B-factory Programme Advisory Committee

Review of the 2024 - 2027 Offline Computing Resource Requirements

Sub-Committee for Computing Resource Request
G. Carlino (Naples), P. Mato (CERN), P. McBride (FNAL),
W. Hulsbergen (Nikhef) and chaired by T. Nakada (EPFL)

4 August 2023

1 Overview

The computing and storage resources needed for the Japanese Fiscal Years (JFYs) 2024 to 2027 have been presented by the Belle II collaboration at the annual B-factory Programme Advisory Committee (BPAC) meeting in February 2023, and updated in a documentation distributed in May 2023. The estimates were produced by the Belle II Computing Steering Group and approved by the Executive Board. The resource request was supported by a detailed accounting document reporting the analysis of computing usage and the pledges by member states in JFY 2022.

The sub-committee for the Computing Resource Review met on May 24th 2023 to review the request. The questions following from the review were answered in a document distributed on June 13th. The review committee met again on June 30th to discuss the answers and evaluate the request. In this meeting, the committee concluded to recommend the resource request for approval. This decision is motivated in this report.

The resource estimates are based on the foreseen integrated luminosity, the number of simulated events requested by the physics analysis groups and computing model parameters such as performance estimates from the most recent releases for reconstruction and simulation software. To understand the reliability of the projections past requests can be compared to actual usage. Table 1 shows request, availability and actual use of resources in JFY2022. The provided CPU consists partially of so-called opportunistic resources, which are not reserved for the Belle II experiment.

	requested (as estimated in 2021)	pledged	provided	used
tape [PB]	8.8	8.8	7.4	7.4
disk [PB]	16.5	16.6	15.6	11.5
CPU [kHEPSpec]	385	385	853	361

Table 1: Requested, pledged, realized and used computing resources for JPY2022.

With the expertise gained after several years of running, the evaluation of the estimates has become a well-established routine. The largest uncertainty is the luminosity. The SuperKEKB luminosity profile for the coming years, adjusted to the latest projections, is summarised in Table 2. Computing and storage needs in the years 2023 to 2026 have been separately estimated for the different activities, namely prompt processing and calibration of real data, reprocessing of the real data, production of simulated events, skimming of the real and simulated data, and physics analysis. Important computing model parameters are the event sizes and the processing times. These parameters have been measured with the latest software release for different classes of events and both for simulated and real data.

JFY	2023	2024	2025	2026	2027
	Apr 2023	Apr 2024	Apr 2025	Apr 2026	Apr 2027
period	Mar 2024	Mar 2025	Mar 2026	Mar 2027	Mar 2028
by year [ab^{-1}/year]	0.06	0.35	0.5	0.7	1.2
cumulative [ab^{-1}]	0.55	0.9	1.4	2.1	3.3

Table 2: Expected SuperKEKB integrated luminosity $\int \mathcal{L} dt$ in ab^{-1} for JFY 2023 to 2027.

The total resource requirements for the JFYs 2024 to 2027 are summarised in Table 3. It is assumed that the computing activities are evenly distributed over the year. Losses due to inefficiencies of the sites, experimental software and middleware for the distributed computing have been taken into account. This request represents an approximately 30% increase in CPU and disk space for 2024 compared to 2023.

JFY	2024	2025	2026	2027
	Apr 2024	Apr 2025	Apr 2026	Apr 2027
period	Mar 2025	Mar 2026	Mar 2027	Mar 2028
Tape [PB]	9.6	11.8	14.8	20.1
Disk [PB]	25.0	31.9	39.6	49.3
CPU [kHEPSpec06]	520	465	464	519

Table 3: Estimated requirements on computing resources for JFYs 2024 to 2027.

2 Evaluation and recommendations

The committee would like to express its appreciation for the detailed documentation provided by the Belle II collaboration. The JFY2022 accounting report gives an excellent overview of last year's usage. The requested resources for 2024 are well motivated. The largest uncertainty is the expected luminosity.

Based on its evaluation of the request the committee makes the following recommendations to the Belle II collaboration:

- The accounting report for 2022 shows that as much as 30% of the used CPU resources came from local resources at KEKCC, DESY, and BNL. It is not

clear whether they were all part of the national pledges or not. The local CPU usage is split between production and for local analysis groups. Those resources used for production should be accounted for in the request and pledges. The committee recommends those local resources to be categorised in a transparent way.

- The accounting report shows that not all national facilities deliver their pledged resources. These national facilities are probably inefficient, due to poor connectivity or manpower. In previous years this did not lead to problems because the experiment used less resources than requested and because some national facilities provided a substantial amount of opportunistic resources. The latter cannot be counted upon for the future. Therefore, the committee encourages the collaboration to develop a model that does not rely on the incidental opportunistic resources. The collaboration needs to decide what to do with national facilities that are not fulfilling their commitments. A mitigation plan should be developed.

- The Monte Carlo requests includes both run-independent (riMC) and run-dependent (rdMC) simulation. During the June BPAC review, the committee learned that currently only 20% of analyses are using the rdMC. The Belle-II physics coordination is strongly encouraging the use of rdMC and phasing out as much of the riMC by spring 2024. The committee strongly supports this strategy. Future computing requests should reflect this change in the usage of MC.

- The collaboration has indicated that the analysis skims were not efficiently used. The majority of the analysts still run on the full mDST samples. Running analyses on mDSTs is less CPU efficient and not according to the analysis model. The committee recommends that the collaboration investigates why the skims are not used. Eventually, it may be necessary to put measures in place that to discourage users to process the mDSTs.

3 Conclusions

The computing and storage resources requested by the Belle II collaboration for the activities foreseen in JFY2024 are well motivated. They present a mild increase compared to resources provided in JFY2022. The committee recommends that the new requests will be granted by the Funding Agencies.

As in the past, the projections for the future still have large uncertainties, in particular due to uncertainties in SuperKEKB machine performance. Therefore, the committee recommends that the resource requirements remain to be updated and reviewed on a yearly basis.