

Belle II Computing Resource Review

6 – 7 December 2016 at CERN

G. Carlino (Naples), W. Hulsbergen (Nikhef), P. Mato (CERN),
and chaired by T. Nakada* (EPFL)

* Chair of the B-factory Programme Advisory Committee

21 December 2016

In the review, various computing and software issues, which are necessary to understand and judge the validity of the Belle II computing resource requirement, were presented. Here are the comments and questions which should be taken into account in the final resource requirement document to be submitted to the B-factory Programme Advisory Committee meeting in February 2017.

Organisation

Organisation for the computing and software has been greatly improved since the last review meeting in June. Coordination between the off- and on-line software, computing, DAQ and physics has received a particular attention and the committee encourages further development in this direction.

The committee would like to see the overall project plan for the computing over the next two to three years in order to understand the relation between various activities and to judge whether there is any items on the critical path.

Data model

The committee understood that the persistent RAW data will be stored in the form of ROOT objects in ROOT files in order to benefit from the data compression this format provides. Reading very old RAW data files with newer ROOT versions could be problematic. Since long-term availability of data is a concern, the committee suggests to consider a custom binary format with zero-suppression plus optimal encoding or standard compression algorithm. Creation of ROOT objects can be done as the first step in the data processing. Note that LHC experiments have adapted this strategy.

In the presentation, it was noticed that there was no explicit mentioning of the alignment and calibration data flow.

In the Mini- and MicroDST, no information on the machine background condition is stored. Some of the analyses, with neutrinos in the final states for example, could be sensitive to the level of the machine background. Therefore, including some information on this in the DST's might be useful.

Luminosity evolution

While the luminosity evolution (and resource requirements) are given as a function of the calendar years, the Belle II run period is from the autumn to early summer. The committee suggests that the luminosity evolution and all the process to estimate the resource requirements are described following the run period, while the final numbers to be presented to the funding agencies are transformed to calendar years.

Processing time

The committee noticed that the comparison of the processing times for event reconstruction between the Belle II and LHCb shows a significant improvement potential for the Belle II experiment.

Dealing with the background uncertainties

In the discussion of the size of simulated events, only the pixel detector occupancy is changed from 1 % to 3 % in order to take uncertainties in the machine related background into account. The committee wonders whether the silicon strip detector and electromagnetic calorimeter which have similar occupancies to that of the pixel detector should also be treated in a similar way.

In the process time estimation for simulation, the committee thinks that the gain from using real data for the machine related background hits in the detector is underestimated.

Computing resource requirement for 2018

The committee understands the physics argument for the required amount of simulated events for the analysis of the Phase 3 physics data taken in 2018. On the other hand, the committee fails to see the reason for producing additional 3 ab^{-1} of generic events in 2018, which will not be used for physics analysis. For testing the computing model and production framework, the large amount of simulated data, including 5 ab^{-1} of generic events planned to be produced in 2017, should be used rather than producing a new data set. Along the same line, production of 2 ab^{-1} of simulated data for "MC production rehearsal" is ill motivated. A project plan recommended above might help to understand better.

No explanation is given for the Phase 2 data related resource request. In particular the large amount of simulated data must be well justified by a clear programme.

In the resource planning 2018, no change in the amount of data for the Phase 3 in 2018 is foreseen although the Phase 2 run is now postponed to 2018. The committee recommends to revisit the Phase 3 related resource requirements in view of this schedule change. It must be noted that production of a large sample of simulated events for physics analysis should be launched after the detector performance is well understood.