FTS summary of the XRootD and FTS Workshop 2023
Jozef Stefan Institute, Ljubljana, Slovenia

Grid Deployment Board (GDB) meeting

Steven Murray on behalf of the CERN FTS team

Wednesday 12th April 2023

Introduction

- **Co-hosted with XRootD – Many thanks to Andrew Hanushevsky**

- **From Monday 27th to Friday 31st March 2023**

- **2 and a half days for FTS: Monday, Tuesday and Wednesday morning**

- **45 registered participants: ~35 in-person and ~10 over Zoom**

- **Majority of talks given in-person with a few presented over Zoom**

- **Indico: https://indico.cern.ch/event/875381/**

FTS timetable - Monday

State of affairs

27/04/23 14:15	FTS3: State of affairs	Mihai Patrascoiu	Slides
27/04/23 15:00	Tape REST API and more	Joao Pedro Lopes	Slides
27/04/23 15:50	FTS & Tokens	Shubhangi Misra	Slides

Site reports

27/04/23 16:20	FTS3 @ CERN	Steven Murray	Slides
27/04/23 16:45	FTS3: BNL Deployment	Hironori Ito	Slides
27/04/23 17:00	Update on FTS at RAL	Rose Cooper	Slides
27/04/23 17:15	FTS3 at FNAL	Lorena Lobato Pardavila	Slides

FTS timetable - Tuesday
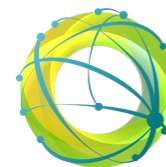
Communities and Collaborations

28/04/23 9:30	FTS Community Talk: ATLAS	Mario Lassnig	Slides
28/04/23 9:55	FTS Community Talk: CMS	Katy Ellis	Slides
28/04/23 10:25	FTS Community Talk: LHCb	Ben Couturier, Christophe Haen	Slides
28/04/23 10:50	EGI Data Transfer Activities	Andrea Manzi	Slides
28/04/23 11:15	The journey of a file in Rucio	Radu Carpa	Slides
28/04/23 11:40	FTS-Alto project	Y. Richard Yang	Slides

Planning

28/04/23 14:00	FTS: Cloud Storage Transfers	Mihai Patrascoiu	Slides
28/04/23 14:00	FTS Planning Session	Mihai Patrascoiu	Slides

Monitoring

28/04/23 16:00	FTS3: The Monitoring Zoo	Joao Pedro Lopes	Slides
28/04/23 16:35	FTS3@CERN: Service Health Monitoring	Mihai Patrascoiu	Slides
28/07/23 17:00	FTS-Noted Project	Edoardo Martelli and Maria Del Carmen Misa Moreira	Slides

FTS timetable - Wednesday

Discussion

|	
29/04/23 9:30	Open Discussion
29/04/23 11:00	The Great Buffer
29/04/23 11:30	Meet the developers: FTS, XRootD, ROOT, EOS, OSG and discuss future needs

FTS3: State of Affairs
Mihai Patrascoiu

- **Increased database performance in 2021 using an optimized tape staging query**

- **Deployed a database replica in 2021 to increase robustness**

- **Added health monitoring and alarming in 2021/2022 so that the team could be proactive in problem solving**

- **Migrated to MySQL 8 to get the ability to run on-line schema upgrades to reduce downtimes**

FTS releases		
FTS v3.10 (2020)	**FTS v3.11 (2021)**	**FTS v3.12 (2022)**
Archive monitoring	**Destination file integrity report**	**FTS-REST-Flask**
FTS-QoS daemon	**Support for SE-issued tokens**	**MySQL8**
Initial OIDC token support	**New tape stage query**	**Tape REST API**

- **The next major FTS release is expected in Autumn 2023 and will be nick names "Spring Cleaning"**

- **The `fts_bringonline` daemon, the `fts_bdii` reporting system and the C++ FTS clients will be removed**

- **Long overdue code consolidations will be made such as moving to a newer C++ compiler for better static analysis and code quality**

Tape REST API and more
Joao Pedro Lopes
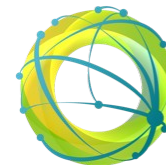
- The tape REST API will replace the tape part of the SRM protocol

- This work involves the EOSCTA, dCache and StoRM storages and theFTS/Gfal2 developers

- There is a DOMA task force: `wlcg-tape-rest-api-discussions@cern.ch`

- The tape REST API reference document [May 2022] can be found here: Link

- Tape storage requires 4 more file transfers operation that disk

1. Has the file been archived to tape?

    `davix-http ... https://eosctaatlaspps.cern.ch:8444/api/v1/archiveinfo/ ...`

2. Please stage the file from tape to disk

    `davix-http … https://eosctaatlasspps.cern.ch:8444/api/v1/stage ...`

3. Has the file been staged on disk?

    `davix-http … https://eosctaatlaspps.cern.ch:8444/api/v1/stage/xxxx ...`

4. Please release the disk copy (as long as there is a copy on tape):

    `davix-http … https://eosctaatlaspps.cern.ch:8444/api/v1/release/xxxx ...`

FTS & Tokens
Shubhangi Misra

- FTS token support is a pre-requisite for migrating the WLCG from X509 certificates to tokens

- This work will be used to validate the data workflows of the main LHC experiments

- This work will hopefully highlight and address any unforeseen shortcomings and scalability issues

- This is not only the work of the FTS team but also:

  - WLCG

  - LHC Experiments

  - Storage providers

  - Authentication and Authorization Services

  - EGI community

  - OSG community

FTS3 @ CERN
Steven Murray

- **The FTS service at CERN installs two extra machines:**
  - **A watchdog machine to monitor and alarm the health of the service**
  - **A database backup machine that extracts disaster recover data from the FTS databases, encrypts it and stores in the EOS disk storage service at CERN**
- **Running the FTS database on a "cloud" like database service requires correctly setting these two parameters:**
  - **`innodb_buffer_pool_size` – RAM cache**
  - **`innodb_online_alter_log_max size` – Log of transactions made during on-line DDL changes**
- **The FTS service has its data privacy notice here: Link**
- **The disaster recovery strategy of the CERN FTS service is**
  - **Plan A - Try to recover here at CERN**
  - **Plan B - Ask experiments to redirect to alternative FTS sites**
- **The importance of the CERN FTS service is demonstrated by the fact that it has been responsible for transfering between 90 and 150 PB of data per month during 2022**

FTS3: BNL Deployment
Hironori Ito

- **The main clients of the FTS service at BNL are the ATLAS and Belle II experiments**
  - **BNL FTS was used as a backup for CERN ATLAS FTS**
  - **BNL FTS transfers 2.5 M files per week totalling 4 PB of data**
- **The future deployment of FTS at BNL will be to put all of the frontends on VMs and the FTS database on physical hardware**
- **FTS is being monitored through the use of Grafana running against the FTS MySQL database**
- **The monitoring can be access here: Link**
- **US ATLAS has started a mini Data throughput test**
- **The test has concluded that the throughput requirements of High Luminosity LHC should be met if slow transfers and storages are not allowed to interfere with faster data transfers**

Update on FTS at RAL
Rose Cooper

- FTS at RAL is being used by EGI

- FTS at RAL is beginning to work with the Square Kilometere (SKA) radio telescope

- RAL plan to make a proposal to involve IRIS

- RAL plan to deploy an FTS instance dedicated to SKA

- The dedicated SKA instance will execute test transfers before being scaling to meet the needs of SKA

- RAL plan to enable IPv6 on all of their production and test FTS instances

- RAL plan to integrate their EGI instance with EGI Check-in and their SKA instance with SKA-IAM

- RAL plan to migrating their FTS databases to a new infrastructure

- The architecture of the FTS instances will be made more consistent

FTS3 at FNAL
Lorena Lobato Pardavila

- Fermilab (FNAL) run two configurations of FTS3

- For CMS, FNAL run the FTS servers and database on a single physical host

- For public experiments, mostly DUNE, FNAL run a containerized version of the FTS servers on OKD with the database running on a VM

- FNAL have found FTS to be low maintenance if used with the recommended configuration

- FNAL hope to expand on the idea of having a containerized public FTS instance for transfers involving the Americas

- FNAL see that database errors cause the the FTS server to crash and they hope that the FTS developers can work on thisons

- Have to learn everything the hard way, bothering FTS devs quite often

FTS Community Talk: ATLAS Data Flows & FTS Mario Lassnig

- ATLAS has noted the significant increase in WLCG network performance over the last two decades

- As a result ATLAS now uses a single WORLD cloud site concept

- ATLAS currently uses approximately 130 sites

- ATLAS uses Rucio to manage the creation, location, transfer, deletion, annotation and access of it data

- ATLAS sites use different storage systems and transfer protocols

- ATLAS Rucio relies on FTS, GFAL and Davix to handle these differences

- ATLAS has R&D projects with Cloud services provides by Amazon, Google and SEAL Storage

- Integrating Cloud services into the ATLAS Distributed Computing (ADC) systems involves PanDA, Rucio and in turn FTS, GFAL and Davix

- ATLAS want to adopt the HTTP TAPE REST API earlier than later

- There are Four volunteer sites: CERN, FZK, DESY and BNL

- ATLAS will deploy and test at scale token-based authentication during the 2024 Data Challenge

- FTS is absolutely essential for ATLAS

- ATLAS find the FTS software to be stable and efficient

- The FTS development and operations team are friendly, diligent and quick

- ATLAS believe it is crucial that CERN IT management continue their strong long-term support of the FTS team

CMS/FTS Community Talk
Katy Ellis

- CMS provide Rucio with data placement rules in order to manage their data placement

- In turn Rucio submits file transfer requests to FTS

- CMS also make streamed data reads that do not use FTS

- CMS uses 4 FTS instances:

  - CERN

  - Fermilab (doing most of the work)

  - Imperial College

  - RAL

- Rucio selects which FTS instance to use based on the final destination of a data transfer

- CMS asked FTS to introduce an archiving status

- CMS have used this feature to quickly determining if tape systems are working correctly

- CMS would like to improve the handling of 'destination file exists'

- CMS would like to know if the transfer of 100 GB files would be feasible for physics Run 4

- CMS thank the FTS team for their continued collaboration and support

- CMS has not risen its use FTS above 2019/2020 levels

- CMS do not expect file transfer rates to significantly rise until physic Run 4

FTS Community Talk: LHCb
Ben Couturier, Christoph Haen

- **LHCb use FTS to execute all of their Third Party Copy transfers**

- **All file transfers are orchestrated by DIRAC**

- **LHCb have almost achieved the use of the HTTPS protocol everywhere**

- **XRootD is still used in some cases**

- **LHCb have successfully used the HTTP REST Tape API for 1 week**

- **LHCb use a single FTS instance at CERN**

- **This instance has been reliable and sufficient**

- **A single instance means no configuration discrepancies between instances**

- **A new FTS instance will be created if the current one fails**

- **LHCb are very happy with the CERN FTS service**

- **LHCb look forward to using the new "explain" feature of the FTS optimizer on the FTS Web Monitoring pages during the next data challenge**

- **LHCb stated that FTS is absolutely paramount for LHCb computing**

EGI Data Transfer Activities
Andrea Manzi

- The EGI e-Infrastructure is an international Federation

- It has evolved from the high-energy physics compute grid (WLCG) in 2010 to a multi-disciplinary, multi-technology infrastructure in 2023
  - VIRGO
  - IceCub
  - WeNMR
  - LSST
  - ENES
  - 250 +

- EGI uses FTS to carry out its data transfer: Link

- EGI has two FTS installations:
  - UKRI-STFC – For three VOs outside of the WLCG
  - CERN FTS Public and CERN X509 based WebFTS - piloting AAI integration+HTTP TCP and new EC Projects

- CERN FTS Public is configured to use all of the EGI Check-in instances

- UKRI FTS is currently being configured

- EGI and CERN have made a new project proposal called EOSC Beyond which plans to continue support and extend the functionalities with include:
  - More storage types and protocols
  - The ability to allow the source and destination storages in a single file transfer to use different authentication methods
  - Improved integration with EGI Check-in

The journey of a file in Rucio
Radu Carpa

- Rucio is a data catalogue that references and organizes files stored on multiple storages. Specifically Ruco can:
    - Upload files to storages
    - Group files them into datasets
    - Enforce replication rules
    - Recover from the loss of replica files
- In production enviroments Rucio uses GFAL2 to access storage
- Rucio uses FTS for third-party copy transfers in over 99% of it production cases
- Advanced features such as the following require joint Rucio and FTS development
    - The automatic detection and overwrite of corrupted files
    - Support for Cloud storage
    - Waiting for files to be safely archived to tape
    - Support for multiple source and multiple hop transfers

- Rucio and FTS plan to provide:
    - Scitags packet marking
    - Full ODIC token support
- Rucio have developed Kubernetes based demonstrator:
    - Available at: Link
- ATLAS operators say that FTS works very well and gets the job done
- ATLAS operators also say that FTS sometimes makes questionable decisions
- ATLAS find the FTS development team to be responsive and helpful.
- ATLAS ask how we can reduce sub-optimal FTS decisions?
- ATLAS would like first-party FTS containers

ALTO/FTS: FTS Control with Deeper Network Visibility
Y. Richard Yang

- The ALTO /FTS project wishes to extend FTS to be an application-defined networking controller, through deeper network (resource) state visibility.

- This should enable FTS to improve its efficiency and flexibility

- ALTO is an effort of the ALTO working group in the Transport Area of Internet Engineering Task Force (IETF)

- ALTO plans to integrate into FTS version 3.12

- ALTO will extend the FTS database schema to support resource control specification

- ALTO will implement an ALTO/TCNOptimizer class for FTS

- Slide 17 – Milestones and Main Remaining Tasks

  - ALTO wish to wrap up implementation and testing during summer 2023

  - ALTO wish to finalize the language used to specify the resource model

  - ALTO want to work with sites and higher-level workflows in order to be able to specify control goals

  - ALTO would like to finish an initial design for the coordination and control of multiple FTS instances

  - ALTO would like to integrate with infrastructure control systems such as the NOTED project

FTS: Cloud Storage Transfers
Mihai Patrascoiu

- **The key aspects of Cloud storage are**
  - **They don't support third party copy HTTP transfers**
  - **They don't support directory listing, checksuming and the creation of directories**
  - **Pre-signed URLs are not helpful because:**
    - **There is now guarantee of the time window in which they will be used**
    - **URLs get posted everywhere**
  - **FTS needs to configured with signing secrets and more**
- **FTS has been redesigned to put all of the Cloud configuration in one intuitive place**
- **Many thanks to Eraldo for his work on this area!**

FTS Planning Session

- **FTS will replace steering meetings with:**

  - **An FTS Community Forum**

  - **Regular "check-up" meetings**

- **The next focus areas for FTS are**

Service operations	Community requests	Project evolution
True microservice model	FTS global config	High Luminosity LHC
Reduce components coupling	Aggressive Optimizer	Modernise codebase
Add built-in service protection	User-friendly Cloud configuration	Deterministic + Global view scheduler
Add built-in service health monitoring	Improved HTTP(TPC) error reporting	DMC clients evolution
		Transfers for non-WLCG environments
		Activity & Priority discussion

- **Impromptu topics raised during the workshop:**

  - **Tokens**

  - **Kubernetes**

FTS: The Monitoring Zoo
Joao Pedro Lopes

- FTS Web Monitoring
  - An interactive web application
  - Developed and maintained by the FTS team
  - Python3 and the Django framework
  - Used to visualize:
    - The status of individual transfer jobs
    - Statistics about queues and finished transfers
    - Link utilization and transfer performance
    - The configuration of storages and links

- Monitoring FTS transfers via the command-line
  - `yum install fts-rest-client`
  - Monitors individual file transfers
- FTS ActiveMQ Monitoring Messages
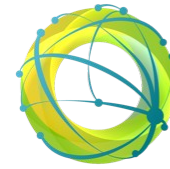 - A dedicated FTS daemon sends messages to CERN ActiveMQ
- Grafana monitoring
  - ElasticSearch indexes can be found at: Link
- FTS Service Health Monitoring
 - Follow the workshop talk given by Mihai Patrascoiu

FTS3@CERN: Service Health Monitoring
Mihai Patrascoiu

- Service health monitoring is for FTS operators
- Some monitoring data are collected indirectly via logs
- Some are collected directly via probes and scripts
- The FTS code was enhanced to print key moments in FTS execution
- Service health monitoring is deployed in every CERN FTS instance
- FluentBit is used to collect monitoring data from logs
  - Tails the log files
  - Parses the desired data
  - Performs 1 minute aggregations
  - Sends aggregations once per minute to Graphite

- Custom written probes:
  - Execute a cron jobs
  - Never run faster than once a minute
  - Send their output to Graphite
  - Some probes generate alarms
    - Send an e-mail to an SMS gateway if a threshold is reached
    - Timestamp the last alarm
    - Have a configurable delay between alarms
  - Probes are run in the dedicated watchdog machine
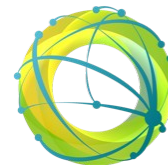
# Network Optimized Transfers for Experimental Data (NOTED)
Carmen Misa Moreira and Edoardo Martelli

- Large HEP data transfers easily saturate WAN network links

- Routing protocols don't take into account link utilization and may force traffic to go over already congested paths, while alternative paths may be left idle

- NOTED plans to reduce the duration of file transfers

- NOTED plans to optimize network utilization by avoiding congestion and reducing idle time

- NOTED is an intelligent network controller that:

  - Continuously checks the status and the logs of the applications that generate large data transfers

  - Knows the links that may get congested and the source-destination sites that can originate the congesting transfers

  - Tries to predict the duration of large transfers

  - Takes corrective actions only if the transfers are long lived

- NOTED has been tested with:

  - SENSE to request direct circuits between source and destination sites

  - Juniper APIs to re-configure BGP metrics and load balance traffic over multiple links

- Version 2 of NOTED has already released

  - Packages: Link

  - Docker container: Link

- NOTED makes decision by watching and understanding the behaviour of transfer services.

- Transfer Applications do not need to be modified to work with NOTED

- NOTED capabilities have been demonstrated with production FTS transfers

Discussions

- FTS community wants tokens and containers

- Tokens:

  - Discussed how to FTS could handle arbitrary token scopes (from top directory to sub-directory to full path)

  - Discussed how specific FTS token solutions should be made in order to take some of the development burden away from Rucio, Dirac and small experiments

- Containers:

  - Already at least two third-party sources of FTS containers:

    - Service Level At the Edge (SLATE)

    - Rucio Kubernetes demo

  - FTS project should consider providing first-party images

The future of FTS discusssions
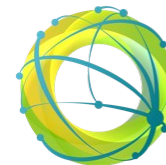
- **The FTS team plan to have a "check-up" workshop or meeting probably every 12 or 18 months**

- **The team also plan to put in place an FTS Community Forum**

- **If you have any feedback we can be reached at:**

  - **fts-devel@cern.ch**