Belle II Computing Model

F. Bianchi
INFN & University of Torino

Pisa Meeting, April 10th 2017

Introduction

- The Belle II computing model
- Accounting: CPU and storage used in 2015 and 2016
- Resource estimate for the years 2018 to 2021.

Belle II Computing Model

Belle II Computing Model (1)

- Raw data coming out of the DAQ system, are permanently stored (two replica), calibrated and processed.
  - A second copy of raw data will be permanently stored for safety
- Fully reconstructed events are stored in the miniDST format.
- Monte Carlo events are simulated and reconstructed using the same software used to process detector events and then also stored in miniDST format.
  - The MC/data luminosity ratio will be 4 in 2018, 3 in 2019, 2 in 2020 and 1 afterwards.
- Two replica of data and MC mDST will be stored for safety and to avoid processing bottleneck

Belle II Computing Model (2)

- Detector and Monte Carlo events miniDST are then "skimmed" to create set of selected events that suit a specific group of physics analysis.
  - During the skimming step, additional information are computed and added to the events.
  - The output of the skimming step will consist of either index files that contain "pointers" to events in miniDST format or deep copies of events in microDST format.
  - The feasibility of using the index file technology is currently under investigation and we plan to use deep copy skims in microDST format at least for the first few years of data taking.

- Two copies of microDST will be stored for safety and to avoid bottleneck in data access.

Belle II Computing Model (3)

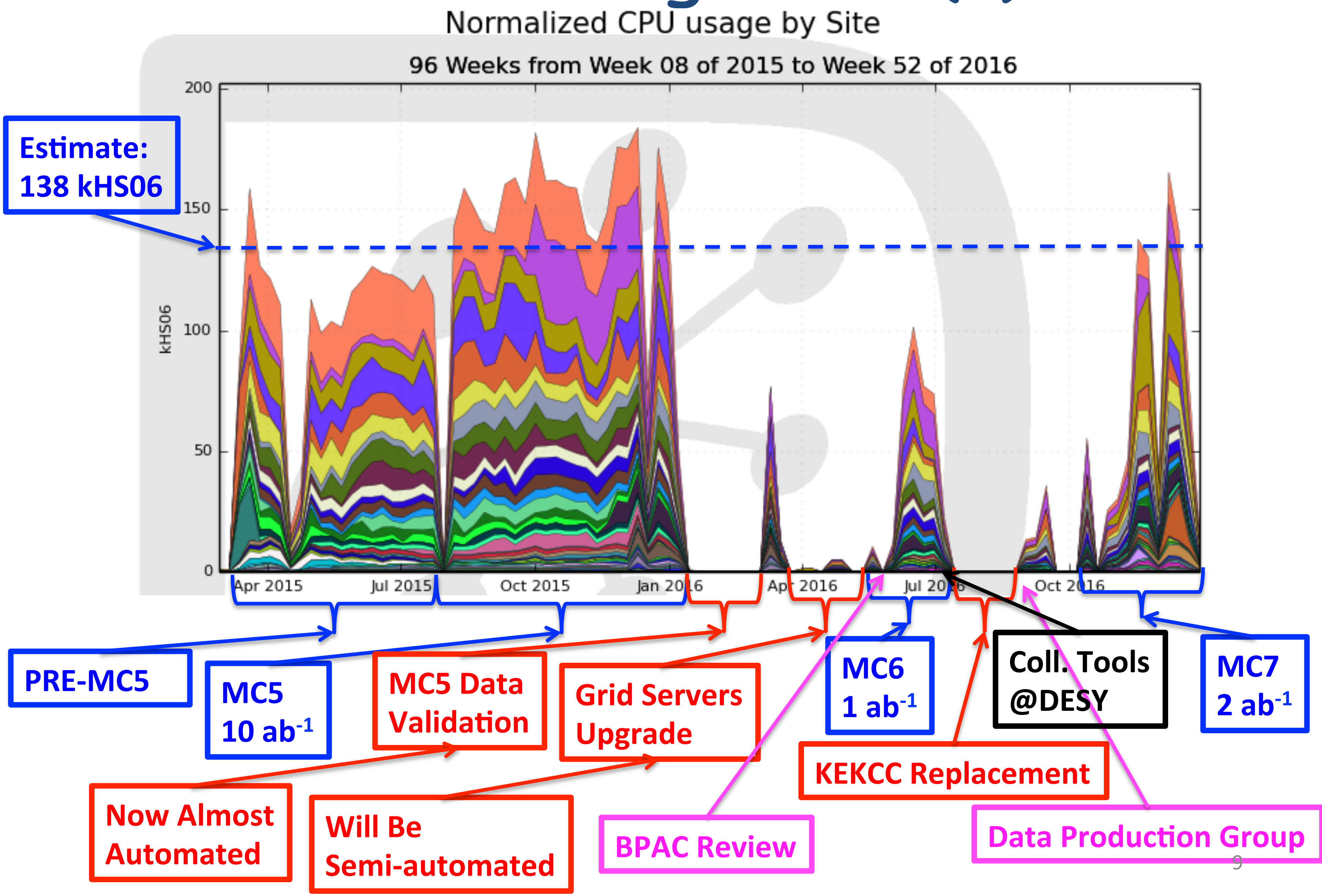
- The understanding of the detector and the quality of the software will increase over time, resulting in new releases of the software that will require reprocessing of the data to exploit the improvements.
  - Reprocessing of detector data is expected to trigger the re-creation of the corresponding Monte Carlo data samples and the skimming of these new data samples.
  - We plan to have two reprocessing per year in 2018 and 2019, one reprocessing in 2020 and a reprocessing every two years afterwards.
- The user analysis will run on skimmed events, unless a suitable skim is unavailable.
  - User analysis job will produce n-tuples that will be transferred over the network to local farms for further analysis
- Detector data processing, simulation and skimming will be centrally managed, while physics analysis will be users responsibility.

Classification of Belle II Computing Centers

- KEK is the host laboratory of the Collaboration. It receives raw data from the Belle II's online computing farm and records them on permanent mass storage. It also performs reconstruction of the data, Monte Carlo production and provides resources for end-user analysis.
- Raw Data Centers receive the raw and reconstructed data, providing a distributed permanent backup of the raw data, permanent storage and management of data needed during the analysis process, and offer a Grid-enabled data service. They also perform re-processing of raw data, Monte Carlo production and provide resources for end-user analysis.
- Regional Data Centers provide Grid-enabled disk storage to host a partial copy of the reconstructed data and concentrate on tasks such as simulation, end-user analysis.
- Monte Carlo Production Centers provide resources for Monte Carlo production and optionally end-user analysis.
- Local farms provide resources for n-tuple level analysis

Accounting

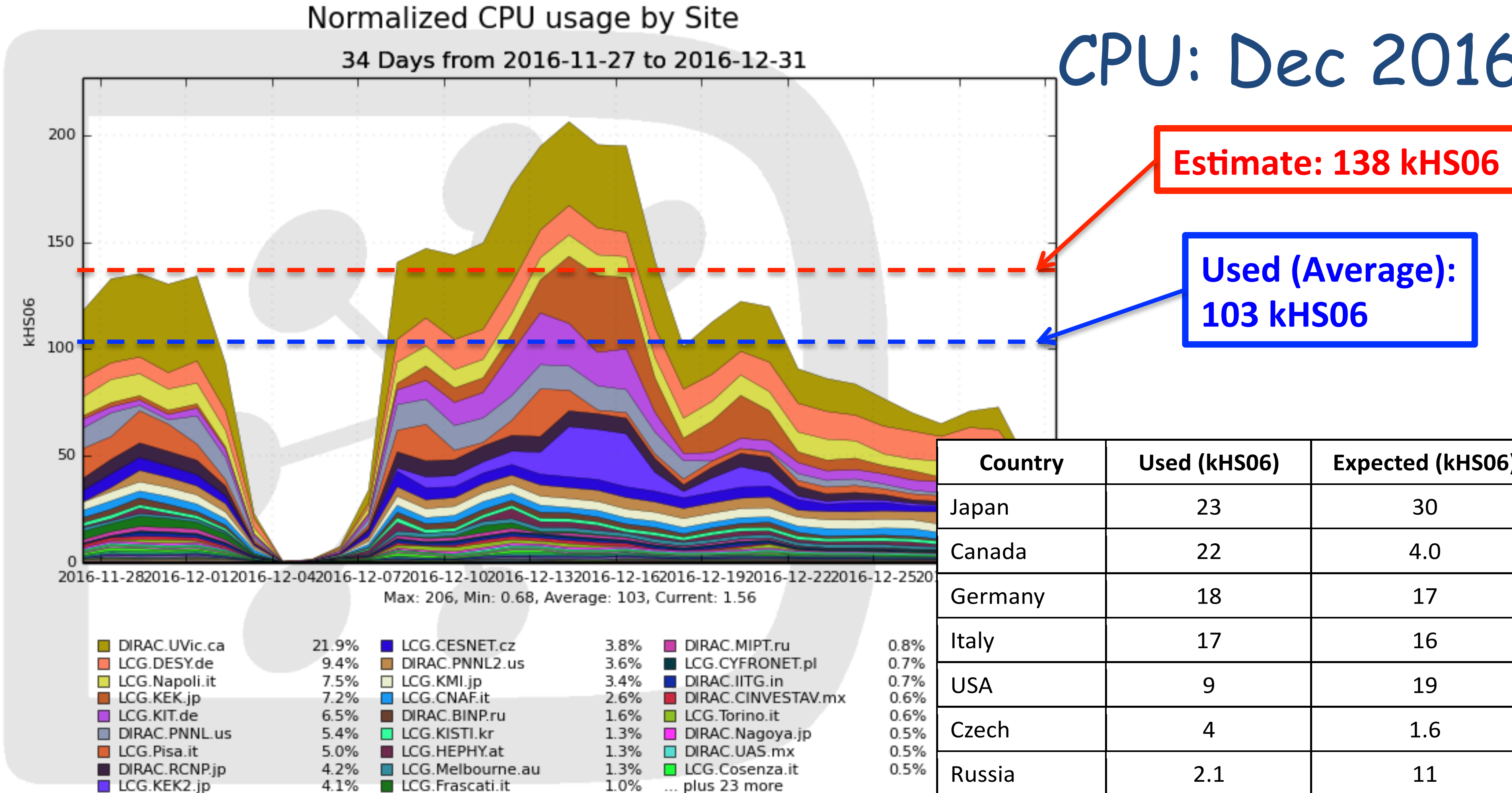
Accounting: CPU (1)



Accounting: CPU (2)

- According to the estimate, with the current input parameters, we should have used 138 kHEPSpec06 for 10 months.
- According to Dirac accounting in 2016 we used an average of 30.4 kHEPSpec06 for 12 months, that is 36.5 kHEPSpec06 for 10 months
- However many of the planned activities were not done on the grid.
- The CPU power required by what we did on the grid should have been around 45 kHEPSpec06 for 10 months.
  - But more then 50% of the MC production was done with older releases that were faster because included less functionality
- Two messages:
  - The CPU required by what we did on the grid compares with our prediction at the 20% level
  - We were able to keep the production system running only for a fraction of time
- But we are improving….

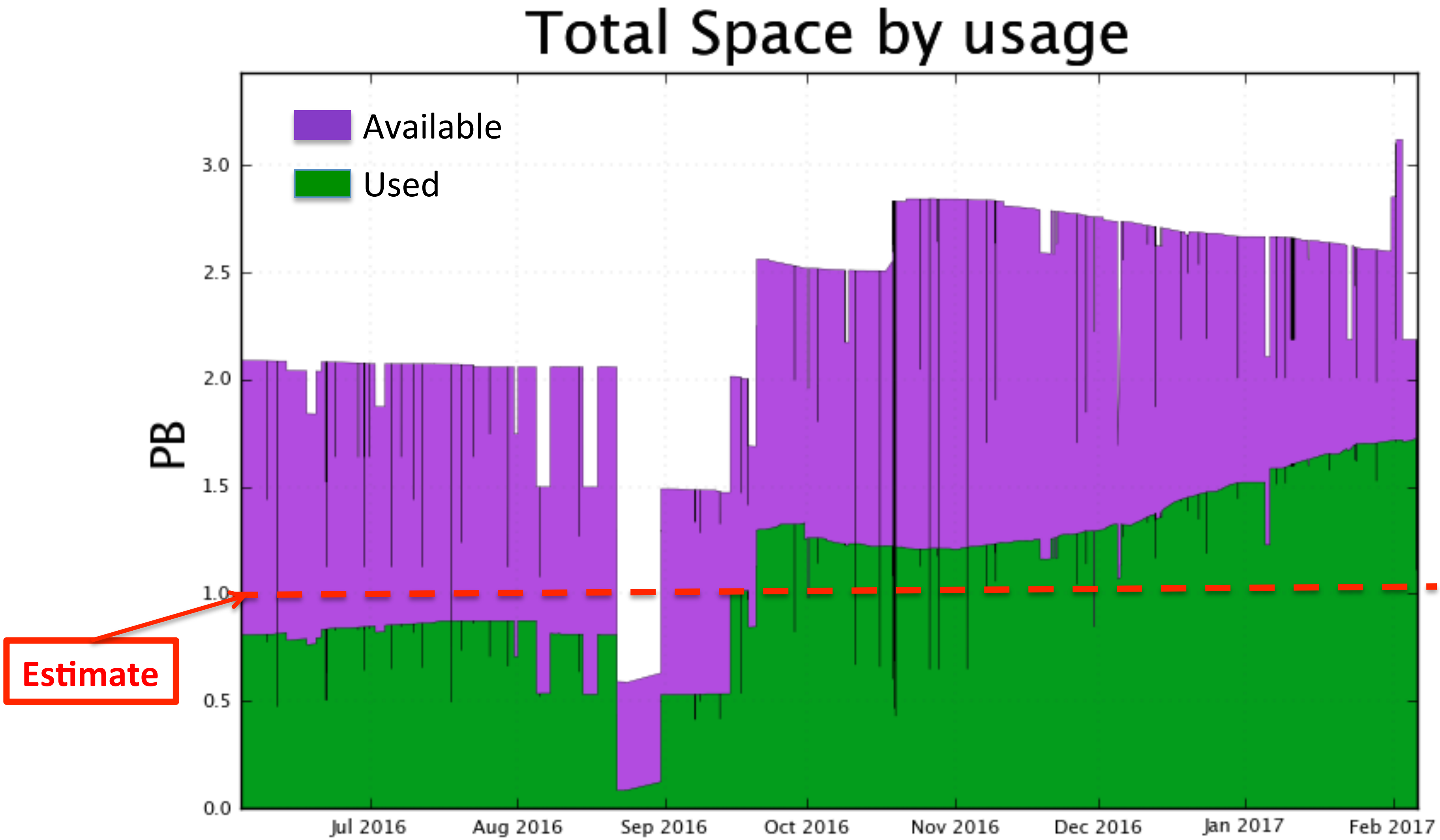
CPU: Dec 2016

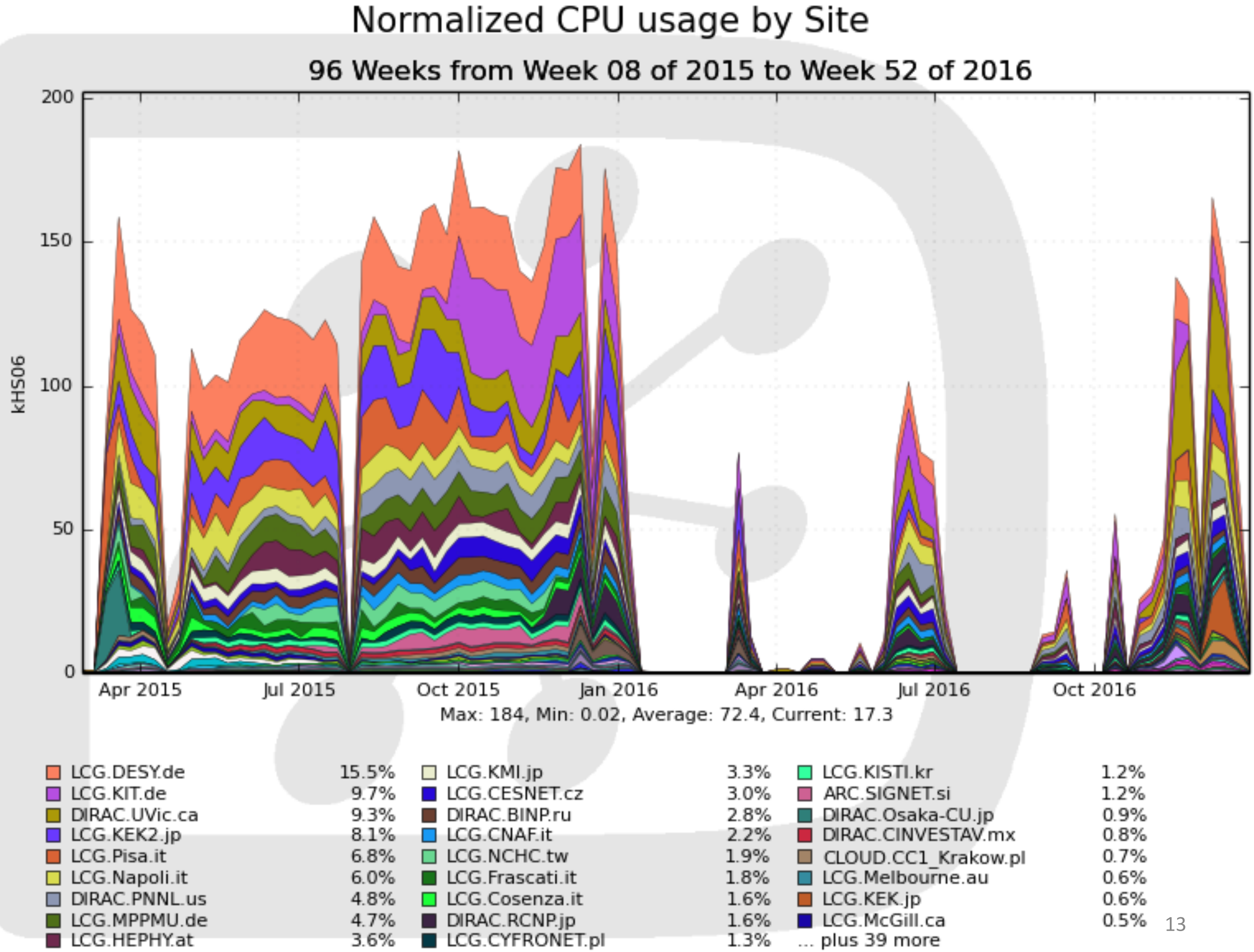


Country	Used (kHS06)	Expected (kHS06)
Japan	23	30
Canada	22	4.0
Germany	18	17
Italy	17	16
USA	9	19
Czech	4	1.6
Russia	2.1	11
Korea	2.0	6.8
Austria	1.4	2.4
Australia	1.4	4.0
Mexico	1.1	2.4
Poland	0.8	2.8
India	0.7	4.4
….		
Total	103	138

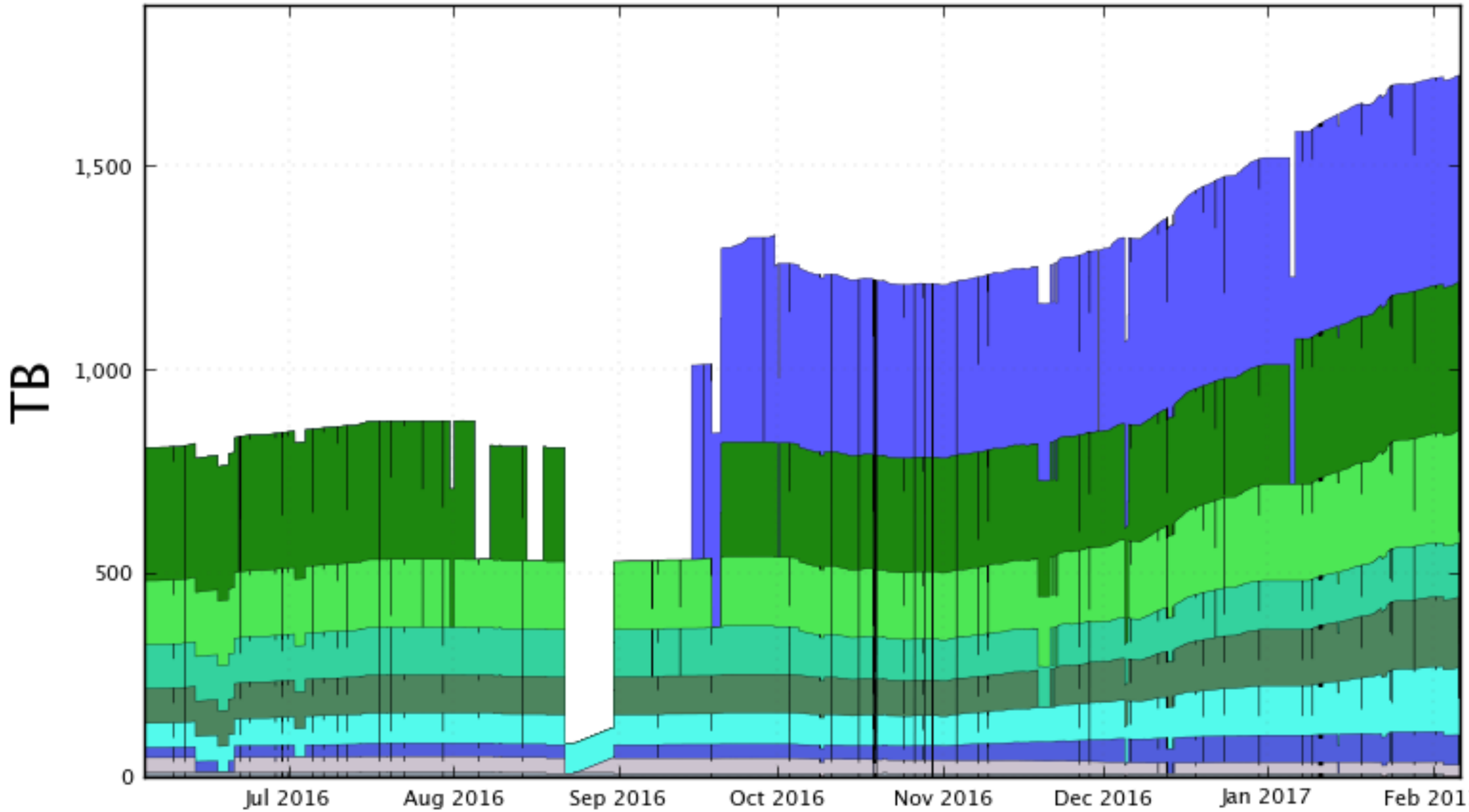
- Our peak use has been 40% more then our expected level
- Continuity looks better then previously, but there is still room for improvement

Accounting: Storage

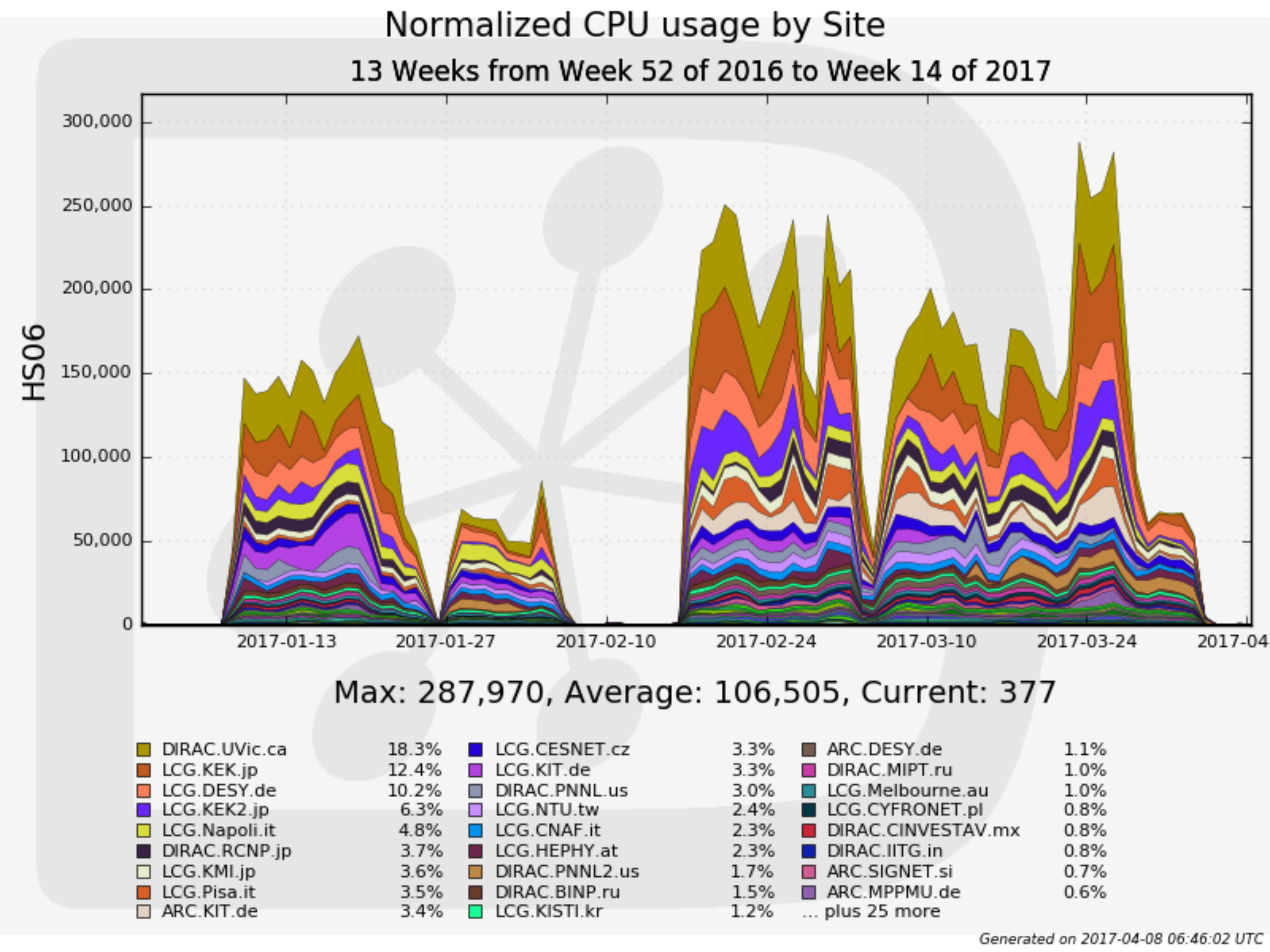



Normalized CPU usage by Site
96 Weeks from Week 08 of 2015 to Week 52 of 2016
kHS06
0
50
100
150
200
Apr 2015
Jul 2015
Oct 2015
Jan 2016
Apr 2016
Jul 2016
Oct 2016
Max: 184, Min: 0.02, Average: 72.4, Current: 17.3
LCG.DESY.de 15.5%
LCG.KIT.de 9.7%
DIRAC.UVic.ca 9.3%
LCG.KEK2.jp 8.1%
LCG.Pisa.it 6.8%
LCG.Napoli.it 6.0%
DIRAC.PNNL.us 4.8%
LCG.MPPMU.de 4.7%
LCG.HEPHY.at 3.6%
LCG.KMI.jp 3.3%
LCG.CESNET.cz 3.0%
DIRAC.BINP.ru 2.8%
LCG.CNAF.it 2.2%
LCG.NCHC.tw 1.9%
LCG.Frascati.it 1.8%
LCG.Cosenza.it 1.6%
DIRAC.RCNP.jp 1.6%
LCG.CYFRONET.pl 1.3%
LCG.KISTI.kr 1.2%
ARC.SIGNET.si 1.2%
DIRAC.Osaka-CU.jp 0.9%
DIRAC.CINVESTAV.mx 0.8%
CLOUD.CC1_Krakow.pl 0.7%
LCG.Melbourne.au 0.6%
LCG.KEK.jp 0.6%
LCG.McGill.ca 0.5%
... plus 39 more

Used Space by site



DESY-TMP-SE	24.9%	CNAF-TMP-SE	10.0%	SIGNET-TMP-SE	3.7%
KEK2-TMP-SE	24.5%	KMI-TMP-SE	9.1%	CESNET-TMP-SE	2.9%
Napoli-TMP-SE	16.2%	KIT-TMP-SE	7.9%	PNNL-TMP-SE	0.8%



Normalized CPU usage by Site
13 Weeks from Week 52 of 2016 to Week 14 of 2017
HS06
300,000
250,000
200,000
150,000
100,000
50,000
0
2017-01-13
2017-01-27
2017-02-10
2017-02-24
2017-03-10
2017-03-24
2017-04
Max: 287,970, Average: 106,505, Current: 377
DIRAC.UVic.ca 18.3%
LCG.KEK.jp 12.4%
LCG.DESY.de 10.2%
LCG.KEK2.jp 6.3%
LCG.Napoli.it 4.8%
DIRAC.RCNP.jp 3.7%
LCG.KMI.jp 3.6%
LCG.Pisa.it 3.5%
ARC.KIT.de 3.4%
LCG.CESNET.cz 3.3%
LCG.KIT.de 3.3%
DIRAC.PNNL.us 3.0%
LCG.NTU.tw 2.4%
LCG.CNAF.it 2.3%
LCG.HEPHY.at 2.3%
DIRAC.PNNL2.us 1.7%
DIRAC.BINP.ru 1.5%
LCG.KISTI.kr 1.2%
ARC.DESY.de 1.1%
DIRAC.MIPT.ru 1.0%
LCG.Melbourne.au 1.0%
LCG.CYFRONET.pl 0.8%
DIRAC.CINVESTAV.mx 0.8%
DIRAC.IITG.in 0.8%
ARC.SIGNET.si 0.7%
ARC.MPPMU.de 0.6%
... plus 25 more
Generated on 2017-04-08 06:46:02 UTC

Resource Estimate

Change in Luminosity Profile

June 2016

Year	2017	2018	2019	2020	2021
Luminosity (ab^{-1}/year)	0.23	0.31	2.82	6.31	10.26
Integrated Luminosity (ab^{-1})	0.23	0.54	3.36	9.67	19.93

December 2016

Year	2017	2018	2019	2020	2021
Luminosity (ab^{-1}/year)	0.00	0.54	2.82	6.31	10.26
Integrated Luminosity (ab^{-1})	0.00	0.54	3.36	9.67	19.93

February 2017

Year	2017	2018	2019	2020	2021
Luminosity (ab^{-1}/year)	0.00	0.21	2.82	6.31	10.26
Integrated Luminosity (ab^{-1})	0.00	0.21	3.03	9.34	19.60

Assuming 9 months running per year and nominal luminosity profile

Phase 2: 0.04 ab^{-1}
Phase 3: 0.17 ab^{-1} assuming Phase 3 start in Nov 2018

Accepted Cross Sections (1)

Class of events	Accepted Cross Section (nb)
Y(4S)	1.05
ccbar	1.30
uds	2.39
$\tau^+\tau^-$	0.86
Subtotal hadronic + τ	5.60
$\mu^+\mu^-(\gamma)$	1.15
$\gamma\gamma(\gamma)$	0.50
$e^+e^-(\gamma)$	2.00
$e^+e^-e^+e^-$	0.50
$e^+e^-\mu^+\mu^-$	0.50
Subtotal low multiplicity	4.65
Sum (all)	10.25

Accepted Cross Sections (2)

- We plan to accept all the hadronic and τ events, while the other low multiplicity events will be pre-scaled.
- At BaBar and Belle essentially all the hadronic and τ events have been used to extract interesting physics results and we expect the situation to be similar at Belle II.
- Preliminary studies done on Belle II skimming code show that between 70% and 80% of hadronic and τ events will be selected by some skim.
- The low multiplicity events will be used to extract physics results and for detector and for calibration studies.
  - They will use around 15% of the CPU and storage resources.

CPU Power and mDST size

- Measured in release-00-08-00 for different classes of events and different background levels

- Additional factors to account for planned improvement of the code and background uncertainties.

- See BELLE2-NOTE-TE-2016-011 for details

History of Changes on Input Parameters

Raw Data Compression
HLT data reduction

	Jun 2016	Dec 2016	Feb 2017
Raw Data Size (kB/ev)	300	100 ± 25	100 ± 25
CPU for Reconstruction (HEPSpec *s /ev)	45	22.0 ± 4.9	21.5 ± 4.4
CPU for MC (HEPSpec *s /ev)	100	63.2 ± 13.3	60.0 ± 12.7
Detector mDST (kB/ev)	20	5.0 ± 1.8	6.8 ± 3.2
MC mDST (kB/ev)	25	7.0 ± 2.6	9.0 ± 4.2

Software measurements on rel 00-07-01

Software measurements in rel 00-08-00
Increase in mDST size comes from a new much improved clustering algorithm in ECL

Resource Estimate Evolution

(errors from uncertainties on impact of software developments)

June 2016

	2018	2019	2020	2021
Tape	4.2	13.8	39.6	81.6
Disk	5.4	8.2	20.4	33.9
CPU	474	609	708	881

LHCb (2017)

Tape (PB)	60
Disk (PB)	28
CPU (kHS06)	375

	2018	2019	2020	2021
Tape	2.2±0.3	6.4±1.6	18.2±4.5	36.5±9.3
Disk	5.5±1.3	11.9±2.9	20.3±4.8	20.4±4.7
CPU	387±41	478±75	541±79	643±91

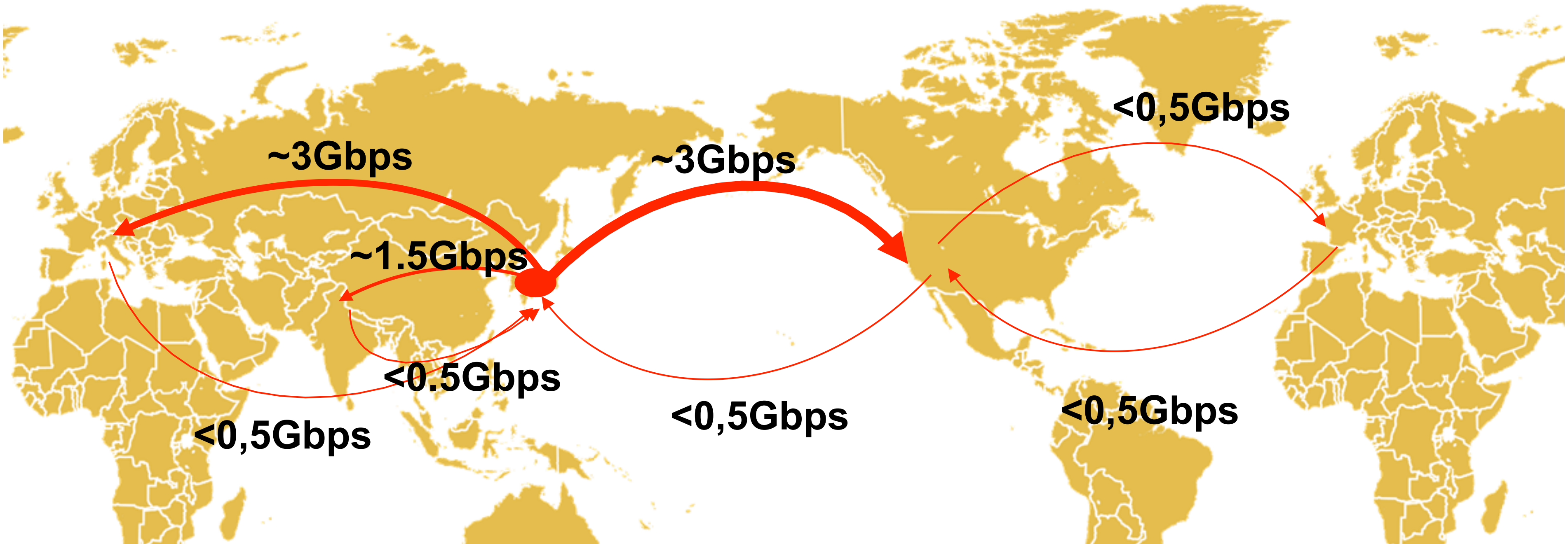
Dec 2016
Raw data compression
Soft Meas on rel 00-07-01
Skim + Analysis Model

	2018	2019	2020	2021
Tape	1.6±0.1	5.8±1.4	17.6±4.3	35.9±9.1
Disk	3.5±1.3	14.1±3.1	25.9±5.5	26.3±5.6
CPU	187±27	400±67	473±75	563±86

Feb 2017
Luminosity Update
Soft Meas on rel 00-08-00

Total traffic per regions

N.B. USER ANALYSIS TRAFFIC NOT INCLUDED



	2017	2018	2019	2020	2021	2022	2023	2024
KEK to ASIA	0,08	0,08	0,06	0,06	1,13	1,37	1,39	1,42
KEK to EU	0,19	0,18	0,12	1,82	2,27	2,74	2,81	2,87
KEK to USA+MEX+CANADA	0,08	0,24	1,56	1,75	2,18	2,64	2,67	2,70
EU to KEK	0,19	0,18	0,10	0,11	0,12	0,15	0,24	0,30
USA+MEX+CANADA to KEK	0,08	0,09	0,07	0,06	0,07	0,09	0,14	0,17
ASIA to KEK	0,08	0,08	0,05	0,05	0,06	0,07	0,11	0,14
USA+MEX+CANADA to EU	0,16	0,17	0,13	0,12	0,13	0,16	0,26	0,32
EU to USA+MEX+CANADA	0,16	0,17	0,09	0,10	0,12	0,14	0,22	0,28

Procurements in 2018

- Year X procurements should cover Belle II needs until when the Year X+1 procurements come online
  - The exact date can be different for different countries
- Assuming that new procurement are online on April 1st, then Year X procurement should cover needs until April 1st of year X+1

	2018	25 % of 2019 increment	Total
Tape (PB)	1.6	1.0	2.6
Disk (PB)	3.5	2.7	6.2
CPU (kHEPSpec06)	187	53	240

Backup

CPU Power for Data Reconstruction (1)

Class of events	HEPSpec06 * s / ev
Υ(4S)	27.56
ccbar	20.60
uds	15.45
$\tau^+\tau^-$	7.41
$\mu^+\mu^-(\gamma)$	5.45
$\gamma\gamma(\gamma)$	5.45
$e^+e^-(\gamma)$	5.45
$e^+e^-e^+e^-$	5.45
$e^+e^-\mu^+\mu^-$	5.45
Average on classes of events	12.13
Including foreseen software upgrade	17.6 ± 4.0
Including background uncertainty	19.6 ± 4.3
Scale factor for calibration step	1.10
Processing power for raw data reconstruction	21.5 ± 4.4

CPU Power for Data Reconstruction (2)

- “Average on classes of events” is the average of the measurements for different classes of events weighted with the accepted cross section.
- “Including foreseen software upgrade” is the average processing power scaled to include an estimate of the impact of the software developments.
- “Including background uncertainty”. To take into account the uncertainty on background level, we use a processing power that is halfway between the values with nominal background and with x2 background
- “Scale factor for calibration step “ is an educated guess for the CPU power required by the calibration step (the corresponding software is under development)

Foreseen Improvement in Reconstruction Software

- Add a more realistic magnetic field map.
- Implement a "region of interest" algorithm in the PXD.
- Implement a new track finder in the VXD.
- Add a cross detector track finder
- Optimize the track fitter.
- Perform track fit with multiple mass hypothesis.
- Implement a new clustering algorithm in the electromagnetic calorimeter (done !)

CPU Power for MC Simulation (1)

Class of events	HEPSpec06 * s / ev
Y(4S)	74.90
ccbar	64.14
uds	58.43
$\tau^+\tau^-$	39.46
$\mu^+\mu^-(\gamma)$	31.89
$\gamma\gamma(\gamma)$	31.89
$e^+e^-(\gamma)$	31.89
$e^+e^-e^+e^-$	31.89
$e^+e^-\mu^+\mu^-$	31.89
Average on classes of events	47.21
Including foreseen software upgrade	52.4 ± 10.1
Including background uncertainty	60.0 ± 12.7
Processing power for Monte Carlo	60.0 ± 12.7

CPU Power for MC Simulation (2)

- "Average on classes of events" is the average of the measurements for different classes of events weighted with the accepted cross section.
- "Including foreseen software upgrade" is the average processing power scaled to include an estimate of the impact of the software developments.
- "Including background uncertainty". To take into account the uncertainty on background level, we use a processing power that is halfway between the values with nominal background and with x2 background

Foreseen Improvement in MC Software

- Add level 1 trigger simulation.
- Perform HLT simulation only for a fraction of events.
  - We'll run the same reconstruction code on HLT as on offline, but the used calibration constants and software version might be different.
  - To do a full HLT simulation means to run the reconstruction algorythms twice.
  - Some CPU power might be saved if a HLT decision can be made after part of the reconstruction chain. We plan to perform a full HLT simulation in 2018 and 2019 and do it only for a fraction of events in the following years.
- Use background overlay.
  - So far signal and background hits are merged before digitization to simulate the effect of background.
  - To have a more realistic simulation of the background it is planned to merge random trigger data and simulated signal events after digitization, a technique called background overlay.

miniDST size (1)

Class of events	Detector events (kB/event)	MC events (kB/event)
Y(4S)	6.75	10.13
ccbar	5.75	8.29
uds	5.03	7.00
$\tau^+\tau^-$	3.41	4.46
$\mu^+\mu^-(\gamma)$	2.75	3.04
$\gamma\gamma(\gamma)$	2.75	3.04
$e^+e^-(\gamma)$	2.75	3.04
$e^+e^-e^+e^-$	2.75	3.04
$e^+e^-\mu^+\mu^-$	2.75	3.04
Average on all classes	4.13	5.47
Including software upgrade and optimization	5.4 ± 2.4	7.1 ± 3.0
Including background uncertainty	6.8 ± 3.2	9.0 ± 4.2
mDST size (kB)	6.8 ± 3.2	9.0 ± 4.2

miniDST size (2)

- "Average mDST size" is the average of the measurements for different classes of events weighted with the accepted cross section.
- "Including foreseen software upgrade" is the mDST size scaled for foreseen software upgrade and packing optimization.
- "Including background uncertainty". To take into account the uncertainty on background level, we use a mDST size that is halfway between the values with nominal background and with x2 background

Comparison with Belle and LHCb

	Belle II release-00-08-00	Belle
Simulation	47 (sim 23, PXDdigi 7.8, bgmix 5.8, CDCdigi 4.6	7.3 (3.9 gsim, 3.2 acc)
Tracking	21 (CombFit 4.0, CDCFit 3.7, V0 3.4, Ext 3.1)	5.4 (3.1 trasan, 1.2 trak)
PID	5.6 (KLMExpert: 2.2, TOP 1.9)	0.7

$\Upsilon(4S)$, $B \to DD$, on KEKCC, 1xBG

	Belle II	**LHCb**
Raw Data Size (kB/ev)	100 ± 25	60
CPU for Reconstruction (HEPSpec *s /ev)	21.5 ± 4.4	20
CPU for MC (HEPSpec *s /ev)	60.0 ± 12.7	1000
Detector mDST (kB/ev)	6.8 ± 3.2	20
MC mDST (kB/ev)	9.0 ± 4.2	200

CPU Power for Reconstruction (1)

- “The committee noticed that the comparison of the processing times for event reconstruction between the Belle II and LHCb shows a significant improvement potential for the Belle II experiment. “

- Our comparison with Belle software leads to the same conclusion.

- However....

CPU Power for Reconstruction (2)

- Belle II is a detector more complex then Belle and works in an higher background environment.
  - It has sub-detectors that require complicated reconstruction algorithms:
    - Pattern recognition in Pixel Detector
    - Tracking of optical photons in the TOP.
- We have much lower momentum (curling) tracks than LHCb. This makes pattern recognition and track fitting more challenging.
  - We need to be more aggressive in fitting curling tracks and rejecting fake tracks because they affect missing energy analyses.
- BaBar was quite aggressive in pushing tracking performances and had to track optical photons in the DIRC
  - BaBar reconstruction required a CPU processing power of the order of 30 HEPSpec06*sec/ev
- Reconstruction software is still under development and less mature than the one of LHCb and Belle
  - Main target now is to provide the required functionality. Not all potential for optimizations has been exploited.
  - We plan to be more aggressive in code optimization from now on

Skimming (1)

- Skimming software is evolving rapidly, the existing code is only a fraction of the final one and large extrapolations are needed to estimate the amount of computing resources needed for skimming.
- Baseline format of skimmed events is microDST
  - miniDST of skimmed events with additional data objects:
    - 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.
    - Results from B and D meson full reconstruction, continuum suppression and other complex algorithms can be saved into dedicated analysis objects.
  - This allows for preprocessing that reduces the CPU requirements of the final analysis step.

Skimming (2)

- Parameters considered in the calculation for skim CPU and disk usage:
  - The number of skim groups.
    - One skim refers to the criteria associated to the analysis of a single publication.
    - Based on experience from Belle and BaBar, the use of more inclusive skims, grouping many decay modes of similar topology, can be very convenient.
    - We assume one skim group per analysis working group for this calculation, including systematics.
  - The total skim acceptance for each skim group.
  - The multiplicity of combined particle candidates analyzed while skimming on the miniDST.
  - The fraction of the HLT output that is from low-multiplicity events.
  - The ratio of the time taken for skimming algorithms to run on hadronic versus low-multiplicity events.

Skimming (3)

- The size of the microDST is dependent on the candidate multiplicity.
  - Varies greatly among analysis types with different selected event topologies.
  - We have estimated multiplicities for each skim group.
- Taking into account the total number of events found by at least one skim, the fraction of events selected by more then one skim, and the acceptance fractions of the skimming step, we expect the number of events in the microDST files to be approximately 50% larger than the miniDST events they are based upon.

LHCb: 4 HEPSpec * s /event

Processing power for skimming (HEPSpec06 * s / ev)	3.50± 2.0
(Size of a microDST event/(Size of a miniDST event)	1.15
(Events in microDST sample) / (Events in miniDST sample)	1.50

Analysis (1)

- User analysis jobs are run on skimmed microDST data to produce n-tuples that will be used to perform the final fit and systematic studies.
- Some steps such as vertexing, particle combinations, and flavor tagging, do not need to be run again due to preprocessing in the skimming stage.
- To estimate the amount of CPU power needed for analysis, we made assumptions on:
  - The average CPU power to run one analysis on one event
  - The number of concurrent analysis
  - The number of analysis cycles
- Uncertainties associated with analysis jobs are large.
  - We set them at the 50% level.
- The user n-tuples size scales with the number of concurrent analysis and of analysis cycles. We assume that running one cycle of one analysis on the skimmed data sample will produce on average 0.007 ± 0.004 kB/event of n-tuple data.

Analysis (2)

	2018	2019	2020	2021
CPU power to run one analysis (HEPSpec06 * s / ev)	0.04	0.04	0.04	0.04
Number of concurrent analysis	40	50	75	100
Number of analysis cycles	3.00	3.00	2.50	2.00
CPU power to run all analysis all cycles (HEPSpec06 * s / ev)	4.8 ± 2.4	6.0 ± 3.0	7.5 ± 3.8	8.0 ± 4.0
N-tuples size (kB/(ev*analysis*cycle)	0.007± 0.004	0.007± 0.004	0.007± 0.004	0.007± 0.004

Fraction of CPU used for analysis is 5 % of the total in 2018 and 2019, 10% in 2020 and 15% in 2021 to be compared with the 10% of LHCb

Activities in 2018 (1)

- Recreate the MC samples with nominal background requested by Physics Groups for 2017 to take advantage of the improvement of the software.
  - Corresponding data sample produced in 2017 will be kept to allow comparison of the code performances, while the data sample produced in 2016 will be deleted to save storage.
  - Two of the 4 ab^{-1} of generic MC will be produced during the MC production challenge (see below).
- Continue analysis of the 2.4×10^{10} Phase 2 (Beast) background events produced in 2017.
- Rehearsal of the processing and reprocessing of Phase 3 data processing using 1 ab^{-1} of raw data stored during the 5 ab^{-1} generic MC production.
- 2 ab^{-1} MC production challenge in parallel with the processing and reprocessing of the rehearsal of the first year of Phase 3 data processing
  - To test the computing system's capability of handling detector data processing and MC production running in parallel.

Activities in 2018 (2)

- Processing, simulation, and analysis of the 2018 cosmic run data.
- Reprocessing and re-simulation of the 2017 cosmic run data to take advantage of the improvement of the code. The raw data and the miniDST data from the 2017 processing are kept.
- Phase 2 data:
  - Process and reprocess the detector data. Due to the incomplete understanding of the detector, of the Data Acquisition and of the software we expect to reprocess the data at least once.
  - Produce, re-produce, and analyze a MC data sample that is 4 times the detector data sample.
  - Skim detector and MC Phase 2 data.
  - Analyze Phase 2 data.

Activities in 2018 (3)

- Phase 3 data:
  - Process, reprocess, and analyze the detector data. We expect to reprocess at least once the first Phase 3 data as our understanding of the Detector, of the DAQ and of the software improves.
  - Produce, re-produce, and analyze a MC data sample that is 4 times the detector data sample.
  - Skim detector and MC Phase 3 data.
  - Analyze the Phase 3 data.

Activities from 2019 to 2021

- Timely processing of the detector data.
- MC production.
  - At the beginning of the Phase 3 run, when the integrated luminosity is still low, it is important to have a MC sample that is many times the detector data sample.
  - This requirement will be relaxed in the following years of the data taking. The ratios of the MC event sample over the detector event sample is planned to be 3 in 2019, 2 in 2020 and 1 in 2021.
- Skimming of detector and MC data.
- Reprocessing of the detector data.
  - It will trigger the re-creation of MC event sample and the re-skimming of detector and MC data.
- Physics Analysis of data.

Tape (PB)

Year	2018	2019	2020	2021
Tape for Raw Data	0.4±0.1	5.6±1.4	17.4±4.3	35.7±9.1
Tape for Data Challenges	1.06±0.0	0.00	0.00	0.00
Tape for Cosmic Run	0.15	0.15	0.15	0.15
Total tape	**1.6±0.1**	**5.8±1.4**	**17.6±4.3**	**35.9±9.1**

Numbers in this table refers to storage capacity, it is immaterial if it actually is tape or disk

LHCb tape (2017): 59.7 PB

Disk (PB)

Year	2018	2019	2020	2021
Disk for Data Processing Buffer	0.01±0.00	0.06±0.00	0.23±0.00	0.37±0.00
Disk for data DST	0.01±0.00	0.11±0.00	0.12±0.00	0.10±0.00
Disk for data mDST	0.07±0.04	1.0±0.49	2.4±1.1	3.7±1.7
Disk for MC mDST	0.35±0.27	3.8±1.1	6.3±2.0	4.9±1.5
Disk for Data + MC microDST	0.72±0.49	8.2±2.6	14.9±5.1	14.7±5.0
Disk for user ntuplas	0.07±0.05	0.9±0.23	2.0±0.7	2.5±0.9
Disk for MC before Data Taking	2.0±0.9	0.00	0.00	0.00
Disk for Data Challenges	0.19±0.06	0.00	0.00	0.00
Disk for Cosmic	0.01±0.00	0.01±0.00	0.01±0.01	0.01±0.01
Total Disk	**3.5±1.3**	**14.1±3.1**	**25.9±5.5**	**26.3±5.6**

LHCb disk (2017): 28 PB

CPU (kHEPSpec06)

	2018	2019	2020	2021
CPU for data processing (kHEPSpec06)	2.9±0.6	23.7±4.9	52.9±10.9	86.1±17.9
CPU for data reprocessing (kHEPSpec06)	2.9±0.7	27.2±5.6	25.4±5.2	39.2±8.0
CPU for MC production (kHEPSpec06)	65.6±13.9	288.8±62.3	288.6±62.3	298.0±64.5
CPU for Skimming (kHEPSpec06)	4.8±2.7	33.4±19.1	42.4±24.2	47.1±26.9
CPU for analysis (kHEPSpec06)	6.5±3.2	26.9±13.8	64.2±32.6	93.2±47.1
CPU for MC Before Data Taking (kHEPSpec06)	63.1±18.5	0.00	0.00	0.00
CPU for Data Challenges (kHEPSpec06)	35.3±15.0	0.00	0.00	0.00
CPU for Cosmic (kHEPSpec06)	6.0±3.0	0.00	0.00	0.00
Total CPU (kHEPSpec06)	**187±27**	**400±67**	**473±75**	**563±86**

LHCb CPU(2017): 375 kHEPSpec