

## Bookkeeping meeting, December 6th 2007

Present: R.McNulty, T.Kechadi, Z.Mathe (Dublin), M.Bargiotti, Ph.Charpentier (chair, minutes), B.Koblitz, E.Lanciotti, A.Maier, S.Paterson, R.Santinelli, A.C.Smith (CERN), A.Tsaregorodtsev (Marseille, on the phone)

The aim of the meeting was to set the scope of the project, review the current situation and existing developments and define the topics of work. The agenda can be found here [?](#)

### Current bookkeeping system (Marianne)

Marianne presented the current status of the system [?](#), which is also in slides from Carmine Cioffi.

There were discussions on the understanding of the actual content of the views (*roottree* and *JobFileInfo* tables) that lasted until after I released the summary ;-). I hope it is now correct... Each *JobFileInfo* table contains a list of files that have the same parameters in terms of queries (Config, evtType, fileType, program version). *roottree* contains as many rows as there are *JobFileInfo* tables and is used by the query in order to match the selection. If more than one row of *roottree* matches, the union of the *JobFileInfo* tables is used.

AMGA is used as an access layer to the actual Oracle database. It implements authentication as well as an insulation from the actual backend. The actual BK schema is based on the generic AMGA schema.

Marianne reviewed the services that form the BK system:

- BkkReceiver: receives XML files from jobs, already using the DISET transport. Currently the implementation for storing the queries is based on the file system. *This should be ported to a RequestDB within DIRAC3*
- BkkManager: uses the *NewConfirm* servlet which checks the consistency of the XML file, in particular if referring to jobs not yet entered in the BK, the XML file is put in hold (in a special directory). This request will be treated after all requests have been treated. The main role (besides checking for inconsistent or missing info) is to sort the entries in the order in which they should be inserted. *In the DIRAC3 framework, this should become an agent* BkkManager inserts the data in the warehouse and is responsible for the views creation every night. *Is this all in a single script?*
- Tomcat servlets: they are responsible for giving access to the browsing web page
- FileCatalog: XMLRPC service that integrates BK and LFC for getting replica information (used by the BK browser web page and stand-alone client that creates XML catalogs or converts Gaudi options from LFN to PFN).
- BkkMonitor: verifies the other services are alive, triggers alarms otherwise

### Bookkeeping Working Group (Philippe)

Philippe presented the same slides he had prepared for the May '07 Software week [?](#) that contains the main recommendations of the BKWG. The BKWG was formed during Spring '07 in order to review the current BK from a user's perspective and identify issues related to real data.

The current warehouse schema seems adequate for real data if a DAQ run is assimilated to a *job*, with several output files. MMinor adaptations might be necessary.

Several new features or concepts for querying the BK for real data need to be introduced (query by date, data

taking period, processing pass). Periods and passes can be (re-)defined a posteriori, hence should be expressed in terms of the query criteria. The definitions could be additional tables or any other source of information (implementation detail).

One of the requests from users is to introduce flexibility in the queries compared to the current rigid web interface. One should be able to chain criteria in any order and get the choice of possible other criteria according to the already selected ones. There is a need to have a browser that can be integrated into ganga and also run in stand-alone mode for producing job options. The web browsing capabilities are not considered as a "must". The WG therefore recommends to implement in priority a stand-alone browser (in python and pyQT for the GUI). The web interface could be added if easy to maintain. The web interface should however be used for providing statistics (e.g. for management of productions).

After considering the prototype browser developed in Dublin within *feicim*, which presents search criteria as a file search tree, the WG was in favour of such a presentation to users. The default behaviour would however basically reproduce the current behaviour (with the caveat of replacing application versions by "processing pass" for non experts). It is also clear that the browsing criteria are different for real data and MC.

## Current new developments (Elisa)

Elisa presented the status of the work she had been doing (slides<sup>☞</sup>) in the past month.

She has developed a python class that allows querying the BKDB directly through the AMGA interface after having prompted the user for his search criteria (in any order). The interface discovers the allowed values for the next criterion, from the previous ones. Checking this method against the web interface+servlets is giving identical results. When asking for files replicated at a given sites, there is one more replica found which pointed to a bug in the web implementation (one file missing)...

There was a long discussion on how to evolve this prototype. One possibility would be to implement the interface as in the prototype directly in the client, while the alternative is to turn it into a DIRAC service that would be interrogated by a DIRAC API client within the framework. The direct implementation is faster for testing while the service implementation is much neater and allows a clear definition of the interface, that could be stable under various implementations.

It was finally decided to first define this interface, then prototype it in a client which then should be made a DIRAC service together with a light client API. It was suggested by Andrew (outside the meeting) that in case the functionality of the interface is similar to that of the generic DIRAC File Catalog interface, the same method and signature be used. This would simplify the implementation (for example for removing a file).

## Short term plans

Elisa will be away (conference) until Christmas break. Zoltan would possibly move to CERN in January.

They would work on defining the interface and providing a prototype (Elisa for the query methods, Zoltan for a GUI starting from what he had developed for *feicim* but in python this time). In parallel they would familiarise with the DIRAC framework.

Marianne will re-implement the manager and related tools in python.

A machine will be dedicated for installing test services (volhcb07). Marianne will enquire whether we still have a test instance of the DB and a connected AMGA service that we can use for non-destructive tests. Philippe (with Roberto) will ask for a PPS AMGA service for this purpose. After a suitable PPS stage, this service could be declared "in production" and be used by the production BK while the AMGA service on arda01 would be used for tests.

## Meeting071206 < LHCb < TWiki

Tahar declared he was very much interested in following the topic as well as participating in further architectural discussions. Philippe said that this collaboration should not necessarily be restricted to BK (most urgent need) but hopefully will enlarge to other domains of distributed computing.

Thanks to everybody who participated! Season's Greetings!

-- PhilippeCharpentier - 10 Dec 2007

---

This topic: LHCb > Meeting071206

Topic revision: r5 - 2007-12-12 - 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