BELLE2-NOTE-TE-2024-001
Version: 1.5

Belle II Computing Resource Estimate for 2025-2028

The Belle II Computing Steering Group

February 1, 2024

1 Introduction

In this note we present an estimate of the computing resources needed by the Belle II collaboration in Japanese fiscal years (JFY) 2025 to 2028 and describe the planned activities and the assumptions that lead to this estimate. This estimate supersedes the one presented at the June 2023 BPAC review. Compared to the estimate presented in June 2023, this revised estimate uses the same luminosity profile for JFY 2023 through 2027 and an estimate for JFY 2028, but is revised to account for the data taking efficiency, which averages about 88%, determined as the recorded luminosity divided by the delivered luminosity. The processing scheme is also revised based on the anticipation that reprocessing will happen in 2024 and then in the two year period of 2026-2027. Another major revision is to update how the storage and CPU resources are estimated, based on the prompt and reprocessing scheme described below. In the following, dates refer to JFY unless stated otherwise.

An outline of the Belle II computing model is presented, followed by a description of the planned computing activities. The input parameters used for the resource estimate are discussed and the resource estimate for years 2025 to 2028 are presented. 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 mDSTs are stored in regional data centers.

Detector and MC mDST samples are then “skimmed” at regional data centers to create a sample of events that suit a specific set of physics analyses. During the skimming step, additional information is computed and may be added to the events. The output of the skimming step is the full mDST information for a selected subset of events plus additional information useful for analysis. These files are stored in uDST (user DST) format at regional data centers. Starting in 2024, the uDST information will no longer contain additional analysis-related objects for most samples, except for the FEI skim, which still requires the extra information.

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 of the skimming of the new data and MC 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 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. The core computing team also uses the trigger decisions made online by the HLT to produce HLT-skimmed raw data samples (hRaw) that are registered to the grid and replicated to raw data centers for processing. The hRaw samples that are useful for calibration are replicated to BNL for prompt calibration. Only calibration skims, delayed bhabhas used to produce the background overlay files needed for MC production, and the HLT hadron skim are produced as hRaw.

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 hRaw 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 KEK 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. Most hadronic analyses can use the HLT hadron sample, Prompt raw data processing of the full dataset begins immediately thereafter.

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 hRaw 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 2026, we plan to reprocess the full data sample every two years. The reprocessing of detector data with new software releases will trigger the reproduction of corresponding MC samples, to have consistent data sets, and the re-skimming of detector data 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. The reprocessing efforts should start late in the calendar year, while prompt processing efforts begin as new data are collected. Since no improvements are planned for the software in 2024, there is no longer an expectation to reprocess the data starting in 2025. Instead, the next reprocessing campaign will begin in 2026.

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. 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 computing activities depends on the SuperKEKB schedule. The integrated luminosity collected in 2018 Phase 2 data taking was 0.5 fb^{-1} and that delivered

Year	2024	2025	2026	2027	2028
Number of raw data (re)processings	1	0	0.5	0.5	0.5
Number of MC reproduction campaigns	1	0	0.5	0.5	0.5

Table 1: Number of raw data reprocessing and of MC Campaigns per year. From 2026, we plan on a full data reprocessing and on a full MC data reproduction every 2 years.

by SuperKEKB in Phase 3 since 2019 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.

The SuperKEKB luminosity profile for years 2025 to 2028 is presented in Table 2. Note that this shows the delivered luminosity, while the recorded luminosity used to for the resource estimate assumes a data taking efficiency of 88%. 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.

Year	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
Luminosity (ab^{-1}/year)	0.01	0.09	0.2	0.19	0.06	0.35	0.5	0.7	1.2	1.2
Integrated Luminosity (ab^{-1})	0.01	0.1	0.3	0.49	0.55	0.9	1.4	2.1	3.3	4.5

Table 2: The SuperKEKB (delivered) luminosity profile: February 2024

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 (delivered) luminosity profile: February 2022

5 Size of the MC samples

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

The expectation is that this requirement may be reduced to twice the detector data sample up to an integrated luminosity of 5 ab^{-1} and to an equivalent size of the detector data sample for higher integrated luminosities. It is possible that significantly larger generator-level skimmed samples may be necessary to produce analysis-specific generic MC samples that are manageable in size. 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.

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 (delivered) luminosity profile: June 2021

Class of events	2024	2025	2026 - 2027	2028 - 2029
hadronic	4	4	4	2
τ	8	8	4	2
$\mu^+\mu^-(\gamma)$	4	4	4	2
$\gamma\gamma(\gamma)$	2	2	2	1
Bhabha	0.2	0.2	0.1	0.05
$e^+e^-e^+e^-$, $e^+e^-\mu^+\mu^-$	2	2	1	0.5
Other e^+e^-X	2	2	1	0.5

Table 5: Ratios of the sizes of the run-dependent generic MC event samples to the detector event sample per year. Data reprocessing and MC reproduction will be performed in two-year periods starting in 2026. No reprocessing is extected in 2025, though prompt data will be processed and the corresponding run-dependent MC produced as usual.

Run-dependent MC samples match the detector conditions on the basis of each run and use background events from delayed bhabha triggers, collected during the data taking, overlaid on the generated events to simulate the effect of machine background. The size of the run-dependent generic MC samples for each event type relative to the size of the detector data sample is 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 size of the run-independent MC samples that are planned. In addition, samples of signal events for specific studies will be centrally produced upon request of the physics analysis groups.

In the previous resource estimate, the number of streams of run-dependent MC (each stream has an equivalent luminosity equal to that of the data sample) decreased each year starting in 2025. However, since the data will be reprocessed over a two-year period starting in 2026, the MC reproduction will also happen only once in a two-year period. Since it is illogical to reduce the number of streams in 2027, while production is ongoing, the number of streams in 2027 is set equal to that in 2026. For the same reasons, the run-independent MC sample does not need to be reproduced when data and run-dependent MC production is ongoing and no new software release is expected.

6 Inputs to resource estimates

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

Class of events	2024	2025	2026	2027	2028
hadronic	1	0	0.5	0	0.25
τ	1	0	0.5	0	0.25
$\mu^+\mu^-(\gamma)$	1	0	0.5	0	0.25
$\gamma\gamma(\gamma)$	1.3	0	0.6	0	0.3
Bhabha	0.17	0	0.08	0	0.04
$e^+e^-e^+e^-$, $e^+e^-\mu^+\mu^-$	1	0	0.4	0	0.2
Other e^+e^-X	1	0	0.4	0	0.2
Non-Υ(4S)	0.7	0	0.5	0	0.5

Table 6: Equivalent luminosity (ab^{-1}) of the run-independent generic MC event samples, which are only produced when a new software release is created.

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 one skim.

The low multiplicity events will be used to extract physics results and for detector and 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 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 was 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. In particular, efforts are ongoing to develop better Bhabha vetoes for the HLT. This should result in a reduction for the accepted Bhabha cross section starting in 2025. Studies of run-dependent MC samples with different background conditions suggest an uncertainty of 7% on the predictions for the accepted cross sections. The level 1 trigger pass-through, which is used for random trigger overlays and trigger diagnostics was not accounted in the previous accepted cross section, but is accounted starting in 2025.

Class of events	2021	2022	2023	2024	2025	2026	2027	2028
$\Upsilon(4S)$	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
$c\bar{c}$	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
uds	2.15	2.15	2.3	2.31	2.31	2.31	2.31	2.31
$\tau^+\tau^-$	0.86	0.86	0.84	0.85	0.85	0.85	0.85	0.85
$\mu^+\mu^-(\gamma)$	0.88	0.88	1.03	1.03	1.03	1.03	1.03	1.03
$\gamma\gamma(\gamma)$	3.74	3.74	3.17	3.18	3.18	3.18	3.18	3.18
Bhabha	20	18	14.19	15.11	12	9	9	9
$e^+e^-e^+e^-$	1.14	1.14	1.66	1.97	1.97	1.97	1.97	1.97
$e^+e^-\mu^+\mu^-$	0.82	0.82	1.82	2.19	2.19	2.19	2.19	2.19
Other e^+e^-X	2.2	2.2	0.72	0.84	0.84	0.84	0.84	0.84
L1 passthrough	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
Sum (all)	34.14	32.14	28.08	29.83	27.82	24.82	24.82	24.82

Table 7: Accepted cross sections (nb) for different classes of events. The accepted cross section for 2025 and later is the proposed one, subject to potential revisions.

6.2 Simulated cross sections

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 were utilized in the MC15 campaign, which was completed in JFY 2023. The cross sections in Table 8 will be used to estimate the computing resources required for MC production during and after JFY 2025.

6.3 Raw data size

The raw data are transferred from the online to the offline system 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 to the raw data size, resulting in an estimated size of 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

Process	cross section (nb)
Υ(4S)	1.05
$c\bar{c}$	1.3
uds	2.13
$\tau^+\tau^-$	0.92
$\mu^+\mu^-(\gamma)$	1.15
$\gamma\gamma(\gamma)$	5.8
Bhabha	74.5
$e^+e^-e^+e^-$	5.88
$e^+e^-\mu^+\mu^-$	5.87
Other e^+e^-X	2.2
Sum (all)	100.8
Non-Υ(4S)	5.8

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

Site	2024	2025	2026	2027	2028
BNL	0.3	0.3	0.3	0.3	0.3
Italy	0.2	0.2	0.2	0.2	0.2
Germany	0.2	0.2	0.2	0.2	0.2
France	0.15	0.15	0.15	0.15	0.15
Canada	0.15	0.15	0.15	0.15	0.15
hRaw Data (one copy)	1	1.3	1.7	2.3	2.9
hRaw Data (including replicas)	2.1	2.6	3.3	4.6	5.8
Raw Data (one copy)	3.5	4.5	5.8	7.9	10
Raw Data (including replicas)	7.1	9.1	11.5	15.7	19.9

Table 9: Sharing of the copy of raw data and size of the raw data sample (PB = 1×10^{15} bytes).

hRaw data and account for roughly the 30% of the size of the total raw data sample. They 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 the US have accepted the sharing according to the top-half of 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

Year	2024	2025	2026	2027	2028
Storage for prompt calibration (PB)	0.2	0.2	0.2	0.2	0.2
CPU power for prompt calibration (kHS06)	4	4	4	4	4
Total cDST data (PB)	0.5	0.75	1	1.25	1.5
CPU power for recalibration (kHS06)	10	10	10	10	10
MC cDST (PB)	0.4	0.5	0.6	0.7	0.7

Table 10: Parameters of the calibration model

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. The skimmed raw data samples are deleted after cDST production.

For each data taking period corresponding to consistent conditions, a set of events equivalent to an integrated luminosity of 9 fb^{-1} is used to produce the corresponding cDSTs and to calculate the prompt calibration payloads. These samples are referred to as "buckets" and currently correspond to two weeks of data taking, though this may increase to three-week periods as instantaneous luminosity increases. The amount of computing resources required for cDST production and prompt calibration does not scale with luminosity. They are 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. The total size of the cDST sample scales with the number of weeks of data taking, but does not depend on luminosity. It is estimated to be 0.5 PB/year. The prompt calibration is done at BNL and, through JFY 2023, recalibration has been performed at DESY. Starting in calendar year 2024, the recalibration will instead be handled at KEK. 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.

An internal review of the Belle II calibration scheme was recently completed[1]. Several options for recalibration were considered, as well as the estimated resource needs, projected into future data taking periods. The conclusion of the review was that recalibration at a single site is preferable, at least through 2028.

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. The software group prepared a note[2] that 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.

---

[1] BELLE2-NOTE-TE-2023-013: `https://docs.belle2.org/record/3579`

[2] BELLE2-NOTE-TE-2016-011: `https://docs.belle2.org/record/371`

Class of events	Release 5	Release 6	Release 7	Release 8
Υ(4S)	64.5	64.8	62	64.8
$c\bar{c}$	53.3	55.4	51.8	57.4
uds	46.8	42.6	40.2	45.9
$\tau^+\tau^-$	36.2	28.2	27.5	38.4
$\mu^+\mu^-(\gamma)$	26.1	23.4	23.4	30.4
$\gamma\gamma(\gamma)$	26.1	23.4	23.4	30.4
$e^+e^-(\gamma)$	26.1	23.4	23.4	30.4
$e^+e^-e^+e^-$	26.1	23.4	23.4	30.4
$e^+e^-\mu^+\mu^-$	26.1	23.4	23.4	30.4
e^+e^-X	26.1	23.4	23.4	30.4
Average	29.8	27.4	27.6	33.3
Scale factor for software upgrade	1.05	1	1	1
Average including software upgrade	31.3	27.4	27.6	33.3
Scale factor for background uncertainty	1.1	1.1	1.1	1.1
Average with all scale factors	35 ± 2	30 ± 2	30 ± 2	37 ± 2

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

The processing power necessary to reconstruct a detector event with releases 7 and 8, based on preliminary measurements, are given in Table 11. The resource estimate is based on the expected processing power for release 8.

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 estimated CPU power required to fully reconstruct a detector event is 37 ± 2 HEPSpec06 * s / event (see Table 11).

The calibration scheme requires a set of events equivalent to an integrated luminosity of 9 fb^{-1} for each calibration period. The calibration and reconstruction is performed with this data set within eight weeks, which 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. Previously, the processing scheme required three weeks from data taking to processing, but experience has shown this to be insufficient. Buffer space is also required for raw data staging from tape during reprocessing. 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. The total buffer space for raw data processing is accounted in Table 16 and in the appendix.

Additional buffer space is needed for intermediate files produced during mDST production, which cannot be removed until the final merge step is completed and verified. The buffer space for intermediate files is estimated to require approximately 30% of the

Class of events	Release 5	Release 6	Release 7	Release 8
$\Upsilon(4S)$	85	93.6	89.5	93
$c\bar{c}$	71.1	83.1	77.7	86.7
uds	62.5	65.7	62.1	74.3
$\tau^+\tau^-$	45	40.5	39.3	43.2
$\mu^+\mu^-(\gamma)$	35.1	29.9	29.9	30.9
$\gamma\gamma(\gamma)$	35.1	29.9	29.9	30.9
$e^+e^-(\gamma)$	35.1	29.9	29.9	30.9
$e^+e^-e^+e^-$	35.1	29.9	29.9	30.9
$e^+e^-\mu^+\mu^-$	35.1	29.9	29.9	30.9
e^+e^-X	35.1	29.9	29.9	30.9
Average	48.6	47.2	42.5	49.7
Scale factor for software upgrade	1.05	1	1	1
Average including software upgrade	52.9	48.8	42.5	49.7
Scale factor for background uncertainty	1.1	1.1	1.1	1.1
Average with all scale factors	58 ± 4	53 ± 4	47 ± 3	55 ± 3

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

final mDST disk space.

In the previous resource estimate, buffer space was allocated for a full copy of the mDST data, since Rucio immediately duplicates the data while the production is ongoing. Rather than allocate this as buffer space, the revised estimate properly accounts for both copies of mDSTs for the ongoing production in the disk required for data mDST samples.

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 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 background levels. The processing power necessary for simulation and reconstruction with releases 7 and 8, based on preliminary measurements, are given in Table 12. The resource estimate is based on the expected processing power for release 8.

An estimate for the processing power necessary to reconstruct a detector event with release 07 based on preliminary measurements.

The average processing power, weighted by the accepted cross sections, is 55 ± 3 HEPSpec06 * s / event). In release 05-01-08, the average processing power was 48.6 HEPSpec06 * s / event.

To take into account the possibility of higher backgrounds 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 is 48 ± 3 HEPSpec06 * s / event (see Table 12) 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 reproducing the MC samples. 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 48 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 results in a set of mDST files that are too small for convenient handling. These intermediate files are merged, in a multi-step process, to create the final mDST files. Thus, the MC production generates an amount of intermediate files that is typically 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. The software note[3] describes the procedure in detail, including a breakdown of the sizes of different branches. An estimate for the event sizes produced with releases 7 and 8, based on preliminary measurements, are given in Tables 13 and 14 for data and MC events, respectively.

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 10.3 ± 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.

---

[3] BELLE2-NOTE-TE-2023-013: `https://docs.belle2.org/record/3579`

Class of events	Release 5	Release 6	Release 7	Release 8
Y(4S)	17.8	14.9	15	15.2
ccbar	16.3	13.4	13.5	13.6
uds	15.3	12.4	12.5	12.7
t+t-	13.7	10.2	10.4	10.8
m+m-(g)	12.9	9.3	9.5	10.3
gg(g)	12.9	9.3	9.5	10.3
e+e-(g)	12.9	9.3	9.5	10.3
e+e-e+e-	12.9	9.3	9.5	10.3
e+e-m+m-	12.9	9.3	9.5	10.3
e+e-X	12.9	9.3	9.5	10.3
Average	13.4	9.9	10.1	10.1
Scale factor for software upgrade	1.05	1.05	1	1
Average including software upgrade	14	10.4	10.1	10.1
Scale factor for background uncertainty	1.02	1.02	1.02	1.02
Average with all scale factors	14.3 ± 1.3	10.6 ± 1.1	10.3 ± 1.1	10.3 ± 1.1

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

Class of events	Release 5	Release 6	Release 7	Release 8
Y(4S)	20.9	18.4	18.5	18.7
ccbar	18.7	16.1	16.2	16.3
uds	17	14.4	14.5	14.7
t+t-	14.7	11.3	11.4	12
m+m-(g)	13.2	9.6	9.8	10.6
gg(g)	13.2	9.6	9.8	10.6
e+e-(g)	13.2	9.6	9.8	10.6
e+e-e+e-	13.2	9.6	9.8	10.6
e+e-m+m-	13.2	9.6	9.8	10.6
e+e-X	13.2	9.6	9.8	10.6
Average	15.5	12.1	11.9	11.9
Scale factor for software upgrade	1.02	1.05	1	1
Average including software upgrade	15.8	12.7	11.9	11.9
Scale factor for background uncertainty	1.02	1.02	1.02	1.02
Average with all scale factors	16.1 ± 1.4	13 ± 1.2	12.1 ± 1.2	12.1 ± 1.2

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

	Release 5	Release 6	Release 7	Release 8
Average processing power for skimming data (HS06*s/evt)	2.9	4.4	4.6	4.5
uDST detector event size (kB)	16.9	12.8	12.6	12
(Events in uDST)/(events in mDST) data	1.2	1.2	1.2	1.2
Average processing power for skimming MC (HS06*s/evt)	4.6	5.9	5.6	5.4
uDST MC event size (kB)	21.3	17	15.1	14.5
(Events in uDST)/(events in mDST) MC	1.2	1.2	1.2	1.2

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

	2024	2025	2026	2027	2028
Disk for raw data processing	1.09	1.29	1.48	2.18	2.18
Disk for data mDST in production	0.09	0.04	0.11	0.2	0.26
Disk for MC in production	0.39	0.09	0.33	0.47	0.34
Disk for random trigger files	0.06	0.08	0.11	0.19	0.19
Disk for skimming in production	0.67	0.2	0.69	1.06	0.94

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

6.8 Skimming

The skimming software has continually evolved as new skims have been added and some of the skims that selected largely overlapping samples have become single skims. Different skim output files can be created from a single job. Overall there is an optimization process ongoing that will continue in the future as experience is gained. The uDST file format is defined as a skimmed mDST with the potential to include 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. Based on the experience gained so far, the expectation is that the uDST output for most analyses will not include additional data objects, keeping the output size relatively low, while the Full Event Interpretation skim will include additional data objects.

The amount of CPU power required for skimming, the skim retention factor (defined as the number of events in uDST divided by the number of events in mDST), and the size of one event in uDST format 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 then 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 by subtracting the size of MC truth information stored in MC uDST from the size of a MC uDST event.

	2024	2025	2026	2027	2028
Size of prompt data mDST (latest version)	0.19	0.24	0.3	0.52	0.52
Size of reprocessed data mDST (latest version)	0.39	0	0.41	0.83	1.24
Size of prompt MC mDST (latest version)	1.74	0.61	1.33	1.46	1.02
Size of reprocessed MC mDST (latest version)	0.84	0	0.85	1.69	1.27
Size of prompt data mDST (all versions)	0.34	0.47	0.74	1.07	1.34
Size of reprocessed data mDST (all versions)	1	0.75	1.16	1.58	2.45
Size of prompt MC mDST (all versions)	4.3	3.54	3.68	3.4	3.8
Size of reprocessed MC mDST (all versions)	1.39	1.23	2.08	2.93	3.81
Size of data uDST (latest version)	0.8	0.34	1.04	1.95	2.55
Size of MC uDST (latest version)	3.66	0.98	3.53	5.1	3.74
Size of data uDST (all versions)	1.84	1.71	2.75	3.83	5.5
Size of MC uDST (all versions)	8.07	7.72	9.33	10.24	12.44

Table 17: Size of mDST and uDST samples (all replicas), in PB, for data and MC samples, both for the latest version (X) and for all stored versions (X, X-1, and X-2).

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 dataset. 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. In previous estimates, it was assumed that only a single copy of the files produced with release X would be stored. However, replication is handled automatically by Rucio, so an additional buffer space equivalent to the size of the full samples was added to the resource estimate. In this revised estimate, the space required for two replicas of files with release X is included and the additional buffer for files in production is removed. Additional details about the mDST and uDST sizes are accounted in Table 17, for reference.

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.

	2025	2026	2027	2028
CPU power to run one analysis (HS06*s/evt)	0.08	0.08	0.08	0.08
Number of concurrent analyses	300	300	300	300
Number of analysis cycles	2	2	2	2
CPU power to run all analyses (HS06*s/evt)	48	48	48	48
ntuple size for all analyses (kB/evt)	12	12	12	12

Table 18: Parameters of the analysis model.

6.10 Analysis

User analysis jobs should be run on skimmed uDST samples to produce ntuples that will be used for offline analysis and systematic studies. The uncertainties associated with analysis jobs are much larger and therefore more difficult to estimate in advance.

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 were adjusted in the previous resource estimate to better reproduce the observed amount of CPU power devoted to analysis. Once again, the actual usage thus far in 2023 was significantly higher than predictions. Therefore, the number of concurrent analyses was increased to 300 and the number of analysis cycles to 2 in order to bring expectations closer to actual usage.

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 analyses in 2023 was about 300. We expect this to remain fairly stable for 2024 and later.
- The number of analysis cycles (defined as the number of times an analysis is run on the data and MC samples) is set to 2 and is expected to remain constant. Even as analysis experience grows, analyses that become limited by systematics may require subsequent processing as each analysis evolves.
- The uncertainty is at the 50% level.

The total processing power to analyze one event (all analyses, all analysis cycles) is shown in Table 18 together with the parameters of the analysis model. The user ntuple 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 ntuple 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

	2024	2025	2026	2027	2028
Total tape (PB)	9.6	11.8	15	20.4	25.9
Total disk (PB)	25	19.4	25.3	30.3	37.7
Total CPU (kHS06)	520	247	492	547	643

Table 19: The resource estimate for years 2024 to 2028. CPU needs drop in 2025-2026 due to the reduced rate or data reprocessing and MC production. The storage units are PB = 10^{15} Bytes.

	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 20: The resource estimate for years 2024 to 2027 from the previous (2023) resource estimate, BELLE2-NOTE-TE-2023-004 `https://docs.belle2.org/record/3412`, for reference.

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.

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 19. 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 21 shows the pledged and approved computing resources for 2019-2024. Approval by the Belle II Financial Oversite Panel occurs in October of the preceding JFY (for example, JFY 2024 resources are approved in October 2023). 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 2025 to 2028 is included in the Appendix A.

	2019	2020	2021	2022	2023	2024
Total tape (PB)	10	1.4	3.2	8.8	8.6	9.6
Total disk (PB)	11	11	11	16.5	19.6	25
Total CPU (kHS06)	299	207	498	385	404	520

Table 21: The resources pledged and approved for 2019 through 2024.

8 Conclusion

We have presented the estimate of the computing resources needed by the Belle II collaboration in the years 2025 to 2028 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.

Appendix A: Resource Breakdown for years 2025 to 2028

Tape (PB)	2025	2026	2027	2028
Tape for raw data	9.06	11.51	15.72	19.93
Tape for hRaw data	2.63	3.34	4.56	5.78
Tape for cosmic data	0.15	0.15	0.15	0.15
Total tape	11.83	15	20.43	25.86

Table 22: Tape resources by data type (in PB).

Disk (PB)	2025	2026	2027	2028
Disk for data processing buffer	1.29	1.48	2.18	2.18
Disk for data mDST buffer	0.04	0.11	0.2	0.26
Disk for calibration	0.95	1.2	1.45	1.7
Disk for data mDST	1	1.47	2.1	3.8
Disk for random trigger files for MC	0.08	0.11	0.19	0.19
Disk for MC production buffer	0.09	0.33	0.47	0.34
Disk for MC mDST buffer	0	0	0	0
Disk for MC cDST	0.5	0.6	0.7	0.7
Disk for MC mDST	4.77	5.76	6.33	7.61
Disk for data and MC uDST	9.43	12.09	14.07	17.94
Disk for data and MC skimming buffer	0.2	0.69	1.06	0.94
Disk for data and MC uDST buffer	0	0	0	0
Disk for user ntuples	1.03	1.42	1.59	2
Disk for cosmic	0.01	0.01	0.01	0.01
Total disk	19.38	25.26	30.35	37.69

Table 23: Disk resources by data type (in PB).

CPU (kHS06)	2025	2026	2027	2028
CPU for data processing	17.1	21.6	37	37
CPU for data reprocessing	0	28.1	28.1	56.1
CPU for calibration	14	14	14	14
CPU for MC production	51.9	186.7	197	197.8
CPU for skimming	7.6	26.2	29.1	32.8
CPU for analysis	156.7	215.9	241.3	305
Total CPU	247	492	547	643

Table 24: CPU resources by usage (in kHS06).