BELLE2-NOTE-TE-2023-003
Version: 2.2

Computing Accounting for JFY 2022

The Belle II Computing Steering Group

May 28, 2023

1 Introduction

In this note we present the accounting information for the year 2022, covering the period from April 2022 through March 2023. In the first section, the computing activities originally planned for 2022 are summarized. In the second section, the activities undertaken are described, along with the accounting information. In the following, dates refer to JFY unless otherwise stated.

2 Planned activities and resource estimate for year 2022

In February 2021, when the plan of activities for 2022 was finalized, the expected integrated luminosity up to 2022 was on the order of 0.73 ab^{-1}. The focus of the computing plan was the processing of detector data, simulation of corresponding MC samples, skimming of detector and MC data samples, and support of analysis studies.

The planned activities were:

- Processing of the detector data.
- MC production. The ratio of the generic run-dependent MC event sample over the detector event sample is 4 for hadronic and τ events and smaller for low-multiplicity events. Another 1 ab^{-1} equivalent of run-independent MC is produced. In addition, samples of signal events for specific studies are centrally produced upon request of the Physics Analysis Groups.
- Skimming of detector and MC data.
- Reprocessing of the detector data. The understanding of the detector and the software quality will improve over time, faster at the beginning of the experiment, resulting in new software releases that will require reprocessing of the data to exploit improvements. One reprocessing campaign was planned. The reprocessing of the detector data with new releases of the software will trigger the reproduction of the corresponding MC data samples, to have consistent data sets, and the re-skimming of the detector and MC data sets.
- Physics analysis based on MC and reprocessed recorded data.

Taking into account the planned activities and the best knowledge of the parameters of their modeling available in February 2021, we estimated to need, for the year 2022, the amount of computing resources listed in Table 1. The rightmost column has the total amount of storage and CPU power actually used for different activities in 2022. Note that 0.2 PB of disk was requested for prompt calibration at BNL, while the 1.5 PB of space required for recalibration was provided as tape at DESY. The recalibration center at DESY also includes 451 TB of disk space, which was purchased for this purpose.

The discrepancy in the pledged and used CPU for user analysis in Table 1 comes primarily from an underestimation in the CPU required per analysis cycle and the number of concurrent analyses. The CPU estimate was revised in the resource estimate created in 2022 to reflect the actual use in 2021.

3 Data Production Activities in 2022

Data processing continued in 2022 with the well established scheme of prompt processing and reprocessing. All the data collected during 2022 have been calibrated and processed promptly, using the basf2 software release-06. This amounts to 174.9 fb^{-1} on-resonance, 17.99 fb^{-1} off-resonance and 19.3 fb^{-1} energy scan data. The latter were collected in fall 2021 and processed promptly with release-06. In addition to this, all the data collected before summer 2021, which were only available in processing performed with release-05, have been re-calibrated and reprocessed with release-06. This reprocessing is known as proc13. The recalibration has only been partial and was made necessary by the changes in the reconstruction introduced in release-06. Overall, 186.75 fb^{-1} were reprocessed in proc13. For future reprocessings we will reduce further the amount of recalibration, as the prompt calibration have been proven to be already of high quality.

On the Monte-Carlo production side we performed the first full campaign of run-dependent MC (MC15rd), taking into account real detector conditions at the time of data taking. The total sample size amounts to 4 times the collected luminosity for all the generic samples, and few million events per signal mode. The production of the generic samples have been completed, while the signal samples continue to be produced upon request from the physics working groups. Overall, the campaign was successful, despite being more time-consuming than anticipated due to a rather steep learning curve. The experience gained will be precious to complete the next MC campaign in a timely manner. Together with the run-dependent MC we also produced run-independent MC, which uses static detector conditions, to a total of 2 ab^{-1} equivalent luminosity.

Skimming of the data was performed following the prompt processing and proc13 and is now completed. Skimming of the run-independent MC is also completed, while it is ongoing for MC15rd.

4 Accounting

The amount of CPU power that the different countries participating in the Belle II collaboration have pledged, have actually provided, and that has been used on average in 2022 are shown in Table 2. The column "Pledged CPU" shows the amount of CPU power

	Estimated	Used
Tape for raw data (PB)	6.86	5.47
Tape for hRaw data (PB)	1.95	0.97
Tape for recalibration (PB)	1.50	0.92
Total tape (PB)	**10.30 (estimate)**	**7.36**
Disk for data processing buffer (PB)	0.08	1.80
Disk for data mDST buffer (PB)	0.35	0.58
Disk for data mDST (PB)	1.20	
Disk for random trigger files for MC (PB)	0.13	0.25
Disk for MC buffer (PB)	0.27	3.80
Disk for MC mDST buffer (PB)	0.89	
Disk for MC mDST (PB)	4.90	
Disk for MC cDST (PB)	0.30	
Disk for data and MC uDST (PB)	4.72	1.60
Disk for data and MC skim buffer (PB)	0.40	
Disk for data and MC uDST buffer (PB)	1.34	
Disk for user ntuples (PB)	0.18	
Disk for cosmic data (PB)	0.01	
Disk for calibration/recalibration	0.20	0.68
Total disk (PB)	**14.97 (estimate)**	**11.48**
CPU for data processing (kHS06)	9.99	12.57
CPU for data reprocessing (kHS06)	25.53	
CPU for calibration/recalibration (kHS06)	90.00	4.05
CPU for MC production (kHS06)	225.33	138.95
CPU for skimming (kHS06)	23.31	17.07
CPU for analysis (kHS06)	11.03	68.29
CPU for local resources (non-production) (kHS06)		120.00
Total CPU (kHS06)	**385 (estimate)**	**361**

Table 1: The resource estimate for year 2022. The rightmost column has the total amount of storage and CPU power used for different activities in 2022.

(in kHEPSpec06) that the different countries have pledged. The pledges are calculated according to the PhD fraction of each country. The column "Provided CPU" reports the amount of CPU (in kHEPSpec06) dedicated to Belle II at the different sites at the time of the last yearly survey (April 2022). While some CPUs are accessible only to Belle II, shared CPUs are used by different experiments according to a fair share algorithm that takes into account the amount of pledged CPU. It is important to note that the resources that a site provides, on top of "pledged" resources that are guaranteed to the Belle II collaboration, may include opportunistic resources or other resources that are only temporarily available and that will not necessarily be available in the future. The column "Used" shows the amount of CPU (in kHEPSpec06) used on average in 2022. These numbers are obtained by taking the average of the HS06 per core times the number of cores times the number of hours they were used for each site and then dividing by the average number of hours per month (720) to get the average HS06 value that can be compared with the pledged and provided CPU per site. Fig 1 displays the CPU power used in different countries per month. Note that for some sites the amount of CPU used is less than that pledged amount due to the fact that there were not enough Belle II jobs to make full use of the available resources.

The used CPU power is extracted from the EGI portal for the sites who export their accounting to it. It is extracted from the KEK Dirac portal for the other sites. From April 2022 through March 2023, Belle II has used an average of 237 kHEPSpec06 of CPU power on the Grid. Prompt Calibration done at BNL and recalibration done at DESY required on average 4 kHEPSpec06. Some detector studies and some physics analysis are done on local resources and 120 kHEPSpec06 have been devoted to that.

The amount of disk storage (in TB) that the different countries participating in the Belle II collaboration have pledged, have actually provided, and the maximum amount in use during 2022 thus far are shown in Table 3. The column "Pledged Disk" shows the amount of disk storage that the different countries should have provided. The pledges are calculated according to the PhD fraction of each country. The column "Provided Disk" reports the amount of disk storage made available to Belle II at the different sites at the time of the last yearly survey (April 2022). The storage available only to local Belle II users and not configured as a storage element is not included. The last column shows the maximum amount of storage in use during 2022. The amount of tape storage (in PB) that the different Raw Data Centers have actually provided, along with its use, is shown in Table 4.

Country	Site	Pledged CPU	Total CPU provided	Total used
Armenia	-	1.6	0	0
Australia	Melbourne	6.4	28	4.8
Austria	HEPHY	4	4.8	2.1
Canada	Uvic	11.6	100	38.6
China	Fudan			0
China	USTC			0
China	Shandong			0.2
China	IHEP			7.3
China	Total	24	15	7.5
Czech	CESNET	3.2	16.4	4.6
France	IN2P3CC			4.2
France	IPHC			0
France	LAL			1.6
France	Total	14	14	5.8
Germany	DESY			21.2 (17.7 local)
Germany	KIT			21.7
Germany	MPPMU			1.1
Germany	Total	61.5	180.5	41.4 (17.7 local)
India	-	9.6	24.7	0
Israel	-	3.2	2.7	0.9
Italy	CNAF			13.4
Italy	Cosenza			0
Italy	Frascati			1
Italy	LNL			0.2
Italy	Napoli			7.5
Italy	Pisa			4.2
Italy	Roma3			1.1
Italy	Torino			3.1
Italy	Total	47.5	150.6	30.5
Japan	KEK			41.4 (105.3 local)
Japan	NDU			1.7
Japan	TMU			0.6
Japan	Niigata			0.4
Japan	KMI			0.2
Japan	Total	72.3	104.1	44.3 (105.3 local)
Korea	KISTI	16	9.9	0.8
Malaysia	-	0.8	0	0
Mexico	-	4.8	2.4	0
Poland	CYFRONET	5.6	2	1.3
Russia	BINP			0.8
Russia	MIPT			0
Russia	Total	20.8	18	0.8
Saudi Arabia	-	1.6	0	0
Slovenia	SIGNET	8	38.5	18.7
Spain	-	0.8	0	0
Taiwan	NTU	5.6	18.3	0
Thailand	-	3.2	0	0
Turkey	ULAKBIM	1.6	0.9	0.5
Ukraine	-	2.4	0	0
USA	BNL			31.5 (0.4 local)
USA	Univ of Miss			0.4
USA	Total	53.6	122.4	31.9
VietNam	-	1.6	0	0
Total All Countries	-	385.2	853.3	237.1 (123.4 local)

Table 2: CPU power (in kHEPSpec06) that the different countries participating in the Belle II collaboration have pledged, have actually provided, and that has been used in 2022.

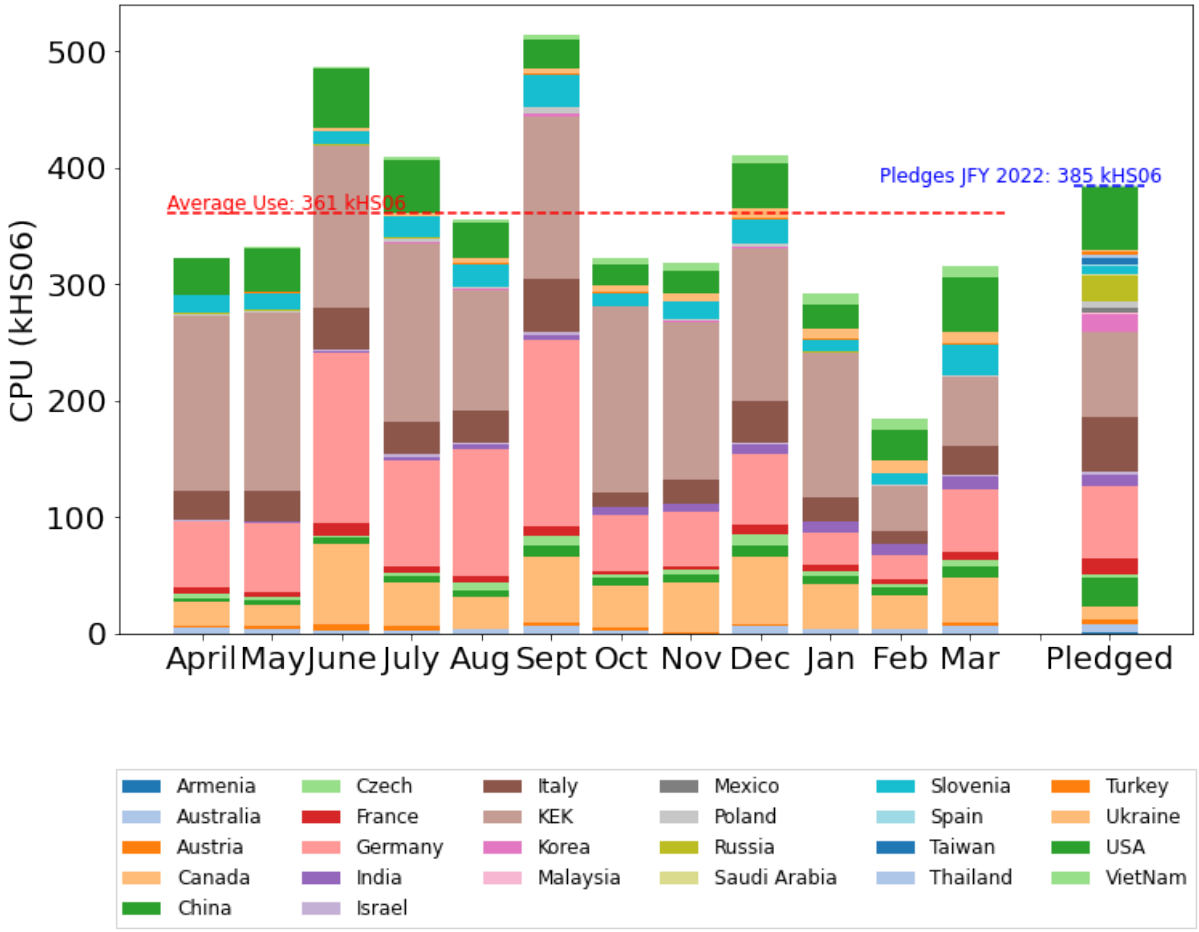

Figure 1: CPU power per month (in kHEPSpec06) used in different countries during 2022, including both grid and local resource usage.

Country	Site	Pledged Disk	Pledged Disk (grafana)	Provided Disk	Used Disk
Armenia		71	0	0	0
Australia	Melbourne	280	50	50	0
Austria	HEPHY	177	250	250	238
Canada	Uvic	320	600	665	630
China	IHEP	1065	250	280	265
Czech	CESNET	142	100	238	227
France	IN2P3CC	320	309	324	277
France	IPHC	100	72	88	12
France	LAL		0	23	22
France	Total	420	381	435	310
Germany	DESY	1720	1589	2000	1200
Germany	KIT	1720	790	790	701
Germany	MPPMU	220	185	200	113
Germany	Total	3660	2564	2990	2014
India		426	0	0	0
Israel		142	60	60	0
Italy	CNAF	820	820	819	665
Italy	Frascati		10	11	10
Italy	Napoli		390	590	353
Italy	Pisa		200	209	11
Italy	Roma3		2	2	1
Italy	Torino		350	350	288
Italy	Total	1810	1772	1981	1328
Japan	KEK		3100	3650	2986
Japan	KMI		300	300	120
Japan	Total	2910	3400	3950	3106
Korea	KISTI	710	100	110	72
Malaysia		35	0	0	0
Mexico		213	0	0	0
Poland	CYFRONET	248	0	11	8
Russia		923	0	0	0
Saudi Arabia		71	0	0	0
Slovenia	SIGNET	355	1210	1210	1100
Spain		35	0	0	0
Taiwan	NTU	248	791	869	36
Thailand		142	0	0	0
Turkey	ULAKBIM	71	130	143	36
Ukraine		106	0	0	0
USA	BNL	1970	2300	2320	2110
USA	Univ of Miss		0	0	0
USA	Total	1970	2300	2320	2110
VietNam		71	0	0	0
Total	-	16624	13958	15562	11481

Table 3: Storage (in TB) that each country has pledged and that has actually provided in 2022. The last column shows the amount of used storage.

Site	Pledge	Raw	hRaw	Other	Total
KEKCC	4.40	2.66	0.398	0	3.058
Canada - Uvic	0.34	0.122 (test)	0	0	0.122 (test)
France - IN2P3	0.34	0.12	0.029	0	0.149
Germany - DESY	0.22	0.089	0.018	0.921 (calib)	1.027
Germany - KIT	0.22	0.103	0.021	0	0.124
Italy - CNAF	0.45	0.208	0.042	0	0.249
USA - BNL	2.83	2.17	0.461	0	2.631

Table 4: Tape storage (in PB) provided by different Raw Data Centers in 2022. For the time being, the University of Victoria is storing raw data on disk with backup on tape.