Science & Technology Facilities Council
e-Science

Summary of the status and plans of FTS and LFC back-end database installations at Tier-1 Sites

Gordon D. Brown

Rutherford Appleton Laboratory

WLCG Workshop
Victoria, BC

1st/2nd September 2007

Science & Technology Facilities Council
e-Science

Overview

- Acronyms
- 10 Sites
- 26 Questions
- Developer overview
- Summary

Acronyms - FTS

- Full-Text Search (filename extension)
- Fourier Transform Spectroscopy
- Fault Tolerant System
- Federal Technology Service
- Fourier Transform Spectrometer
- Federal Telecommunications System
- Flying Training Squadron (USAF)
- Flight Termination System
- Full-Time Support
- Force Transfer System (Mavic)
- Full Time Student
- Federal Telecommunications Service
- Federal Telephone System
- Financial Tracking System
- Federal Customs Service (Russian)

Acronyms - FTS

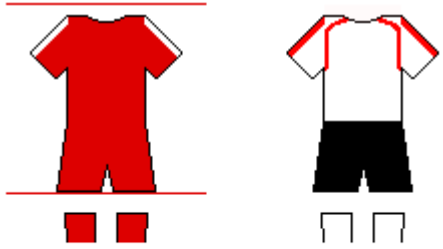
- **Fourier Transform Spectroscopy** is a measurement technique whereby spectra are collected based on measurements of the temporal coherence of a radiative source, using time-domain measurements of the electromagnetic radiation or other type of radiation. It can be applied to a variety of types of spectroscopy including optical spectroscopy, infrared spectroscopy (FTIR), nuclear magnetic resonance, and electron spin resonance spectroscopy. There are several methods for measuring the temporal coherence of the light, including the continuous wave Michelson or Fourier transform spectrometer and the pulsed Fourier transform spectrograph (which is more sensitive and has a much shorter sampling time than conventional spectroscopic techniques, but is only applicable in a laboratory environment).

Acronyms - LFC

- Liverpool Football Club
- Lake Forest College (Lake Forest, IL)
- Level of Free Convection (meteorology)
- Los Fabulosos Cadillacs (Argentina band)
- Large Format Camera
- Land Forces Command (Canada)
- Load Frequency Control
- Laminar Flow Control
- Lightforce
- Lost Foam Casting
- Low Flying Chart
- Looking for Clan (gaming forum)
- Lowest Feasible Concentration
- Local Function Capabilities
- Leesburg Football Club (Virginia, USA)

Acronyms - LFC

- Full name: Liverpool Football Club
- Nickname(s): Pool, The Reds
- Founded: March 15, 1892
- Ground: Anfield, Liverpool, England
- Capacity: 45,362
- Chairman: Tom Hicks (co-chairman)
- George Gillett (co-chairman)
- Manager: Rafael Benítez
- League: Premier League
- 2006–07: Premier League, 3rd



Home colours

Away colours

Overview - FTS

- grid **F**ile **T**ransfer **S**ervice
- FTS Web Service
  - This component allows users to submit FTS jobs and query their status. It is the only component that users interact with.
- FTS Channel Agents
  - Each network channel, e.g CERN-RAL has a distinct daemon running transfers for it. The daemon is responsible for starting and controlling transfers on the associated network link.
- FTS VO Agents
  - This component is responsible for VO-specific parts of the transfer (e.g. updating the replica catalog for a given VO or applying VO-specific retry policies in case of failed transfers). Each VO has a distinct VO agent daemon running for it.
- FTS Monitor
  - This is currently a CERN only element.

Overview - LFC

- **L**CG **F**ile **C**atalog
- The LFC is a catalog containing logical to physical file mappings. Depending on the VO deployment model, the LFC is installed centrally or locally.
- The LFC is a secure catalog, supporting GSI security and VOMS.
- In the LFC, a given file is represented by a Grid Unique IDentifier (GUID). A given file replicated at different sites is then considered as the same file, thanks to this GUID, but (can) appear as a unique logical entry in the LFC catalog.
- The LFC allows us to see the logical file names in a hierarchical structure.

Science & Technology Facilities Council
e-Science

FTS

Do you have FTS running at your site?

- CERN (Geneva): Yes
- PIC (Spain): Yes
- ASGC (Taiwan): Yes
- GridKa (Germany): Yes
- BNL (USA): Yes
- CNAF (Italy): Yes
- Triumf (Canada): Yes
- IN2P3 (France): Yes
- SARA (Holland): Yes
- Rutherford (UK): Yes

If "yes", what back-end database does it run on?

- CERN (Geneva): Oracle
- PIC (Spain): Oracle
- ASGC (Taiwan): Oracle
- GridKa (Germany): Oracle
- BNL (USA): Oracle
- CNAF (Italy): Oracle
- Triumf (Canada): Oracle
- IN2P3 (France): Oracle
- SARA (Holland): Oracle
- Rutherford (UK): Oracle

Would you consider this database development, test or production?

- CERN (Geneva): Production
- PIC (Spain): Production
- ASGC (Taiwan): Production
- GridKa (Germany): Production
- BNL (USA): Production
- CNAF (Italy): Production
- Triumf (Canada): Production
- IN2P3 (France): Production
- SARA (Holland): Production
- Rutherford (UK): Production

If you have a production copy, do you also have a dev or test copy?

- CERN (Geneva): Yes
- PIC (Spain): Pre-production and Test
- ASGC (Taiwan): No
- GridKa (Germany): No
- BNL (USA): No
- CNAF (Italy): Test generic RAC
- Triumf (Canada): Old database as VMware Guest
- IN2P3 (France): No
- SARA (Holland): No
- Rutherford (UK): Test schema on Prod

Is this database dedicated to FTS or does it share it with other schemas/applications?

- CERN (Geneva): Shared
- PIC (Spain): Dedicated
- ASGC (Taiwan): Dedicated
- GridKa (Germany): Dedicated
- BNL (USA): Dedicated
- CNAF (Italy): 3 Node cluster shared with LFC
- Triumf (Canada): Dedicated, same machines as application
- IN2P3 (France): Shared
- SARA (Holland): Dedicated
- Rutherford (UK): Dedicated

Is this database a cluster? If so, how many nodes?

- CERN (Geneva): 8 node cluster
- PIC (Spain): Single instance
- ASGC (Taiwan): Single instance
- GridKa (Germany): Single instance
- BNL (USA): Single instance, will soon be 2 nodes
- CNAF (Italy): 3 node cluster
- Triumf (Canada): Single instance
- IN2P3 (France): 2 node cluster
- SARA (Holland): Single instance
- Rutherford (UK): Single instance

What is the backup policy on this database?

- CERN (Geneva)
  - Weekly full, 6xWeek differential, 1xWeek Cumulative Archive logs everyhour.On Disk backup as well.
- PIC (Spain)
  - Full once a week, mon,tue,wed backup differential...thu cumulative, fri and sat dif...archive logs each hour
- ASGC (Taiwan)
  - None
- GridKa (Germany)
  - Daily diff-backup in TSM
- BNL (USA)
  - Full backup per week, incremental backup per day, archivelogs backup per day. Will be reviewed when moved to RAC.

What is the backup policy on this database?

- CNAF (Italy)
  - L0 backup plus archivelog every week. L1 backup plus archivelog every day. Recovery windows is 31 days.
- Triumf (Canada)
  - Weekly online full
- IN2P3 (France)
  - Full by week. incremental by day.
- SARA (Holland)
  - None
- Rutherford (UK)
  - Full backup every week, incremental every day.

Science & Technology Facilities Council
e-Science

How is this database monitored?

- CERN (Geneva)
  - Home made scripts for DB level + Lemon monitoring for host level
- PIC (Spain)
  - Host with Ganglia, the database with OEM, a local one
- ASGC (Taiwan)
  - Service Monitored. (e.g. nagios)
- GridKa (Germany)
  - DB: OEM (DB host: nagios, ganglia).
- BNL (USA)
  - Nagios, Enterprise Manager.

How is this database monitored?

- CNAF (Italy)
  - Via Grid Control
- Triumf (Canada)
  - The database is not currently directly monitored. However the node is monitored via Nagios (Disk space, host 'up-ness', ssh availability); Ganglia monitors load and swap, and Nagios also looks up the most recent SAM FTS functional tests and runs an ldap query of the FTS BDII.
- IN2P3 (France)
  - oracle tools
- SARA (Holland)
  - Manually.
- Rutherford (UK)
  - Oracle's Grid Control for database/host and nagios for host.

Are you using Data Guard on this database?

- CERN (Geneva): No
- PIC (Spain): No
- ASGC (Taiwan): No
- GridKa (Germany): No
- BNL (USA): No
- CNAF (Italy): No
- Triumf (Canada): No
- IN2P3 (France): No
- SARA (Holland): No
- Rutherford (UK): No

Do you have any other redundancy built in to this database?

- CERN (Geneva): No
- PIC (Spain): No
- ASGC (Taiwan): No
- GridKa (Germany): No
- BNL (USA): Hardware RAID Redundancy

Do you have any other redundancy built in to this database?

- CNAF (Italy)
  - Yes. We are using a EMC CX3-80 storage system with Oracle ASM. The DATA partition is a 4 raid-1 disk group and the flash recovery area is made by 1 raid-5 disk group.
- Triumf (Canada)
  - The archive and online redo log files are replicated over 3 disks.
- IN2P3 (France): No
- SARA (Holland): RAID disks
- Rutherford (UK): No

Do you plan to move FTS to another database vendor? If so, which?

- CERN (Geneva): No
- PIC (Spain): No
- ASGC (Taiwan): No
- GridKa (Germany): No
- BNL (USA): No
- CNAF (Italy): No
- Triumf (Canada): No
- IN2P3 (France): No
- SARA (Holland): No
- Rutherford (UK): No

What are your plans for the FTS database?

- CERN (Geneva)
  - None
- PIC (Spain)
  - 2 node cluster in November 2007
- ASGC (Taiwan)
  - Yes. Installed a new FTS 2.0 server using new RAC database. Tests ongoing. Use Grid control and nagios monitored. Use increment backup everyday. If ready, will change to production very soon.
- GridKa (Germany)
  - Started to set up 3-node FTS/LFC Oracle RAC: 2 preferred FTS nodes (+ 1 preferred LFC node as standby FTS node). FTS DB storage (SAN): 4 x 140 GB, i.e. 140 GB for data + 140 GB for recovery + 2 x 140 GB for mirroring.

What are your plans for the FTS database?

- BNL (USA)
  - Migrate to a 2 node cluster configuration.
- CNAF (Italy)
  - None
- Triumf (Canada)
  - Near term: During the move to FTS version 2 we will purge the DB and start afresh.
  - Medium term: Probably move the DB onto an Oracle RAC after evaluation and presumably when adequate hardware would be in place.
- IN2P3 (France)
  - <no answer>
- SARA (Holland)
  - None. It has just been moved to new, dedicated hardware.
- Rutherford (UK)
  - Produce a test instance and probably migrate to a 2-node cluster.

Does the same DBA looking after the FTS and 3D databases?

- CERN (Geneva): No
- PIC (Spain): Yes
- ASGC (Taiwan): Yes
- GridKa (Germany): After starting up FTS/LFC on RAC, yes
- BNL (USA): The current 3D DBA will take over the FTS database after h/w migration
- CNAF (Italy): Yes
- Triumf (Canada): Yes
- IN2P3 (France): Yes
- SARA (Holland): No
- Rutherford (UK): No

LFC

Do you have LFC running at your site?

- CERN (Geneva): Yes
- PIC (Spain): Yes
- ASGC (Taiwan): Yes
- GridKa (Germany): Yes
- BNL (USA): No
- CNAF (Italy): Yes
- Triumf (Canada): Yes
- IN2P3 (France): Yes
- SARA (Holland): Yes
- Rutherford (UK): Yes

If "yes", what back-end database does it run on?

- CERN (Geneva): Oracle
- PIC (Spain): mySQL in Prod, Oracle in Test
- ASGC (Taiwan): mySQL
- GridKa (Germany): mySQL
- BNL (USA): n/a
- CNAF (Italy): Oracle
- Triumf (Canada): mySQL
- IN2P3 (France): mySQL and Oracle
- SARA (Holland): mySQL
- Rutherford (UK): MySQL and Oracle

Would you consider this database development, test or production?

- CERN (Geneva): Production
- PIC (Spain): Production
- ASGC (Taiwan): Production
- GridKa (Germany): Production
- BNL (USA): n/a
- CNAF (Italy): Production
- Triumf (Canada): Production
- IN2P3 (France): <no answer>
- SARA (Holland): Production
- Rutherford (UK): Production

If you have a production copy, do you also have a dev or test copy?

- CERN (Geneva): Yes
- PIC (Spain): Test copy in Oracle
- ASGC (Taiwan): No
- GridKa (Germany): No
- BNL (USA): n/a
- CNAF (Italy): Test generic RAC
- Triumf (Canada): Old database as VMware Guest
- IN2P3 (France): <no answer>
- SARA (Holland): No
- Rutherford (UK): Test schema on Prod

Is this database dedicated to LFC or does it share it with other schemas/applications?

- CERN (Geneva): Shared
- PIC (Spain): Dedicated
- ASGC (Taiwan): Dedicated
- GridKa (Germany): Dedicated
- BNL (USA): n/a
- CNAF (Italy): 3 Node cluster shared with FTS
- Triumf (Canada): Dedicated, same machines as application
- IN2P3 (France): Shared
- SARA (Holland): Dedicated
- Rutherford (UK): The MySQL server is shared with other schemas/applications, Oracle part of 3D

Is this database a cluster? If so, how many nodes?

- CERN (Geneva): 8 node cluster
- PIC (Spain): Single instance
- ASGC (Taiwan): Single instance
- GridKa (Germany): Single instance
- BNL (USA): n/a
- CNAF (Italy): 3 node cluster
- Triumf (Canada): Single instance
- IN2P3 (France): Yes for oracle, no for mysql
- SARA (Holland): No, two single nodes.
- Rutherford (UK): Single instance

What is the backup policy on this database?

- CERN (Geneva)
  - Weekly full, 6xWeek differential, 1xWeek Cumulative Archive logs everyhour. On Disk backup as well.
- PIC (Spain)
  - Full once a week, mon,tue,wed backup diferential...thu cumulative, fri and sat dif...archive logs each hour.
- ASGC (Taiwan)
  - Backup every week .
- GridKa (Germany)
  - bin-logs; replication; on master: daily diff-backup in TSM.
- BNL (USA)
  - n/a

What is the backup policy on this database?

- CNAF (Italy)
  - L0 backup plus archivelog every week. L1 backup plus archivelog every day. Recovery windows is 31 days
- Triumf (Canada)
  - Hourly full dumps of the LFC database, rsync-ed off to the backup server.
- IN2P3 (France)
  - Full by week, incremental by day.
- SARA (Holland)
  - Daily dump of th database to disk. The disk is backed up to tape.
- Rutherford (UK)
  - MySQL - full and incremental backup for /var/lib/mysql directory.
  - Oracle - full backup every week, incremental every day.

How is this database monitored?

- CERN (Geneva)
  - Home made scripts for DB level + Lemon monitoring for host level
- PIC (Spain)
  - Host with Ganglia, the database with OEM, a local one
- ASGC (Taiwan)
  - Service monitored. (e.g. use nagios)
- GridKa (Germany)
  - DB: MySQL Query Browser (DB host: nagios, ganglia).
- BNL (USA)
  - n/a

How is this database monitored?

- CNAF (Italy)
  - Via Grid Control
- Triumf (Canada)
  - The LFC database in not currently directly monitored. The LFC host is monitored via Nagios in a similar fashion as fts; ie. Disk space, host 'up-ness', ssh availability; Ganglia monitors load and swap, and Nagios also looks up the most recent SAM FTS functional tests.
- IN2P3 (France)
  - oracle tools for oracle hand made software for mysql
- SARA (Holland)
  - Scripts.
- Rutherford (UK)
  - MySQL - Nagios, Ganglia.
  - Oracle - Oracle’s Grid Control for database/host and nagios for host.

Science & Technology Facilities Council
e-Science

Are you using Data Guard on this database?

- CERN (Geneva): No
- PIC (Spain): No
- ASGC (Taiwan): No
- GridKa (Germany): No
- BNL (USA): n/a
- CNAF (Italy): No
- Triumf (Canada): No
- IN2P3 (France): No
- SARA (Holland): No
- Rutherford (UK): No

Do you have any other redundancy built in to this database?

- CERN (Geneva)
  - No
- PIC (Spain)
  - No
- ASGC (Taiwan)
  - No
- GridKa (Germany)
  - Replication (hot standby).
- BNL (USA)
  - n/a

Do you have any other redundancy built in to this database?

- CNAF (Italy)
  - yes. We are using a EMC CX3-80 storage system with Oracle ASM. The DATA partition is a 4 raid-1 disk group and the flash recovery area is made by 1 raid-5 disk group.
- Triumf (Canada)
  - System disk housing the DB is software mirrored; mdmonitor and smartd failures trigger email to our support group.
- IN2P3 (France)
  - No
- SARA (Holland)
  - RAID disks
- Rutherford (UK)
  - MySQL - RAID1 configuration

Do you plan to move LFC to another database vendor? If so, which?

- CERN (Geneva): No
- PIC (Spain): No
- ASGC (Taiwan): Yes, will plan move to new oracle rac db.
- GridKa (Germany): No
- BNL (USA): n/a
- CNAF (Italy): No
- Triumf (Canada): Not sure, but if we move the FTS DB to a RAC then we would probably want to move the LFC DB as well, and therefore to Oracle.
- IN2P3 (France): <no answer>
- SARA (Holland): No
- Rutherford (UK): MySQL -> Oracle - possible, depending on the success with LHCb LFC deployment

What are your plans for the LFC database?

- CERN (Geneva)
  - None
- PIC (Spain)
  - 2 nodes cluster in november 2007
- ASGC (Taiwan)
  - We will migrate data center to new place this year. In that time,will try to migrate LFC DB to Oracle RAC DB. Will use Grid control and nagios monitored. Use increment backup everyday.
- GridKa (Germany)
  - Started to set up 3-node FTS/LFC Oracle RAC: 1 preferred LFC node (+ 2 preferred FTS nodes as standby LFC nodes). MySQL DB to be migrated to Oracle. LFC DB storage (SAN): 4 x 140 GB, i.e. 140 GB for data + 140 GB for recovery + 2 x 140 GB for mirroring.
- BNL (USA)
  - n/a

What are your plans for the LFC database?

- CNAF (Italy)
  - None
- Triumf (Canada)
  - We would probably move to an Oracle cluster setup along with FTS, once the FTS service was moved and stable.
- IN2P3 (France)
  - <no answer>
- SARA (Holland)
  - None. Like the FTS it also has just been moved to new, dedicated hardware.
- Rutherford (UK):
  - Move from mySQL to Oracle if LHCb works well.

Does the same DBA looking after the LFC and 3D databases?

- CERN (Geneva): No
- PIC (Spain): Yes
- ASGC (Taiwan): Yes
- GridKa (Germany): After starting up FTS/LFC on RAC, yes
- BNL (USA): n/a
- CNAF (Italy): Yes
- Triumf (Canada): At this time, Yes
- IN2P3 (France): Yes
- SARA (Holland): No
- Rutherford (UK): Yes

FTS Developer View on Database Plans

- **We leave database design to the DBAs at the moment.**
- **We're planning to sit down with Miguel at CERN and actually *measure* what we use properly, so we can give better estimates to T1 DBAs**
- **Regarding the future – we're keeping more information now for service monitoring purposes. We'd like to process this data into summaries of various service related things, and we'd like to do that on the DB itself (with Oracle analytics) – so I'd expect an increase in data volume and CPU use as we start to introduce this.**
- **It's hard to estimate until we actually write it and benchmark it, but I think the message should be "err on the generous side" when planning your DB hardware purchases…**

Science & Technology Facilities Council
e-Science

LFC Developer View on Database Plans

- **For the LFC, our current estimate is about 1kB per file entry. This is probably a high estimate. So to store 10 million entries, it would be between 5 and 10 GB. It largely depends on the filename lengths (both the Logical File Name and the SURL or replica).**

Personal View

- DBAs need to forge better links with developers
  - Is working with FTS
  - Tools and tuning now in place
- DBAs should work together to tackle issues
- 3D good infrastructure for databases
  - community
  - experience in many areas
  - plenty of people to solve problems

Science & Technology Facilities Council
e-Science

Summary

- Database structure is in place for data management services
- Work more with developers
- 3D community/experience is helping with other database deployment areas
- Availability and resilience is key

Science & Technology Facilities Council
e-Science

Questions and (hopefully) Answers