WLCG FTS Service: database issues

Gavin McCance

CERN IT/GD

Issues

- Raised from:
  - Luis’ document
  - Various discussions on *fts-support* and *grid-service-databases*

Issues raised

- Application cleanup procedure
- Schema and tool versioning
- What’s coming soon?
- DB parameters and procedures
- Resource requirements

Cleanup procedure

- Luis’ document references the most expensive query – this is the query used to scan for pending jobs to be served.
  - It’s big because the job prioritisation is done inside the query

- If we don’t clean up the active table, it’s rather expensive – since we (index) scan millions of rows to find the few that are of interest
  - Finished jobs are not of interest to this query

Options

1. The ‘history’ tool: this moves terminal jobs over 7 days old to another table that is not scanned
   - This is what we currently have – but not as part of the proper release

2. Plan for FTS 2.(1) deployment
   - The schema now has a partitioning key (the time the job went into a terminal state)
   - Queries are being updated to use this
   - Partitioning should help
   - Even without partitioning, this index is highly selective
     - Should significantly reduce CPU and IO requirements for this query

History tool

- We’re stuck with this for a while
  - Until we understand the partitioning, test it and write the DB procedure for it
- 2 issues
  - Fragmentation – this was seen on the CERN-PROD RAC. Not fully understood. It causes problems upon schema upgrade.
  - Unbounded growth of history table
    - We currently never throw anything away

Retention proposal

- To solve the unbounded growth problem

- History table is ‘interesting’ for audit
  - Minimum WLCG audit requirement is 90 days
  - I propose another ‘tool’ to clean up the table after X days (with X defaulting to 90)
  - But data is useful for post-analysis of a LHC ‘run’, so we should coordinate with LHC running schedule

- Eventually we should drop the history table (sic) and use partitioning on the primary table.
  - This will come as a patch to FTS 2.0 (once we’ve tested it properly at CERN).

Dirty ‘tools’

- The primary schema versioning is controlled by YAIM and well managed (I think)
  - It‘s upgraded in line with the service release

- The ‘odd bits of PL/SQL’ and DBMS jobs that you get from *fts-support* are not well managed
  - Not versioned
  - Not controlled
  - Not properly certified

- We need to improve the latter
  - Proper versioning and schema checks
  - Better release process and cleaner procedures

What’s coming to the DB

- Next release: New schema for FTS 2.0 (you’ll need to upgrade this together with your FTS service admins)

➢ WLCG milestone for Tier1 sites to have upgraded by end-September

- CMS would like this at their Tier1s well prior to CSA07
  - ASGC, CNAF, FNAL, FZK, IN2P3, RAL, PIC
  - CSA07 runs from September 10th for 30 days
  - Expect release to be available < end July 2007

- DBMS jobs coming soon
  - The bug-fixed ‘history’ script
  - A new summarisation table with a row for every completed transfer (and the trigger to fill it)
    - This is to drive a Gridview monitoring plug-in requested by the LCG management board
  - A new ‘cleanup’ tool to prevent unbounded growth of these tables

What else…

- The FTS monitoring framework now stores much more in the database
  - This is necessary for the stable operations of the WLCG transfer service

- This monitoring processing will be driven from within the database
  - Not hundreds of perl scripts connecting every minute
  - This is the other reason we need to ‘regularise’ the deployment of all these little bits of PL/SQL
  - Expect CPU increase as we make use of Oracle’s analytic functions

Improved procedures

- The DB is part of the overall WLCG FTS service
- We can make available our FTS service administration procedures (e.g. service upgrade)
  - These involve more procedural cooperation between DBA and service admins for general service maintenance, e.g. service upgrade
- The DB will also be running more of the application (monitoring)
  - It’s not just where we keep the application’s state

- Expect to have more coffee between the DBAs and the FTS service admins ☺
  - General trend for stronger integration of DB ops with Grid-site ops…
  - You too may want to know things that are discussed at Grid meetings, such as Grid Deployment Board, WLCG collaboration workshops, weekly joint operations meetings etc.
  - This is where you learn about schedules, interventions, new versions, problems etc.

DB parameters?

- This is DBA question
- 3D can advise: block-size, memory, cache size, redo log parameters
  - …and can translate the benchmarks of the application to what you need in terms of hardware

- AFAIK, we ~happily run LFC, FTS, Gridview and VOMS on the same RAC with the ~same settings
  - But 3D can advise on this

Backup policy

- 30-day “flashback” (or otherwise) retention is not needed
  - If the schema becomes logically corrupt, we start from a fresh schema

- You need to be able to recover the DB to when it failed
  - i.e. a full standard recovery
  - In the (bad) case of a partial recovery (e.g. only to the last backup), an additional application procedure is needed before the service can go back into production
    - To avoid the “replay” of previously “Done” transfers
    - We will define this and make it available

Resource requirements

- We’ll work with DB team at CERN to determine these out better
  - The ‘cleanup’ should prevent unbounded growth, so we should reach a steady state
- Expect core-application CPU requirements to decrease as queries becomes more efficient (when we move to partitioning)
- But.. expect CPU requirements to increase as we deploy more service analytics in the database to improve the (poor) service monitoring situation

Process

- We will test all the ‘features’ on the validation database RAC at CERN first
  - This benchmarks should be made available as soon as we have them, so you can update your planning
  - The database advice should be integrated more closely with the rest of WLCG operations
  - N.B. we can’t benchmark what we don’t yet have

Final remarks

- We'll provide clearer documentation for FTS
  - Including pointers to the 3D "advice" pages
- The WLCG operations group (together with 3D) will provide regular updates on the status of the FTS application
  - New things coming – benchmarks of new monitoring
  - Any updates to DBA recommendations
    - Procedural changes
    - DB setup and deployment parameters

- The DB is a critical part of the *service*
  - Expect more interaction with the FTS service admins at your site
    - as we make the database do more for us
    - and as we integrate the database more closely with our service operational procedures
  - Expect closer integration with general WLCG operations infrastructure