BELLE2-NOTE-TE-2023-004
Version: 2.1

Belle II Computing Resource Estimate for 2024-2027

The Belle II Computing Steering Group

February 20, 2023

1 Introduction

In this note we present an estimate of the computing resources needed by the Belle II collaboration in Japanese fiscal years (JFY) 2024 to 2027 and describe the planned activities and the assumptions that lead to this estimate. This estimate supersedes the one presented at the June 2022 BPAC review. Compared to the estimate presented in June 2022, this revised estimate is based on an updated luminosity profile for JFYs 2024 to 2027 and an updated HLT trigger menu for the JFY 2023 running period. In the following, dates refer to JFY unless otherwise stated.

We first present an outline of the Belle II computing model followed by a description of the planned computing activities, then we discuss the input parameters used for the resource estimate and finally we present the resource estimate for years 2024 to 2027. Appendix A shows a breakdown of the resource estimate contribution of each of the planned activities.

2 Outline of the Belle II computing model

Belle II uses a three-level hierarchical structure of computing sites: raw data centers, regional data centers, and MC production centers. Raw data centers are also regional data centers and MC production centers. Regional data centers are also MC production centers.

The raw data coming out of the online system are permanently stored, calibrated and processed at raw data centers. The fully reconstructed events, produced from the raw data processing step, are stored in mDST format, which includes the minimum information necessary for analysis. Monte Carlo (MC) events are simulated and reconstructed using the same software used to process detector events and then also stored in mDST format. Detector and MC mDST are stored in regional data centers.

Detector and MC mDST samples are then “skimmed” at regional data centers to create a sample of selected events that suit a specific set of physics analyses. During the skimming step, additional information is computed and then added to the events. The output of the skimming step consists of deep copies of events stored in uDST (user DST) format, which includes mDST information plus analysis level objects, that are stored at regional data centers.

The understanding of the detector and the quality of the software will improve over time, resulting in new releases of the software and better knowledge of the detector conditions and calibrations that will require reprocessing of the data to exploit the improvements. Reprocessing of detector data with a different software release is expected to trigger the re-creation of the corresponding MC samples and the skimming of these new data samples.

User analysis will run on skimmed events, unless a suitable skim is unavailable. Detector data processing, simulation and skimming will be centrally managed, while physics analysis will be the responsibility of users.

3 Outline of the data flow

Two trigger levels are employed in the online system. The level 1 (L1) trigger includes multiple lines and prescales. Events that pass at least one L1 trigger line are processed by the High Level Trigger (HLT), which provides a different set of trigger lines and prescales. Events that pass at least one trigger line are kept, while those that pass no HLT trigger lines are not recorded if the HLT is running in filtering mode, which is now the default.

Raw data are transferred from the online system to KEKCC by the core computing team and then registered to the grid for processing at raw data centers. While raw data are accessible at KEKCC, the Data Production team uses the trigger decisions made online by the HLT to produce HLT skimmed raw data samples (hRaw). The hRaw samples necessary for calibration are also registered to the grid and transferred to BNL for prompt calibration. The hRaw samples of delayed bhabhas used to produce the background overlay files needed for MC production as well as the HLT hadron skim are registered to the grid and transferred to the raw data centers.

Data calibration proceeds according to the following scheme:

- Detector-based calibrations are performed using dedicated “local runs” during data taking.
- Concurrently, tracking and alignment calibrations are performed using the HLT-skimmed raw data.
- When the required payloads are available, the raw data are partially processed to add tracking information and produce the cDST (calibration DST) samples, which include the raw data and tracking objects. This reduces the amount of time required to run calibration algorithms, but retains the raw data needed for calibration.
- All other calibrations use the cDST files as input to produce the necessary payloads.

The cDST samples produced during prompt calibration at BNL are replicated to DESY for subsequent recalibration campaigns.

Raw data prompt processing of the HLT hadron sample begins immediately after the prompt calibration is finished in order to provide samples for analysis as quickly as possible. Prompt raw data processing of the full dataset begins immediately thereafter.

Year	2024	2025	2026	2027
Number of raw data reprocessings	1	0.5	0.5	0.5
Number of MC campaigns	1	0.5	0.5	0.5

Table 1: Number of raw data reprocessing and of MC Campaigns per year. From 2025, we plan on a full data reprocessing and on a full MC data re-generation and processing every 2 years.

Analysis skimming to produce uDST samples begins as soon as the data is ready, with the highest priority skims (systematics) produced concurrently with data processing. Systematics skims consist of decay modes that are used to determine correction factors and are otherwise used for physics performance studies that produce required input for physics results.

Run-independent MC samples, using simulated backgrounds and static conditions, are produced as soon as the major software release is ready and beam background overlay files are prepared and distributed on the grid. Run-dependent MC samples, using random trigger background overlay files and real conditions, are produced on the grid soon after random trigger overlay files are prepared from the delayed Bhabha trigger HLT-skimmed raw files. Skimming of MC samples begins as soon as the samples are available on the grid.

During early data taking, when significant improvements to software and detector configurations are expected, major reprocessing campaigns will happen each year. When conditions are stable, this will be reduced to once every two years. The planned number of reprocessing campaigns per year are listed in Table 1. Starting from 2025, we plan to have a reprocessing of the full data sample every 2 years. The reprocessing of the detector data with new releases of the software will trigger the reproduction of the corresponding MC samples to have consistent data sets and the re-skimming of the detector and MC samples.

A new major software release is typically tagged in July and tested in August and September to be ready for the prompt processing of the new data collected from the fall and to recalibrate and reprocess the data previously collected. Both prompt processing and reprocessing should start in October.

Only the data for which improvements are expected from new calibration algorithms or for which problems were discovered with prior calibrations will be recalibrated. Otherwise, the existing payloads will be reused when the data is reprocessed with a new release. The reprocessing begins with the earliest data taking periods, with a target completion by May of the following year to be available for the summer conferences. Data are made available as soon as they are reprocessed, meaning that the reprocessed data set does not become available all at once upon completion.

In parallel with the reprocessing, prompt calibration and processing is performed using the same major release. In this way, the datasets available in the summer of each year are processed with the same major release. This allows for consistent use of performance corrections and systematic studies.

4 The schedule of SuperKEKB

The planning of the computing activities depends on the SuperKEKB schedule. The integrated luminosity collected in 2018 Phase 2 data taking was 0.5 fb^{-1} and the one collected in Phase 3 since 2019 until now is shown in Table 2. The Long Shutdown 1 (LS1) for the installation of the PXD began in June 2022 and will continue until the beginning of calendar year 2024.

Year	2019	2020	2021	2022	2023	2024	2025	2026	2027
Luminosity (ab-1/year)	0.01	0.09	0.2	0.19	0.06	0.35	0.5	0.7	1.2
Integrated Luminosity (ab-1)	0.01	0.1	0.3	0.49	0.55	0.9	1.4	2.1	3.3

Table 2: The SuperKEKB Luminosity profile: February 2023

Year	2019	2020	2021	2022	2023	2024	2025	2026
Luminosity (ab-1/year)	0.01	0.09	0.2	0.18	0.19	0.44	0.55	0.69
Integrated Luminosity (ab-1)	0.01	0.1	0.3	0.48	0.67	1.11	1.66	2.35

Table 3: The SuperKEKB Luminosity profile: February 2022

Year	2019	2020	2021	2022	2023	2024	2025
Luminosity (ab-1/year)	0.01	0.09	0.38	0.25	3.09	4.14	4.72
Integrated Luminosity (ab-1)	0.01	0.1	0.48	0.73	3.82	7.96	12.68

Table 4: The SuperKEKB Luminosity profile: June 2021

The assumed luminosity profile for years 2024 to 2027 is presented in Table 2. Tables 3 and 4 show, for comparison, the luminosity profile that was used for the computing resource estimate in February 2022 and June 2021, respectively. The current luminosity profile is based on the 2022 running experience with additional assumptions on the evolution of the machine parameters, running time, and LS1.

5 Size of the MC samples

The B factories experience has demonstrated that it is important to have a generic MC sample that is many times the detector data sample. MC studies of Belle II performance have shown that the generic MC sample for hadronic and τ events should be at least 4 times the detector data sample up to an integrated luminosity of at least 1 ab^{-1} to avoid the case in which the dominant systematic error is the one coming from limited MC statistics.

This requirement can be reduced to twice the detector data sample up to an integrated luminosity of 5 ab^{-1} and to one time the detector data sample for higher integrated luminosities. The amount of MC for low multiplicity events can be smaller without significantly impacting the quality of the physics results. Both run-dependent and run-independent MC samples will be produced. Run-dependent MC samples match the detector conditions

Class of events	2024	2025	2026	2027
hadronic	4	4	3	2
τ	8	8	6	2
$\mu^+\mu^-(\gamma)$	4	4	3	2
$\gamma\gamma(\gamma)$	2	2	1.5	1
Bhabha	0.2	0.2	0.15	0.05
$e^+e^-e^+e^-$, $e^+e^-\mu^+\mu^-$	2	2	1.5	0.5
Other e^+e^-X	2	2	1.5	0.5

Table 5: Ratios of the sizes of the run-dependent generic MC event samples to the detector event sample per year.

Class of events	2024	2025	2026	2027
hadronic	1	0.75	0.5	0.25
τ	1	0.75	0.5	0.25
$\mu^+\mu^-(\gamma)$	1	0.75	0.5	0.25
$\gamma\gamma(\gamma)$	1.3	1	0.6	0.3
Bhabha	0.17	0.15	0.08	0.04
$e^+e^-e^+e^-$, $e^+e^-\mu^+\mu^-$	1	0.75	0.4	0.2
Other e^+e^-X	1	0.75	0.4	0.2
Non-Υ(4S)	0.7	0.7	0.5	0.5

Table 6: Equivalent luminosity (ab-1) of the run-independent generic MC event samples.

on a run by run basis and use background events from delayed bhabha triggers, collected during the data taking, overlaid on the generated event to simulate the effect of machine background. The ratios of the run-dependent generic MC event sample over the detector event sample are summarized in Table 5. Run-independent MC samples are generated using average detector conditions and simulated background events. Table 6 shows the luminosity corresponding to the amount of run-independent MC events that are planned. In addition, samples of signal events for specific studies will be centrally produced upon request of the Physics Analysis Groups.

In comparison to previous resource estimates, the scale factor for the size of MC samples relative to the data is larger in each year, but consistent as a function of integrated luminosity. That is, since the luminosity profile was revised downward, the reduction in the relative size of the MC was delayed to avoid a situation in which the precision of physics and performance results is limited by MC statistics.

6 Inputs to resource estimates

The following details are used to determine the computing resources required for Belle II.

Class of events	2021	2022	2023	2024	2025	2026	2027
Υ(4S)	1.05	1.05	1.05	1.05	1.05	1.05	1.05
ccbar	1.3	1.3	1.3	1.3	1.3	1.3	1.3
uds	2.15	2.15	2.3	2.3	2.3	2.3	2.3
$\tau^+\tau^-$	0.86	0.86	0.84	0.84	0.84	0.84	0.84
$\mu^+\mu^-(\gamma)$	0.88	0.88	1.03	1.03	1.03	1.03	1.03
$\gamma\gamma(\gamma)$	3.74	3.74	3.17	3.17	2.01	2.01	2.01
$e^+e^-(\gamma)$	20	18	14.19	12	8.35	8.35	8.35
$e^+e^-e^+e^-$	1.14	1.14	1.66	1.66	1.66	1.66	1.66
$e^+e^-\mu^+\mu^-$	0.82	0.82	1.82	1.82	1.82	1.82	1.82
e^+e^-X	2.2	2.2	0.72	0.72	0.72	0.72	0.72
Sum (all)	34.1	32.1	28.1	25.9	21.1	21.1	21.1

Table 7: Accepted cross sections (nb) for different classes of events. The accepted cross section for 2023 and later is the proposed one, subject to potential revision.

6.1 Accepted cross section

Table 7 shows the accepted cross sections for different classes of events. We plan to accept all of the hadronic and τ events, while the other low multiplicity events will be pre-scaled. At BaBar and Belle, essentially all of the hadronic and τ events have been used to extract interesting physics results; we expect the situation to be similar at Belle II. Studies done on Belle II skimming code show that more than 90% of hadronic and τ events will be selected by at least a skim.

The low multiplicity events will be used to extract physics results and for detector and for calibration studies. An open trigger for low multiplicity events increases the sensitivity for new physics searches like searches for dark matter candidates. On the other hand, the increase in luminosity will require a more restricted trigger and a smaller accepted cross section for low multiplicity events. The optimization will be done during the data taking, considering the peak luminosity and the machine background levels.

Table 7 shows the cross section accepted by the current HLT menu (measured during the 2021 data taking) and an extrapolation to data taking in 2023 and later that assumes some tuning of the HLT menu. The proposed accepted-cross-section for 2023 was modified to reduce the effective cross section (the number of recorded events divided by the integrated luminosity), which had increased during running by 64% relative to the nominal value, mostly due to how the track triggers were implemented. By tightening some track selection criteria and increasing prescales, the sum of accepted cross sections is predicted to be 28.08 nb in 2023, compared to the 41.6 nb for the HLT configuration in the 2022a/b running period. This estimate was made using data and run-dependent MC15 samples, which were in excellent agreement. Additional reductions in low multiplicity event rates are expected for future running periods and are factored into this resource estimate. Studies of run-dependent MC samples with different background conditions suggest an uncertainty of 7% on the predictions for the accepted cross sections.

Process	MC15
$\Upsilon(4S)$	1.05
ccbar	1.3
uds	2.13
$\tau^+\tau^-$	0.92
$\mu^+\mu^-(\gamma)$	1.15
$\gamma\gamma(\gamma)$	5.8
$e^+e^-(\gamma)$	74.5
$e^+e^-e^+e^-$	5.88
$e^+e^-\mu^+\mu^-$	5.87
e^+e^-X	2.2
Sum (all)	100.8
Non-$\Upsilon(4S)$	5.8

Table 8: Generated cross sections (nb) for different classes of events. Preliminary cuts at generator level are included.

6.2 Simulated cross sections

The MC events are generated according to the cross sections modeled in the generators. To optimize the usage of computing resources, a set of preliminary cuts is applied to the generated events before passing them to the Belle II detector simulation and reconstruction. The goal of these preliminary cuts is to remove, immediately after generation, those events that will be discarded later by the reconstruction and analysis software. Table 8 shows the cross sections of the different processes after applying the preliminary cuts that are utilized in the MC 15 campaign already in progress.

The cross sections in the MC 15 column of Table 8 is used to estimate the computing resources required by the 2024 MC production.

6.3 Raw data size

The raw data coming out of the Belle II data acquisition are an uncompressed sequence of ROOT objects that will be converted to ROOT files and permanently stored on mass storage. The size of one event in ROOT format has now been measured in Phase 3 data taking, with incomplete PXD, to be around 60 kB/event. Once the full PXD is installed, after LS1, the raw data event size will increase. Increased luminosity can possibly result in an increase of the machine background despite the foreseen mitigation actions. Higher background levels result in higher occupancy in the detector and larger raw data size.

To take into account the impact of the full PXD installation and of the uncertainties in the level of machine background, we add 20 kB on top of the previous value to obtain 80 kB/event. We assign a 25% uncertainty on this value and take 80 ± 20 kB/event as the raw data size to estimate the amount of offline storage needed for raw data.

In addition to the raw data corresponding to the all event sample, a subset of events, selected by specific HLT lines, are also stored in raw data format. These data are called hRaw data and account for roughly the 30% of the size of the total raw data sample. They

Site	2024	2025	2026	2027
BNL	0.30	0.30	0.30	0.30
Italy	0.20	0.20	0.20	0.20
Germany	0.20	0.20	0.20	0.20
France	0.15	0.15	0.15	0.15
Canada	0.15	0.15	0.15	0.15
Size of Raw Data (one copy)	3.7 + 1.1	4.5 + 1.3	5.7 + 1.6	7.7 + 2.2
Size of Raw Data (including replicas)	7.3 + 2.1	9.0 + 2.6	11.4 + 3.3	15.4 + 4.5

Table 9: Sharing of the copy of raw data and size of the raw data sample (PB). The added number in the last two rows is the size of the hRaw data.

Year	2024	2025	2026	2027
Storage for prompt calibration (PB)	0.2	0.2	0.2	0.2
CPU power for prompt calibration (kHS06)	4	4	4	4
Total cDST data (PB)	2.5	3	3.5	4
CPU power for recalibration (kHS06)	12	15	18	21
MC cDST (PB)	0.4	0.5	0.6	0.7

Table 10: Parameters of the calibration model

include the raw data needed for prompt calibration, hadronic events that are processed with higher priority, and the delayed Bhabha trigger events used to create the background files for run-dependent MC simulation.

One copy of the raw data will be permanently stored at KEK. To avoid catastrophic loss of data, a second copy is stored elsewhere, now shared among different countries according to Table 9. The funding agencies of Canada, France, Germany, Italy, and US have accepted the sharing according to Table 9.

6.4 Calibration step

A subset of events in raw data format, selected according to the requirements of the calibration algorithms, is stored separately and is centrally processed to produce events in cDST format. The separately stored raw data are named “skimmed raw data” and are a subset of the hRaw data discussed previously. Events in cDST format consist of raw data and reconstructed tracking objects that are used to compute the calibration constants for the various Belle II subdetectors.

For each week of data taking, a set of events equivalent to an integrated luminosity of 9 fb^{-1} is used to produce the corresponding cDST and to calculate the prompt calibrations. This step is performed only for the data collected during the current year of data taking. Clearly the amount of computing resources required for cDST production and prompt calibration doesn’t scale with luminosity. It is estimated to be 0.2 PB of storage and 4 kHEPSpec06 of CPU power.

The cDST are stored and will be used in the recalibration step in one or more sites.

Class of events	Release 5	Release 6	Release 7
Υ(4S)	64.5	64.8	62
ccbar	53.3	55.4	51.8
uds	46.8	42.6	40.2
$\tau^+\tau^-$	36.2	28.2	27.5
$\mu^+\mu^-(\gamma)$	26.1	23.4	23.4
$\gamma\gamma(\gamma)$	26.1	23.4	23.4
$e^+e^-(\gamma)$	26.1	23.4	23.4
$e^+e^-e^+e^-$	26.1	23.4	23.4
$e^+e^-\mu^+\mu^-$	26.1	23.4	23.4
e^+e^-X	26.1	23.4	23.4
Average	29.8	27.4	27.6
Scale factor for software upgrade	1.05	1	1
Average including software upgrade	31.3	27.4	27.6
Scale factor for background uncertainty	1.1	1.1	1.1
Average with all scale factors	35 ± 2	30 ± 2	30 ± 2

Table 11: Processing power (HEPSpec06 * sec / event) needed to fully reconstruct a detector event.

The total amount of cDST scales with the time of data taking and does not depend on luminosity. It is estimated to be 0.5 PB/year. All events will be recalibrated once per year. For the time being, prompt calibration is done at BNL and recalibration is done at DESY, but this may change in the future. In addition, a sample of MC events in cDST format is produced for calibration and detector performance studies. The parameters of the calibration model are shown in Table 10.

6.5 Processing Power and buffer space for raw data processing

The CPU power required for processing one event and writing the reconstructed event in mDST format has been measured with software release 06-00-08 for different classes of simulated events and different background levels. Reference [1] describes the procedure in detail, including a breakdown of the processing power used in the different framework modules. The average processing power, weighted by the accepted cross sections, is 27.4 HEPSpec06 * s / event for the reconstruction of an event simulated with nominal background. This value is 10% lower than the corresponding value of 29.8 HEPSpec06 * s / event measured with release 05-01-08 in February 2021.

Software release 07 is not yet available in official form, though prereleases are available for tests. Table 11 includes an estimate for the processing power necessary to reconstruct a detector event with release 07 based on preliminary measurements. The main foreseen improvements of the code are the following:

- Optimizing the track fitter.
- Optimizing tracking algorithms for high background.

The cumulative effect of these software improvements is expected to have minimal impact on the processing power. The experience gained during Phase 2 and Phase 3 data taking have significantly improved the understanding of the machine background and many mitigation actions have been implemented. Increasing the peak luminosity of SuperKEKB is still expected to increase the machine background levels in the Belle II detector and to take into account this possibility we include a 10% increase of the CPU power and quote half of the corresponding increase as background uncertainty. The errors from the foreseen software improvements and from the background uncertainty are summed in quadrature.

The resulting CPU power required to fully reconstruct a detector event is 30 ± 2 HEPSpec06 * s / event (see Table 11).

The revised plan of selecting a set of events equivalent to an integrated luminosity of 9 fb^{-1} for each week of data taking and performing the calibration and reconstruction of the full set of events within eight weeks requires a buffer space equivalent to twice the size of the raw data collected in eight weeks of data taking (see Table 16), since two replicas are stored for raw data during calibration and processing. Previously, the proposed processing scheme required three weeks from data taking to processing, but experience has shown this to be insufficient. The previous resource estimate did not take into effect that two replicas are stored, one at KEK and another at the other raw data centers.

The previous resource estimate did not include buffer space for raw data staging from tape during reprocessing (only during prompt processing). Based on the average size of data samples used during recent reprocessing campaigns, equivalent to about 50 fb^{-1}, a buffer space of 500 TB is necessary for this purpose. This amount has been added to the “Disk for raw data processing” in Table 16 and in the appendix. The previous estimate also did not include buffer space for intermediate files produced during mDST production. This is estimated to require approximately 30% of the final mDST disk space, since intermediate files are deleted only after productions are complete. The previous estimate also only accounted for buffer space for a single replica of mDST data, but Rucio immediately duplicates the data while the production is ongoing. That means that additional disk space for a full copy of the mDST data is required. The “Disk for data mDST buffer” has therefore been increased by a factor of 2.3.

In principle, it is instead possible to store the raw data on tape and stage them from tape to disk: however, some disk storage will be necessary anyway and additional tape drives will be needed. The savings will be marginal and the operations will be more complicated and reduce the speed at which prompt data become available.

6.6 Processing power and buffer storage for Monte Carlo production

The processing power to perform the full simulation and the reconstruction of one event and write the event in mDST format has been measured on software release 06-00-08 for different classes of events and different background level [1]. An estimate for the processing power necessary to reconstruct a detector event with release 07 based on preliminary measurements.

Class of events	Release 5	Release 6	Release 7
$\Upsilon(4S)$	85	93.6	89.5
ccbar	71.1	83.1	77.7
uds	62.5	65.7	62.1
$\tau^+\tau^-$	45	40.5	39.3
$\mu^+\mu^-(\gamma)$	35.1	29.9	29.9
$\gamma\gamma(\gamma)$	35.1	29.9	29.9
$e^+e^-(\gamma)$	35.1	29.9	29.9
$e^+e^-e^+e^-$	35.1	29.9	29.9
$e^+e^-\mu^+\mu^-$	35.1	29.9	29.9
e^+e^-X	35.1	29.9	29.9
Average	48.6	47.2	42.5
Scale factor for software upgrade	1.05	1	1
Average including software upgrade	52.9	48.8	42.5
Scale factor for background uncertainty	1.1	1.1	1.1
Average with all scale factors	58 ± 4	53 ± 4	47 ± 3

Table 12: Processing power (in HEPSpec06 * sec /event) for Monte Carlo production.

The average processing power, weighted by the accepted cross sections, is 47.2 HEPSpec06 * s / event). In release 05-01-08, the average processing power was 48.6 HEPSpec06 * s / event. This improvement comes mostly from the optimization of the simulation of low multiplicity events.

To take into account the possibility of having a higher background when increasing the peak luminosity, we include a 10% increase of the CPU power and quote half of the corresponding increase as background uncertainty.

The CPU power required to fully simulate and reconstruct one event with software release 7 is expected to be 47 ± 3 HEPSpec06 * s / event (see Table 12), based on preliminary estimates, where the errors from the foreseen software improvements and from the background uncertainty are summed in quadrature.

The event generation and event reconstruction steps of MC production are performed in a single job, without writing a raw-data-like intermediate output. In principle, writing the intermediate raw data after the generation step will allow to perform only the reconstruction step when re-producing the MC data in case of reprocessing. However the saving will be marginal at best, because the reconstruction step requires on the order of 70% of the total CPU power and large storage for intermediate data will be required. In addition, in that case, it will be possible to run simulation only at sites with large storage resources.

It is worth noting that the statement that the reconstruction step requires on the order of 70% of the total CPU power, is only apparently contradicted by using the values of 47 HEPSpec06 * s / event for MC production and 30 HEPSpec06 * s / event for detector data processing. These values are averages over different classes of events and the detector data sample has a much larger fraction of low multiplicity events, which are less CPU intensive then hadronic and τ events, than the MC sample.

The MC production produces a set of mDST files that are too small for convenient

Class of events	Release 5	Release 6	Release 7
Υ(4S)	17.8	14.9	15
ccbar	16.3	13.4	13.5
uds	15.3	12.4	12.5
$\tau^+\tau^-$	13.7	10.2	10.4
$\mu^+\mu^-(\gamma)$	12.9	9.3	9.5
$\gamma\gamma(\gamma)$	12.9	9.3	9.5
$e^+e^-(\gamma)$	12.9	9.3	9.5
$e^+e^-e^+e^-$	12.9	9.3	9.5
$e^+e^-\mu^+\mu^-$	12.9	9.3	9.5
e^+e^-X	12.9	9.3	9.5
Average	13.4	9.9	10.1
Scale factor for software upgrade	1.05	1.05	1.0
Average including software upgrade	14	10.4	10.1
Scale factor for background uncertainty	1.02	1.02	1.02
Average with all scale factors	14.3 ± 1.3	10.6 ± 1.1	10.3 ± 1.1

Table 13: Sizes of detector events in mDST format (kB/event).

handling. These intermediate files are merged, in a two-step process, to create the final mDST files. Thus, the MC production generates an amount of intermediate files that is twice the size of the final mDST data-set. These intermediate files are deleted in bulk during the MC production, but some buffer space, estimated to be of the order of 30% of the final MC mDST size is needed (see Table 16).

In addition, files with random trigger events collected during the data taking to perform background overlay need to be stored. We estimate the need to be 60 GB/fb^{-1} (see Table 16).

6.7 Event size in mDST format

The mDST event sizes have been measured on software release 06-00-08 for different classes of events and different background level. Reference [1] describes the procedure in detail, including a breakdown of the sizes of different branches. An estimate for the processing power necessary to reconstruct a detector event with release 07 based on preliminary measurements. The average sizes are 10.0 kB/event and 11.9 kB/event for detector and MC events, respectively. These values are smaller than the ones measured in release 05-01-08 due to some optimization of the mDST content.

To take into account the possibility of having a higher background when increasing the peak luminosity, we include a 2% increase of the event size and quote half of the corresponding increase as background uncertainty. The errors from the foreseen software improvements and from the background uncertainty are summed in quadrature. Finally, we get 10.2±1.1 kB/event and 12.1±1.2 kB/event for detector and MC events, respectively. Tables 13 and 14 show the sizes of detector and MC events in mDST format.

Class of events	Release 5	Release 6	Release 7
Υ(4S)	20.9	18.4	18.5
ccbar	18.7	16.1	16.2
uds	17	14.4	14.5
$\tau^+\tau^-$	14.7	11.3	11.4
$\mu^+\mu^-(\gamma)$	13.2	9.6	9.8
$\gamma\gamma(\gamma)$	13.2	9.6	9.8
$e^+e^-(\gamma)$	13.2	9.6	9.8
$e^+e^-e^+e^-$	13.2	9.6	9.8
$e^+e^-\mu^+\mu^-$	13.2	9.6	9.8
e^+e^-X	13.2	9.6	9.8
Average	15.5	12.1	11.9
Scale factor for software upgrade	1.02	1.05	1.0
Average including software upgrade	15.8	12.7	11.9
Scale factor for background uncertainty	1.02	1.02	1.02
Average with all scale factors	16.1 ± 1.4	13 ± 1.2	12.1 ± 1.2

Table 14: Sizes of MC events in mDST format (kB/event).

	Release 5	Release 6	Release 7
Average processing power for skimming data (HS06*s/evt)	2.9	4.4	4.6
uDST detector event size (kB)	16.9	12.8	12.6
(Events in uDST)/(events in mDST) data	0.7	0.7	0.7
Average processing power for skimming MC (HS06*s/evt)	4.6	5.9	5.6
uDST MC event size (kB)	21.3	17	15.1
(Events in uDST)/(events in mDST) MC	1.2	1.2	1.2

Table 15: Parameters of the skimming model for Detector and MC events.

6.8 Skimming

The skimming software has evolved in the past year: new skims have been added and some of the skims that selected largely overlapping samples have become a single skim. Different output skims are created from a single job. Currently we have 11 combined skims that run between 2 and 6 skims at once with separate output. Overall there is an optimization process ongoing that will continue in the future as experience is gained. A key aspect of skimming to uDST files is the analysis data format. The uDST is defined as a skimmed mDST with additional data objects that can be exploited in subsequent analysis. The central component of this is the Particle object, which links particle hypotheses with tracks, particle identification and neutral cluster information. Vertex fit results (covariance matrices) for combined particles are also saved in the Particle objects. Furthermore, results from B and D meson full reconstruction, continuum suppression and other complex algorithms can be saved into dedicated analysis objects. This allows for pre-processing that reduces the CPU requirements in the final analysis step.

The amount of CPU power required for skimming, the skim retention factor (defined as (Events in uDST) / (Events in mDST)), and the size of one event in uDST format

	2024	2025	2026	2027
Disk for raw data processing	1.08	1.17	1.44	2.12
Disk for data mDST in production	0.72	0.97	1.33	1.93
Disk for MC production	0.41	0.53	0.64	0.74
Disk for MC mDST in production	1.35	1.76	2.12	2.47
Disk for random trigger files	0.16	0.25	0.37	0.58
Disk for skimming	0.73	0.97	1.21	1.53
Disk for uDST in production	2.43	3.24	4.03	5.1

Table 16: Buffer spaces (in PB) for raw data processing, MC production and skimming. Buffer spaces (in PB) for the mDST and uDST under production. Space for random trigger files for MC production.

have been measured while skimming the MC events produced in MC15.

The parameters of the skimming model are summarized in Table 15. The detector mDST data sample will include a larger fraction of low multiplicity events than the MC mDST data sample. The effect of these different sample compositions is included in the estimate of the processing power and uDST event size in Table 15. The uDST event size in detector data is calculated subtracting the size of MC truth information stored in MC uDST from the size of a MC uDST event.

The skimming produces a set of uDST files that are too small for convenient handling. These intermediate files are merged to create the final uDST files. Thus, the skimming generates an amount of intermediate files that is equivalent in size to the final uDST data-set. These intermediate files are deleted in bulk during the skimming, but some buffer space, estimated to be on the order of 30% of the final uDST data-set size is needed (see Table 16).

6.9 mDST and uDST replica

We know that some level of replication of mDST and uDST data will be required for safety and to have reasonable performance of the analysis jobs. For the time being we assume that two replicas will be sufficient.

Until the production with release X is completed, we plan to keep on disk the replicas of release X-1 and X-2. When production with release X is completed, the replicas of release X-2 will be removed from disk. This implies that a buffer space equivalent to one replica of the mDST and uDST in release X is needed (see Table 16).

A key concept in this context is the “popularity” of the data that helps in optimizing the usage of storage. Data sets that are rarely accessed do not need to be replicated and data that are not accessed at all can be removed from disk storage. As soon as the appropriate tools will become available, we plan to move to a replica policy based on data popularity.

	2024	2025	2026	2027
CPU power to run one analysis (HS06*s/evt)	0.08	0.08	0.08	0.08
Number of concurrent analyses	250	250	250	250
Number of analysis cycles	1.75	1.5	1.25	1.25
CPU power to run all analyses (HS06*s/evt)	35	30	25	25
ntuple size for all analyses (kB/evt)	8.8	7.5	6.2	6.2

Table 17: Parameters of the analysis model.

6.10 Analysis

User analysis jobs are run on skimmed uDST data to produce n-tuples that will be used to perform the final fit and systematic studies. The uncertainties associated with analysis jobs are much larger and therefore more difficult to estimate in advance. Some steps such as vertexing, particle combinations, and flavor tagging, do not need to be run again due to pre-processing at the skimming stage.

The amount of CPU power used for analysis in 2022 was significantly larger than originally estimated from the 2021 resource estimate. The expected number of concurrent analyses was only 100, when in reality more than 200 concurrent analyses ran. The parameters of the model for the analysis step have been adjusted to better reproduce the observed amount of CPU power devoted to analysis.

The values used in the resource estimate are:

- The average CPU power to run one analysis on one event is 0.08 HEPSpec06 * s / event.
- The number of concurrent analysis is currently about 200, which is twice the expected 100 concurrent analyses. In the revised estimate we use 250 for 2024 and later, based on an expected increase in analysis use.
- The number of analysis cycles (defined as the number of times an analysis is run on the data and MC samples) is 2 in 2023, 1.75 in 2024, 1.50 in 2025, and 1.25 in 2026 and 2027.
- The uncertainty is at the 50% level.

The total processing power to analyze one event (all analyses, all analysis cycles) is shown in Table 17 together with the parameters of the analysis model. The user n-tuple size scales with the number of concurrent analyses and of analysis cycles. We assume that running one cycle of one analysis on the skimmed data sample will produce on average 0.02 ± 0.01 kB/event of n-tuple data.

6.11 Cosmic Data

We conservatively assume that the processing and reprocessing of the cosmic data will require a CPU power of 0.5 kHEPSpec06 for 10 months assuming a trigger rate of 100 Hz for 3 months of cosmic data taking. In addition, we assume that the production of the corresponding sample of MC cosmic events will require 1 kHEPSpec06 for 10 months and that another 1.5 kHEPSpec06 will be needed for cosmic data analysis.

	2024	2025	2026	2027
Total tape (PB)	9.6	11.8	14.8	20.1
Total disk (PB)	25	31.9	39.6	49.3
Total CPU (kHS06)	520	465	464	519

Table 18: The resource estimate for years 2024 to 2027. CPU needs drop in 2025 due to the reduced rate of data reprocessing and MC production. The storage units are PB = 10^{15} Bytes.

	2019	2020	2021	2022	2023
Total tape (PB)	10	1.4	3.2	8.8	8.6
Total disk (PB)	11	11	11	16.5	19.6
Total CPU (kHS06)	299	207	498	385	404

Table 19: The resources requested for 2019 through 2023.

6.12 Efficiency

We assume to be able to run the code for 10 months every year because time will be lost due to site downtimes, operational mistakes, software bugs discovered only after the start of the data processing that will require to restart it from scratch, etc. This corresponds to an efficiency in the use of CPU power of 83% that compares well with the 80% efficiency experienced by the different LHC experiments at the beginning of data taking.

7 The Resource Estimate

The input parameters that have been discussed in the previous section are used to obtain the resource estimate summarized in Table 18. CPU and storage for processing and reprocessing of raw data will be shared among the countries hosting raw data, according to the fraction of raw data hosted. CPU and disk storage for MC production and analysis, and storage for mDST and uDST data will be shared among the different countries according to the PhD count. For comparison, Table 19 shows the computing resource levels pledged for 2019 (approved at the October 2018 Financial Oversight Panel), for 2020 (approved at the October 2019 Financial Oversight Panel), for 2021 (approved at the October 2020 Financial Oversight Panel), for 2022 (approved at the October 2021 Financial Oversight Panel), and for 2023 (approved at the October 2022 Financial Oversight Panel). For this resource estimate, the storage units are PB = 10^{15} Bytes.

A detailed breakdown of the different activities and data samples for the years 2024 to 2027 is included in the Appendix A.

8 Conclusion

We have presented the estimate of the computing resources needed by the Belle II collaboration in the years 2024 to 2027 and described the planned activities and the assumptions that lead to this estimate. This estimate has been produced by the Computing Steering

Group and approved by the Executive Board.

References

1. BELLE2-NOTE-TE-2016-011; T. Kuhr; Software Resource Estimates.

Appendix A: Resource Breakdown for years 2024 to 2027

Tape (PB)	2024	2025	2026	2027
Tape for raw data	7.34	9.03	11.39	15.43
Tape for hRaw data	2.13	2.62	3.3	4.48
Tape for cosmic data	0.15	0.15	0.15	0.15
Total tape	9.61	11.79	14.83	20.06

Table 20: Tape resources by data type (in PB).

Disk (PB)	2024	2025	2026	2027
Disk for Data Processing Buffer (PB)	1.08	1.17	1.44	2.12
Disk for Data mDST buffer (PB)	0.72	0.97	1.33	1.93
Disk for calibration (PB)	2.7	3.2	3.7	4.2
Disk for data mDST (PB)	1.07	0.85	1.15	1.68
Disk for RandomTrigger files for MC Prod	0.16	0.25	0.37	0.58
Disk for MC prod buffer (PB)	0.41	0.53	0.64	0.74
Disk for MC mDST buffer (PB)	1.35	1.76	2.12	2.47
Disk for MC cDST	0.4	0.5	0.6	0.7
Disk for MC mDST [PB]	4.72	6.23	7.76	9.17
Disk for Data + MC uDST (PB)	8.36	11.35	14.53	18.25
Data + MC skimming buffer	0.73	0.97	1.21	1.53
Data + MC uDST buffer	2.43	3.24	4.03	5.1
Disk for user ntuples (PB)	0.89	0.88	0.74	0.79
Disk for Cosmic (PB)	0.01	0.01	0.01	0.01
Total Disk (PB)	25.04	31.91	39.63	49.27

Table 21: Disk resources by data type (in PB).

CPU (kHS06)	2024	2025	2026	2027
CPU for data processing (kHEPSpec)	10.6	13.1	18.3	31.4
CPU for data reprocessing (kHEPSpec)	25.3	13.2	13.2	25.3
CPU for Calibration (kHEPSpec)	16	19	22	25
CPU for MC production [kHEPSpec]	289.2	247.1	255.2	267.1
CPU for Skimming (kHEPSpec)	43.3	38.7	42.5	49.6
CPU for analysis (kHEPSpec)	136	133.6	113.1	120.5
Total CPU (kHEPSpec)	520	465	464	519

Table 22: CPU resources by usage (in kHS06).