirectorsSubscription#Transitional_Federation
 - ◆ Check if they correctly report summary and detail monitoring
 - ◊ http://xrootd.t2.ucsd.edu/dump_cache.jsp?pred=%25%2F%25%2F*.cr.cnaf.infn.it%2F%25sit
 - ◆ Use xrdmapc to figure out if site has correctly updated subscription
 - ◊ <https://twiki.cern.ch/twiki/bin/view/CMSPublic/CompOpsCentralServicesXrootd>

How to learn storage backend and version

- It's important from AAA point of view to know site's storage technologies (dCache, DPM etc..) and version. In case a site is using very old version of a storage system, we need to open GGUS ticket and initiate a system update.
- Please go to AAAOps github page [and](#) have a look at README.md
- To run the script, use lxplus. (not vocms037)

This topic: CMSPublic > CompOpsAAAOperationsGuide

Topic revision: r30 - 2020-07-08 - DonataMielaiKaite



Copyright &© 2008-2020 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

Ideas, requests, problems regarding TWiki? Send feedback