B-factory Programme Advisory Committee
Focused Review on Software and Computing
Short Summary

G. Carlino (Naples), G. Cerminara (CERN), M. Elsing (CERN),
F. Gaede (DESY), A. Formica (Saclay), W. Hulsbergen (NIKHEF),
M. Ishino (Tokyo), A. Krasznahorkay (CERN), P. Mato (CERN),
P. McBride (FNAL), H. Tajima (Nagoya)
and chaired by T. Nakada (EPFL)

10 July 2016

A focused review on the offline software and computing for the Belle II experiment took place at KEK on 27th and 28th of June 2016 by a review committee that consists of three members of the Belle Programme Advisory Committee and nine experts. Reconstruction, simulation and data analysis software and database were covered on the first day and data processing and resources on the second day. In this short report, only essential general comments are made and a comprehensive report on the individual subjects will follow.

Development and test of various individual components for simulation, event reconstruction, databases and analysis are well advancing. On the other hand, the committee feels that less attention is being paid to global software coordination and integration aspects. Overall stress tests of the whole reconstruction chain, including calibration and alignment, are missing. The reconstruction time is the major driver for the CPU needs of the experiment and further optimisation work for the software and event data model is needed.

The preparation of framework for the Belle II distributed computing is progressing well. However, essential missing functionality needs to be implemented and the scalability of the production and data management tools up to full Belle II production levels must be demonstrated. The level of automatisation achieved in the production and data management system is also a little concern. The analysis model must cope with a huge number of events efficiently and within available resources. Although those issues are addressed individually by the various software and computing subgroups, problems are tightly connected and should be examined globally. In addition, the overall architecture of the software and computing should also be revisited in order to avoid any builtin limitation.

The committee heard little about the connection to the online subsystem, e.g. database relation and High Level Trigger (HLT). If the HLT makes event selections, the software

implementing the algorithms must be of extremely high quality and practically free from errors and memory leaks. The selections must be offline reproducible in order to measure the efficiencies. Any event reduction there changes the resource requirements for offline. The HLT is not crucial for the Cosmic Ray run nor for the BEAST Phase 2 run. However, a clear communication line with the online project must be established now.

In general, the committee misses the overall planning of the software and computing project. A plan must be established now that clearly defines a performance goal with tasks, milestones, and available and required resources for the Cosmic Ray run, BEAST Phase 2 run, Phase 3 physics run. Expertise and experience of people contributing to the project should be well balanced and matched to the needs. The plan should cover not only the software related to reconstruction and analysis, but also all the infrastructure needed for data processing as well as the HLT. For this purpose, the committee recommends the Belle II management to bring the core software group and subsystem software groups together through regular meetings and to consider introduction of a person having a clear mandate and authority to coordinate the effort and to monitor the progress. This would allow to optimise resources and to set priorities from a more global point of view, as well as to ask the Belle II management necessary resources in a coherent manner, if needed.

The committee was presented with a draft Memorandum of Understanding (MOU) for a share of CPU power and storage to be provided by the participating funding agencies. Furthermore, opinion was asked on the projected resource requirements until 2024. The committee fully supports that such an MOU is essential to ensure the required computing resources. The initial resource requirement for the next two to three years reflects the needs estimated conservatively based on the current status of the offline software and the current plans for the computing and analysis model. The ramping up of the machine luminosities at a startup period is always difficult to predict. The current resource requirement is also subject to necessary improvements in the reconstruction software, as well as to eventual changes and refinements in the analysis model. However, the committee notes that resource requirement for the initial phase will be in any case limited. The long term evolution of the Belle II computing resource needs will follow the evolution of the integrated luminosity delivered by SuperKEKB, which has large uncertainties. One should also expect continuous improvements in the offline software and evolution of the model for distributed computing that takes into account the whole lifetime of the experiment, which are difficult to quantify now. The committee suggests to follow the LHC computing resource model and to restrict the discussion on the concrete resource requirements to a period of the coming two to three years, keeping in mind the overall view on the longer term evolution.