Belle II CDB migration review July 2024 - General notes
July 11 2024, online only

Charge

Charge for review of Belle II conditions database migration
David Jaffe

The review committee is requested to perform an assessment of the basis and plan to migrate the Belle II conditions database (CDB) including addressing the following specific questions:

1. Have the requirements for the Belle II CDB for the foreseeable future been clearly defined and understood by the CDB team?
2. Have the concerns and needs of the relevant Belle II stakeholders, particularly the data production and software groups and users, been taken into account in the design and proposed implementation of the new CDB?
3. What process and consideration of options for a new CDB has led to the proposed solution?
4. Does the proposed new CDB satisfy all requirements? Does the proposed CDB rely on established and supported software and hardware?
5. Is the migration plan technically feasible, sound and reasonable? Does the migration timeline have sufficient contingency? Are the milestones clearly defined and achievable in the specified time given the proposed effort and resources?
6. Are there any deficiencies or shortcomings with the new CDB or the migration plan?

The review committee is expected to provide a report containing their findings, comments and any recommendations.

This review is being undertaken at my initiative as US Belle II Operations Manager because the CDB is a US(BNL) responsibility.

Committee

Wouter Deconinck, Paul Laycock, Chris Pinkenburg, Torre Wenaus (chair)

Notes

Outstanding questions

1. What is the maintenance plan that accompanies the new CDB implementation? Which institution is responsible? What is the plan for training and integration of a broad enough group of experts on the required tools? How will updates be propagated from development to production?
   a. BNL responsibility. BNL is the US lead lab, will remain in Belle II with this responsibility unless US leaves Belle II (or we bomb on a review at the level PNNL did)
   b. BNL is making a hire that will restore dedicated HSF ref implementation client effort (replacing Lino)
   c. good suggestion that the Belle II client specifics would be good to have as a BNL responsibility, we should discuss
   d. plan to include CDB training in HSF training (we are involved in HSF training WG)
2. Do you understand the GT cloning use case? This was mainly used for making a "backup" of a GT before the Data Production manager would make a complicated update to the GT. Because the same GTs are used for specific workflows ad infinitum, they get very big, this could also be worth discussing with the Data Production / software experts (the GT provenance mechanism relies on these GT names not changing, which is .. sub-optimal).
   a. CERN R1 coffee discussion
   b. Chris invited to the meeting, has ideas
   c. not clear there is a single point expert on all workflows
3. Can you compare the demonstrated performance parameters (in particular response frequency) with expected requirements for the Belle II program over the next 5 years of operation? There are values about the number of GTs and IOVs in the backup slides of the first presentation, but no values about the job load that is expected.
   a. addressed. Don't expect growth in requirements in next 5 years. Have headroom via caching, horizontal scaling etc. sPHENIX experience suggests Belle II won't hit limits. Also working on improving for sPHENIX.
4. It would be good to understand why you have 34 (?) calls to the CDB per job, to be more sure that your assumptions about extrapolating to future needs are valid.
   a. "how many calls a regular job makes? No one knows." Michel is investigating. Strange duplication of calls. Caching seemingly not always used when it shouldbe. He is digging more deeply, but need someone to dedicate time.
5. Identifying a person on the basf2 side to make a prototype (it's relatively little code) should be a priority
6. Validation is complicated by the current behaviour using "version number" - validation needs more attention such that you can agree with Belle II when the new CDB is considered to be ready for production. Help from the Data Production team is needed here.