FTS improvements for LHC Run-3 and beyond

CHEP 2019 Adelaide

Edward Karavakis
on behalf of the FTS team
Contents

• Intro
• WLCG DOMA TPC
• Performance improvements
• New features
• XDC
• Conclusion
FTS

- Distributes majority of LHC data across WLCG infrastructure
- 7 WLCG and 13 non-WLCG instances
- ~25 Virtual Organisations
  - ATLAS, CMS, LHCb, AMS, NA62, Compass, ILC, Magic, Belle II, Mice, Xenon, Snoplus, Gridpp, Dune, LZ, Solidexperiment.org, SKA, Ligo, Icecube, Elixir, NP02, CAST, ESCAPE, Eiscat.se, Virgo
- Integrated with experiment frameworks: Rucio, PhEDEx, DIRAC
- Transferred in 2019 so far >750 PB (700 only for ATLAS, CMS and LHCb)
FTS Core Features

Simplicity
- Easy user interaction for submitting transfers. Copy one file from one place to another
- WebFTS portal for end-users, Real Time monitoring and Web Admin

Reliability & Integrity
- Checksums and retries are provided per transfer
- Multiprotocol support (HTTP, gsiftp, xrootd, SRM, S3,..)
- Different clients to access the service (REST APIs, python bindings)
- Transfers from/to different storages
- Support for bringonline
- FTS can be run "zero config"

Flexibility & Scalability
- Transfers from/to different storages
- Support for bringonline
- FTS can be run "zero config"

Intelligence
- Parallel transfers scheduling and optimisation to get the most from network without burning the storages
- Priorities/Activities support for transfers classification
WLCG DOMA TPC activities

See A. Forti's talk here

Towards a replacement for gridftp WAN transfers

- Production ready version of FTS with
  - XRootD TPC with X509 delegation
  - Macaroon acquisition for HTTP TPC

- Provisioning of FTS testbed for HTTP and XRootD TPC testing (functional and stress)

- Few production transfers via HTTP and XRootD so far, waiting for the (ongoing) storage upgrade campaigns
Performance improvements

Motivation

• Scheduler and Optimiser performance affected with >2K links and >2M submitted transfers
  • DB queries do no scale $\Rightarrow$ reduced number of transfers scheduled with increased load

• Improvements needed for experiments to be able to use less FTS instances (even one) and to handle Run-3 load
Performance improvements

• First set of improvements were released already in 3.9 series
  • Adding missing indices by examining all slow queries and passing them through a profiler
  • Optimisations to get rid of expensive joins

• Work on DB table partitioning bringing significant performance gains (to be released soon)
New features

- CTA is the new tape based solution under implementation at CERN that exposes an XRootD interface and supports TPC
  See E.Cano's talk here
- EOS+CTA integration is production ready – interface to FTS doesn’t change for the experiments
  - Implemented staging via XRootD
  - Disk copy eviction on transfer completion, to better handle the reduced buffer size
  - Staging+Multihop supported (to handle data export to T1s)
- Stress-tested during the ATLAS Data Carousel exercise
  See X.Zhao's talk here
New features

- Protection against duplicate transfers by avoiding double submission to the same destination URL as requested by ATLAS

- Archive to tape monitoring
  - Reporting a transfer to tape-backed storage as completed only when the file has been migrated to tape successfully
  - Requested by CMS as part of their migration to Rucio and by CTA as a way to throttle active transfers
  - First prototype implemented and tested both via SRM and XRootD
XDC

See P. Fuhrmann's poster here

- EU Horizon2020 XDC software development project started in Feb 2018 ➔ http://www.extreme-datacloud.eu

- “Developing scalable technologies for federating storage resources and managing data in highly distributed computing environments”

- Funded FTS activities
  - Integration with OIDC (OpenID Connect)
  - CDMI protocol integration to support QoS transitions
XDC

- **OIDC support**
  - FTS supported only X509 authentication and delegation to contact storages
  - First version of FTS with OIDC support released for XDC
  - Offline validation implementation
  - Refresh token acquisition and token refresh daemon implemented
  - Successfully demoed with Rucio as client

- **QoS support**
  - First version of QoS daemon and gfal2 CDMI API released for XDC – bringonline managing two QoS: “disk” and “tape”
  - PoC with basic functionality
  - Request and monitoring of a QoS transition

11/04/2019
Conclusion

• FTS continues to evolve with the infrastructure as WLCG's principal data movement service

• Various performance improvements and new features put in place in preparation for Run-3

• Full support planned for TPC and token auth

• Integration with CERN's new tape archival system CTA

• Expanding community and adoption by upcoming data intensive projects