XRootD
Server Status & Plans

XRootD Workshop

IN2P3

June 11 - 12, 2019

Andrew Hanushevsky, SLAC

http://xrootd.org

The Current State

- XRootD has an aggressive release schedule
  - Perhaps too aggressive for available FTE's
    - Release delays occur
- Current adoption is 4.9.0 or 4.9.1 (preferred)
  - Though many still run 4.8.x
- Latest patch/feature release is 4.10
  - RC-1 cut May 13, 2019

The Future

- Upcoming feature release is 5.0.0
  - Still not fully feature fixed
    - Depends on how much time we have
      - Target is 3Q19 barring 4.10.0 patch releases
- Let's look at server-side highlights
  - 4.9, 4.10, 5.0, and beyond

XRootD 4.9.x Server Basic Set

- Vector write (new request code **kXR_writev**)
  - Reduce net latency for certain workflows
- Propagate xrdcp streams spec to TPC
  - Compliance with FTS expectations
- Enhancements to ease containerization
  - Simplify setup for container management
    - Kubernetes, swarm, etc

XRootD 4.9.x Server HTTP Set

- Support **Xcache** ingest via HTTP
  - Increase source choices
  - Allow **Xcache** as a Squid replacement
    - More on this during the workshop
- xrdcp cross protocol copies (xroot <-> HTTP)
  - Simplify tool set vs. available protocols
- Macaroon support
  - Allow token-based authorization for TPC

XRootD 4.9.x Server Security Set

- Subject Alternative Name (SAN) support
  - X509 and RFC 2818 compliance
- Full delegated proxy support
  - Congruent with RFC 2818
    - Solves convoluted interaction with DNS usage
      - Largely triggered by how sites register DNS names
    - Full security requires reissuance of server certs
- TPC support for delegated proxies

XRootD 4.10.0 The Latest One

- 130 commits!
  - Mostly for new bugs, regressions, & features
    - 24% HTTP protocol
    - 19% Client
    - 12% Build issues
    - 11% Server **Xcache**
    - 9% Python bindings
    - 7% SSI
    - 5% Documentation
    - 5% GSI security
    - 5% Miscellaneous
    - 4% Server other

XRootD 4.10.0 Server Features I

- New prepare plug-in
  - Provides path for SRM replacement (bring online)
    - See **ofs.preplib** directive in R5 manual
      - Initially being used by CTA
- Plug-in for cache context management
  - Allow implementation of new cache semantics
    - See **pss.ccmlib** directive
      - Used for CERNVMFS and Rucio name2name

XRootD 4.10.0 Server Features II

- Direct **Xcache** access
  - Eliminates server overhead for complete files
    - See **pss.dca** directive
      - Used in HPC centers with a DFS cache (e.g. Lustre)
      - Only supported for 4.0.0 and above clients
- Multiple **Xcache** write-back streams
  - Dramatically improves I/O performance
    - See **pfc.writequeue** directive

XRootD 4.10.0 Server Features III

- Additional clustering options
  - Reduces possibility of extra file copies
    - Geared for clustered **Xcache** deployments
      - See **cms.sched** directive in R5 reference
- New POSC tuning option
  - Controls **p**ersist **o**n **s**uccessful **c**lose overhead
    - See **ofs.persist** directive in R5 reference

XRootD 5.0.0 The Next Big One

- Introduces many new features
  - Breaks plug-in ABI in some cases
    - Some external plug-ins will need to recompile
- It's very ambitious & planned for 3Q19
  - So, I will break it down to…
    - Done deal, on track, hopefully will get in
- Focus on server
  - Client covered in another talk

XRootD 5.0.0 Done Deal I

- User settable file extended attributes
  - Allows adding metadata to a file
    - See **ofs.xattr** directive
      - Restricted to user namespace
- New monitoring G-Stream
  - For periodic medium level information
    - Supports JSON, XML, and many others
    - See **xrootd.monitor** directive
      - Currently used by **Xcache** and CCM plug-ins

XRootD 5.0.0 Done Deal II

- Extended stat information
  - Allows future support for uid/gid tracking
- Add additional security entity fields (breaks ABI)
  - Allows future support for uid/gid security
- Trivialize ofs plug-in wrapping (breaks ABI)
  - Avoids disabling XRootD performance features
    - This has happened in the past, sigh.

XRootD 5.0.0 On Track I

- Full TLS support (a.k.a. roots/xroots)
  - Required for authorization token handling
  - Improves security and data integrity
  - Only obstacle is host name verification
    - This is a long-standing OpenSSL issue
    - Current plan is to activate it for ssl 1.0.2 and above
      - Otherwise, the default is to trust DNS (always an option)
        - RHEL6 is at 1.0.1e so this is a concern EOL 11/30/2020
        - RHEL7 is at 1.0.2k which is barely acceptable

XRootD 5.0.0 On Track II

- TPC version 3
  - Support getFile() and putFile() requests
    - Set stage for safe authorization token support
      - Compliance with XRootD server security architecture
  - Current obstacle is putFile
    - Semantics are very messy and error prone
    - Current plan is to go with getFile and delay putFile
      - We don't want putFile being a show stopper

XRootD 5.0.0 On Track III

- State full redirects (i.e. on read)
  - Dramatically enhances **Xcache** bypass
    - Redirect to local data source when file is completed
- Remove old client
  - At least no longer compiled
    - This may impact ALICE

XRootD 5.0.0 Hopefully…

- Allow checksum verification on close()
  - Requires modifications to open() request
    - Server needs to know at file open to set it up
  - Server will indicate if this is supported
    - This is not a straight forward modification
      - It needs to be fool-proof

Beyond XRootD 5 Wish List I

- Recursive delete (mostly for HTTP)
  - Driven by WebDAV semantics
- Full data streaming (i.e. no chunking)
  - Can substantially improve transfer rate
- Apply/Map operation for data pipelining
  - Can dramatically reduce WAN latency
    - Sometimes known as request bundling
      - Good for certain workloads

Beyond XRootD 5 Wish List II

- RDMA support
  - Allow HPC's to run at full speed
- Erasure encoding plug-in
  - Enable another space saving option
- Native data striping across partitions
  - For improved performance
- The uid/gid tracking for files/directories
  - This is a long-standing request (may be moved up)

Beyond XRootD 5 Wish List III

- Allow appends to a zip archive
  - Another long standing request (may be moved up)
- Dynamic data source selection in the client
  - A long-standing ALICE request
    - Already being done by CMS
- Enable mock testing of client
- Docker based distribution
- Implement package config functionality

In The End

- **XRootD** has a plethora of requests
  - Constant struggle to prioritize these
    - **XRootD** is now embedded in practically every HEP data delivery system
      - EOS, DPM, CTA, dCache (Java version), QSERV, native …
    - New experiments rely on the **XRootD** as well
      - E.g. Dune, LSST, LCLS II
  - All of this requires careful threading
    - We must ensure we don't break any of these