B-factory Programme Advisory Committee

Review of the 2023 - 2026 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)

11 October 2022

1 Overview

The computing and storage resources needed for JFY2022 to JFY2026, where JFY stands for the Japanese Fiscal Year starting on April 1st and ending on March 31st year after, have been presented by the Belle II collaboration at the annual B-factory Programme Advisory Committee (BPAC) meeting in February 2022 and elaborated in a document distributed in April 2022. The estimates were produced by the Belle-II Computing Steering Group and approved by the Belle II Executive Board. The resource request was supported by a detailed accounting document reporting the analysis of computing usage and the pledges by the Belle II member countries in JFY2021. The BPAC sub-committee for the computing resource review met on May 5th 2022 to review the request. The questions following from the review were answered in a short document distributed a few days later.

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 of 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 JFY2021 .

	requested	realized availability	usage
tape [PB]	3.4	7	7 (status on 2022/01)
disk [PB]	11	13.8	9 (status on 2022/03)
CPU [kHEPSpec]	498	486	292 (extrapolated 2021/04-2022/01)

Table 1: Requested, realized and used computing resources for JPY2021.

The SuperKEKB luminosity profile for the coming years, adjusted to the latest projections, is summarised in Table 2. Computing and storage needs for JFY2023 to JFY2026 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 6 for different classes of events and both for simulated and real data.

JFY	2022	2023	2024	2025	2026
period	Apr 2022 Mar 2023	Apr 2023 Mar 2024	Apr 2024 Mar 2025	Apr 2025 Mar 2026	Apr 2026 Mar 2027
by year [ab^{-1}/year]	0.18	0.19	0.44	0.44	0.69
cumulative [ab^{-1}]	0.48	0.67	1.11	1.66	2.35

Table 2: Expected SuperKEKB integrated luminosity $\int \mathcal{L} dt$ in ab^{-1} for JFY2022 to JFY2026.

The total resource requirements for JFY2023 to JFY2026 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.

Year	Apr 2023 Mar 2024	Apr 2024 Mar 2025	Apr 2025 Mar 2026	Apr 2026 Mar 2027
Tape [PB]	8.6	11.0	13.3	16.3
Disk [PB]	19.6	24.8	28.3	31.3
CPU [kHEPSpec06]	404	495	340	308

Table 3: Estimated requirements on computing resources for JFY2023 to JFY2026.

The pledges by the national facilities for JFY2022 were 8.8 PB of tape space, 16.5 PB of disk space and 385 kHEPSpec of CPU. Therefore, the request for JFY2023 represents an approximately equal amount of tape space and CPU time, and about a factor 1.2 more disk space.

2 Evaluation and recommendations

The committee would like to express its appreciation for the detailed documentation provided by the Belle-II collaboration. The JFY2021 accounting report gives an excellent overview of last year's usage. The committee is also pleased to see that, contrary to earlier years, the pledges from the national facilities matched well with the requests.

As shown in Tab. 1, CPU usage in JFY2021 was smaller than predicted. This was in part because the experiment collected less data than foreseen. The decrease in the number of events was partly compensated by a larger event size and an increase in the MC production. The usage of tape storage was larger than expected, mainly because the HLT was run in pass-through mode for most of JFY2020. These events were eventually moved to tape in 2021. There was also more space than foreseen needed for calibration samples.

Compared to last year's predictions, the estimated integrated luminosity for JFY2023 and beyond has been scaled down considerably. This also affects the estimated required computing resources for those years. It illustrates that the long

term needs are still difficult to predict. However, the committee has concluded that the current estimates for next year are solid.

The production of Monte Carlo simulated events represents a large part of the requested resources. The number of required events is estimated based on input from the physics working groups. For optimal performance the physics analyses require so-called run-dependent Monte Carlo, which relies on calibration parameters obtained from the data. Currently, these are only available several months after the data taking. As a result, the physics groups also need large samples of run-independent simulation. As these events will eventually be replaced by run-independent Monte Carlo, this leads to inefficiency. The committee recommends that the collaboration continuous their effort to minimize this inefficiency.

3 Conclusions

The committee finds that the computing and storage resources requested by the Belle II collaboration for the activities foreseen in JFY2023 are well motivated. They present a mild increase compared to the resources provided by the member countries in JFY2022. The committee recommends the JFY2023 request to be granted by the funding agencies.

As in the past, the projections for the years beyond JFY2023 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.