

Bookkeeping meeting, January 28th 2008

Present: R.McNulty (by phone), Ph.Charpentier (chair, minutes), E.Lanciotti, Z.Mathe

Apologies: M.Bargiotti

Goal of the meetings

These regular meetings (Mondays at 16:00) are intended at reviewing the work to be done and being done on the bookkeeping system. This includes maintenance and short term fixes on the production system as well as new developments for the new system.

Hardware setup

The existing system uses volhcb05.cern.ch as BK server, that accesses the BKDB via AMGA running on lxarda01.cern.ch. This machine has to be retired and the service has to run somewhere else. It is proposed to run the service on volhcb05 itself as it uses very few resources.

Action (Marianne): see with Birger how to install the service on volhcb05

For the new system, Marianne has obtained a DB instance from IT. She has set up the BK services on volhcb07 as well as an AMGA service in Glasgow. The main problem today is to figure out how to create the tables in the DB. It seems the tables are created directly with SQL command, and Elisa found a .SQL file that looks as if it is the one to use. There is however some hardcoded information for indexing the tables that look pretty strange as they concern old data. Is there no partitioning for DC06? Elisa will continue investigating with Marianne. The goal is to have the development system work with old code, old schema and old data before moving to a new schema or new software...

Maintenance of the existing system

There are two pending issues:

- Java crashes for some stripping event types. it seems linked to the Java AMGA interface as the same query is OK with the python API (Elisa's script). Elisa will ask Birger whether he can have a look
- Duplicate entries of files in the views: if a file has ancestor jobs that were using two versions of a given program, this file is entered in the two JOBFLEINFO corresponding to the two versions. The consequence is that the statistics reported in the BK web page is incorrect (double counting of these files). The Gaudi cards generated however give the correct list and correct number of events. A fix would be to enter each file only once (e.g. with the first version, or with a combined version e.g. Brunel v30r14/v30r15 to indicate there is a mixture). Not only the web page is affected but also the script Carmine gave to the physics coordinator and a few others report that number of events (which is the sum of the events in each matching roottree row).

Items for the new system

Three main threads were discussed concerning the new system

Setting up a test system

Elisa reported the efforts to set up the test system in a first instance with the current schema. The main initial problem is to understand how to create the initial tables. There are no instructions, but Elisa and Marianne found an SQL file that can correspond to this. Looking at the code it was not easy to understand why index

partitioning is implemented with a predefined segmentation (hardcoded) that seems to correspond only to DC04. This cannot scale when new files/jobs are registered... but to start with we can try this out. It is then necessary to migrate a small fraction of the DB (test DB is smallish), e.g. starting from a simulation production and all its descendants.

Action (Marianne and Elisa): understand how to create a DB schema (first warehouse, then how to re-create views). Get a BK interface work on the test system.

Review of the DB schema

The current warehouse schema is a translation of the former schema, but the flexibility of the (*key, value*) pairs has been lost. The columns in the *params* tables are just an "OR" of all the *key* values that were found at the time of the migration. It makes it quite difficult to add new parameters if necessary. It is therefore proposed to include all columns that are needed for creating the views in the */jobs* and */files* tables rather than leaving them in a separate table that has as many entries as the mother table.

The provenance parameters should remain in the *params* table, and we propose to go back to a (*key, value*) scheme. This would allow an easy display of all parameters in the web page by a generic code.

Action (Elisa): prepare a proposal for the */jobs* and */files* tables, as well as for a list of *key* values necessary for converting the existing entries. This proposal should take into account requirements for real data (such as identifying a "job" to a "run" for RAW data).

As suggested by the BKWG, two view schema should be defined separately for real and MC data, and new tables added in order to define in particular "data taking periods" as well as "processing passes". The existing MC views schema should be reviewed in order to address the current problems (limited number of programs in history, multiple versions in the history).

In parallel with the definition of a new schema the necessary script should be written for migrating the existing information to the new schema. One should check which data should be migrated (e.g. is DC04 data still useful?)

Action (Philippe): check with physics groups on the need to keep any non-DC06 data in the BK

A well specified API should be defined in order to implement BK access through a DIRAC service (see next point). This API must set the baseline for all higher level usage of the BK, either for administrating it (BK manager) or querying it. Elisa already defined part of the API for implementing her prototype python interface.

Action (Elisa): document the prototype API and classes, to be reviewed and augmented in order to accommodate the use cases (to be explicitly listed as well)

Architecture of the BK services

The BK service should be integrated in the DIRAC3 framework. The BK services should be handled in a DIRAC service that will interact with the database (through AMGA). The client interface will use this service as a proxy to the DB.

Action (Zoltan): get documentation on how to set up a DIRAC service and client from DIRAC developers. Implement a dummy server/client, and follow soon with porting Elisa's prototype as a service.

The question of CVS repository and access was discussed and since then Andrei has set up the structure.

Next meetings

Wednesday 31 January if Marianne available

Monday 4 February @ 14:00 (note unusual time) -- PhilippeCharpentier - 28 Jan 2008

This topic: LHCb > Meeting080128

Topic revision: r4 - 2008-01-29 - PhilippeCharpentier



Copyright &© 2008-2019 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

Ideas, requests, problems regarding TWiki? Send feedback