

Belle II Computing Update

- 1. Experiment status*
- 2. Belle II Computing*

*May 6, 2020
@ 2020 May GDB
Takanori HARA (KEK IPNS)*

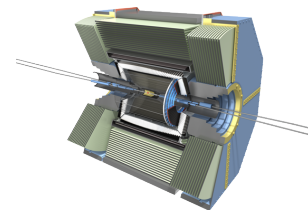
Belle II Computing Update

- 1. Experiment status*
- 2. Belle II Computing*

*May 6, 2020
@ 2020 May GDB
Takanori HARA (KEK IPNS)*

Motivation from Physics

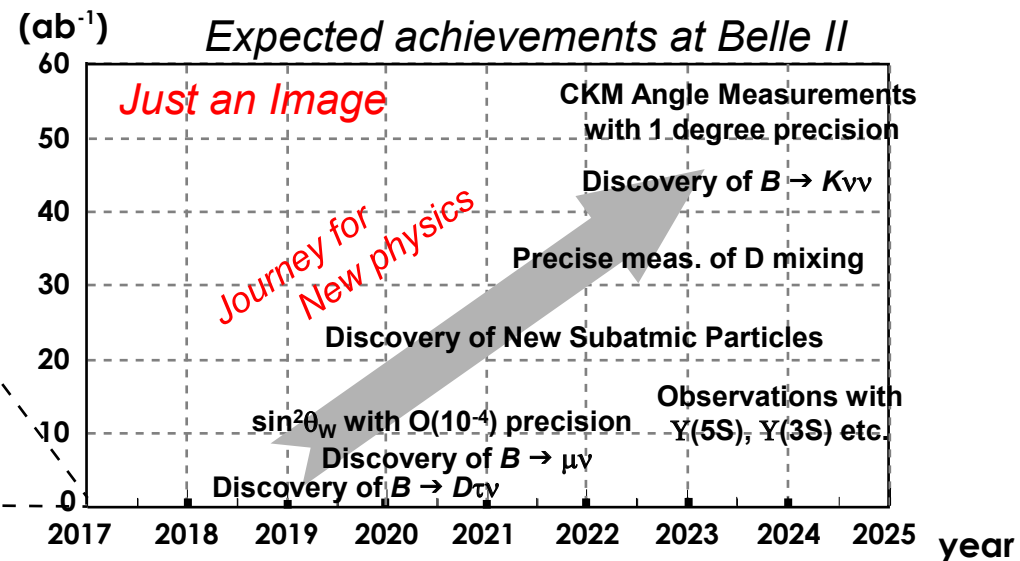
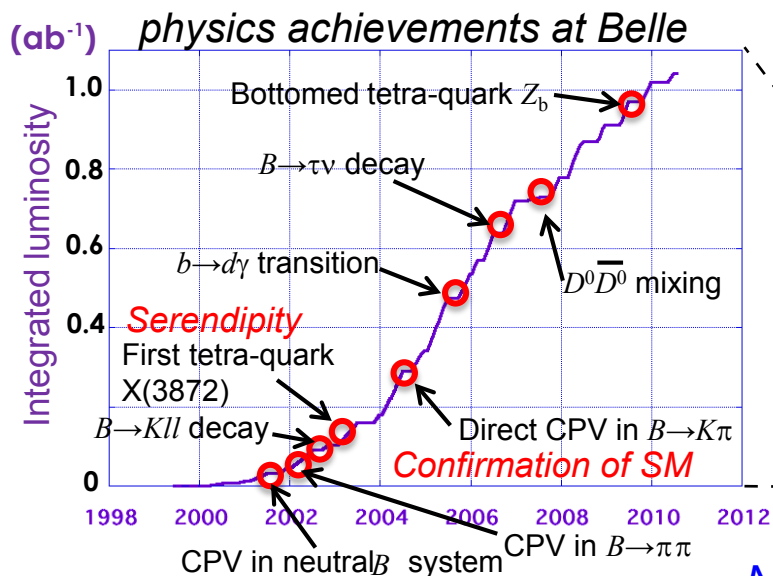
Accelerator	KEKB	SuperKEKB
Beam Energy (GeV)	3.5 x 8 (g = 0.425)	4 x 7 (g = 0.28)
CM energy, Y(4S),, Y(4S),
Luminosity ($\text{cm}^{-2} \text{s}^{-1}$)	2.1×10^{34}	8×10^{35}
Total data (ab^{-1})	1	50
	raw data : ~1PB mDST data/MC : 0.14/0.6 PB (for one version)	raw data : ~50PB (in total ~200PB including physics datasets and replicas)



x 40
x 50
Higher intensity

Centralized Computing one big center @ KEK (non-grid)

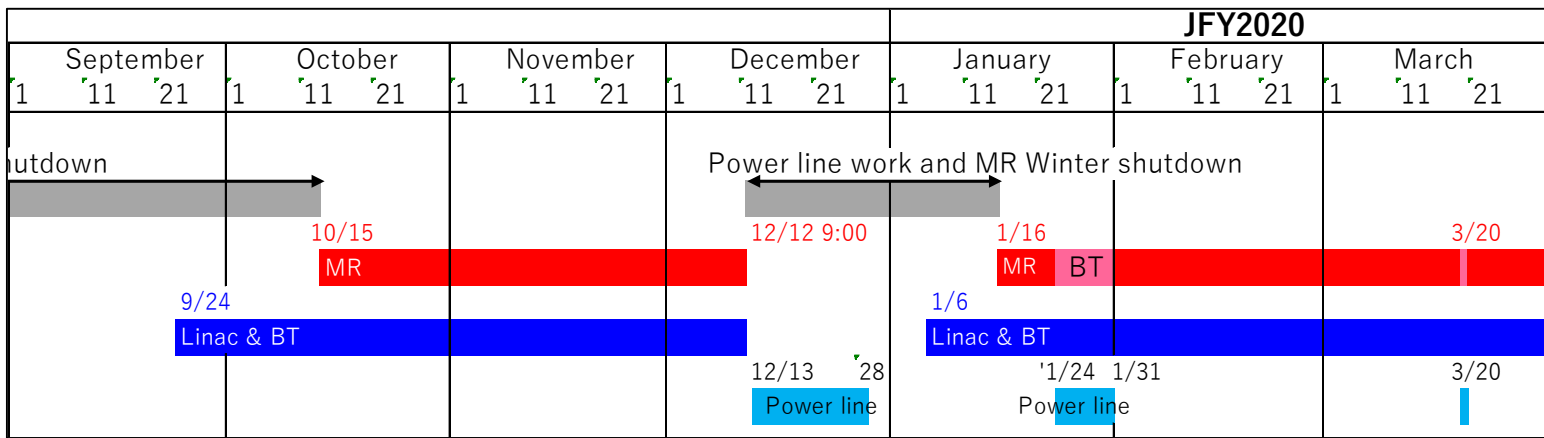
World-wide Distriuted Computing



Need large-scale computing resources and effective operation

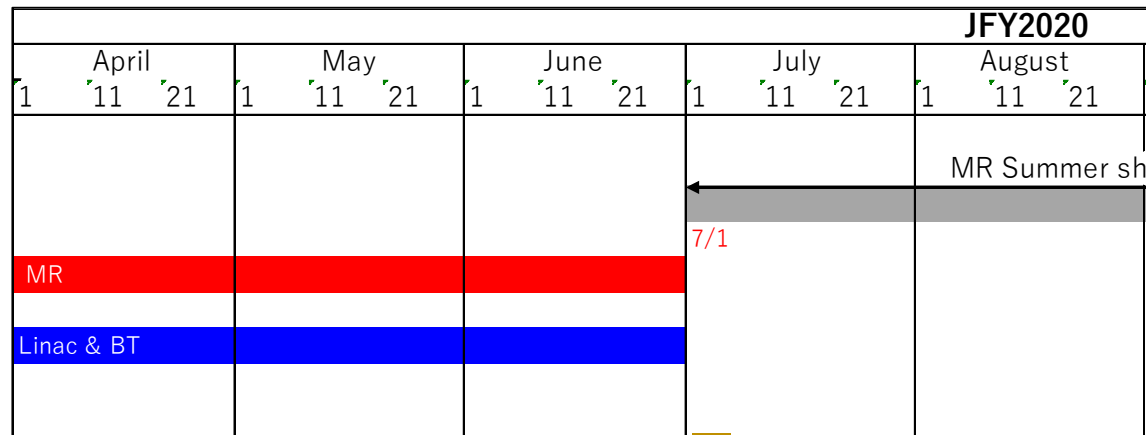
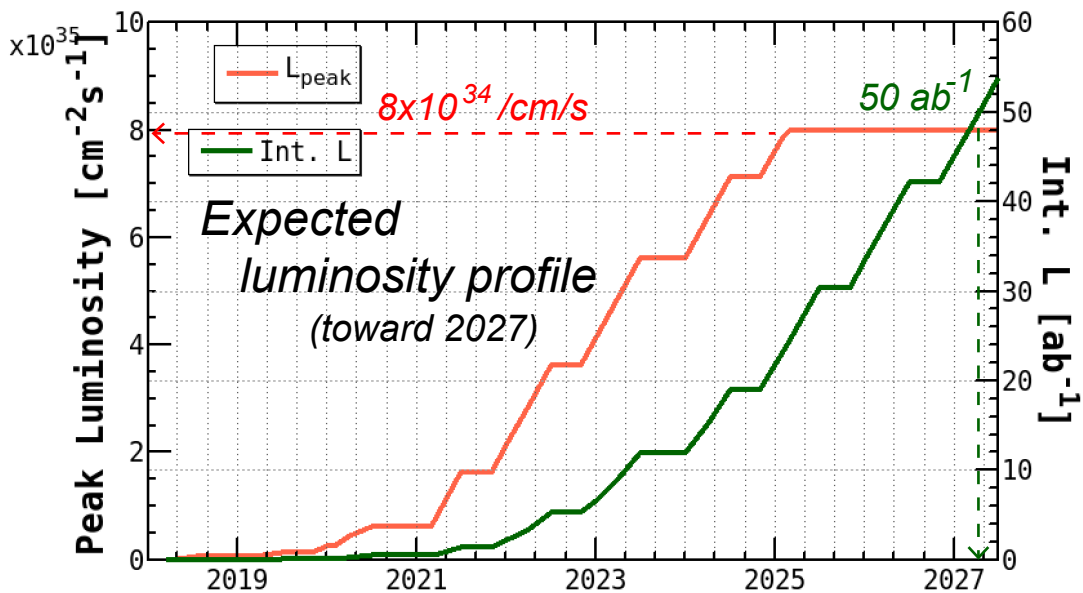
Belle II Experiment : run plan (until 2020 Summer)

(and toward 2027)



Luminosity goal
 Integrated : 200 fb^{-1}
 peak : $2\sim 3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$

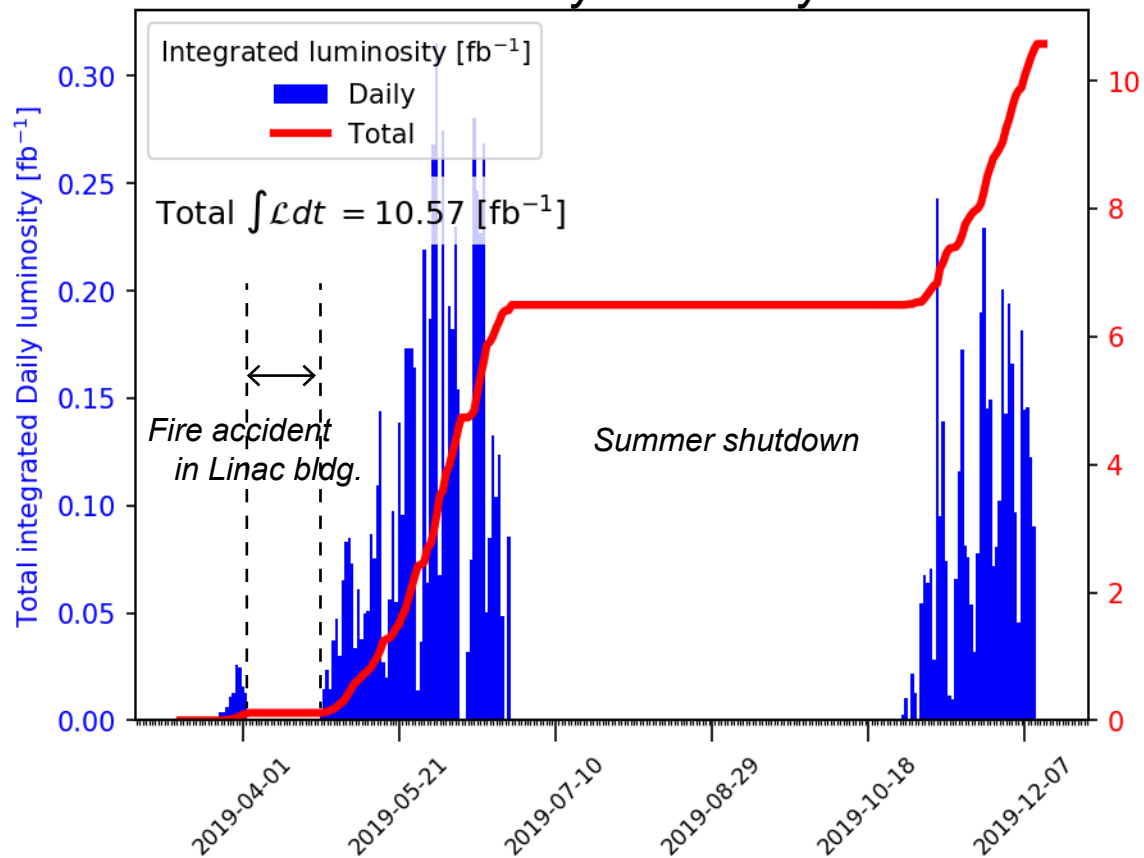
Current status (May 4)
 Integrated : $\sim 35 \text{ fb}^{-1}$
 peak : $\sim 1.8 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$
 (not official values)



Even under restricted situation because of COVID19, SuperKEKB accelerator, Belle II detector and Belle II Distributed Computing are being operated as usual

Run operation status

Belle II Online daily luminosity in 2019



c.f. Belle $Y(4S)$: 711 fb^{-1}

Plot on 2020/01/28 15:24

First paper is being published

<https://arxiv.org/abs/1910.05365>

arXiv.org > hep-ex > arXiv:1910.05365

Search...

Help | Advan

High Energy Physics - Experiment

Measurement of the integrated luminosity of the Phase 2 data of the Belle II experiment

Belle II Collaboration: F. Abudinén, I. Adachi, P. Ahlburg, H. Aihara, N. Akopov, A. Aloisio, L. Andricek, N. Anh Ky, D. M. Asner, H. Atmacan, T. Aushev, V. Aushev, K. Azmi, V. Babu, S. Baehr, S. Bahinipati, A. M. Bakich, P. Bambade, Sw. Banerjee, S. Bansal, V. Bansal, M. Barrett, J. Baudot, A. Beaulieu, J. Becker, P. K. Behera, J. V. Bennett, E. Bernieri, F. U. Bernlochner, M. Bertemes, M. Bessner, S. Bettarini, V. Bhardwaj, F. Bianchi, T. Bilka, S. Bilokin, D. Biswas, G. Bonvicini, A. Bozek, M. Bračko, P. Branchini, N. Braun, T. E. Browder, A. Budano, S. Bussino, M. Campajola, L. Cao, G. Casarosa, C. Cecchi, D. Červenkov, M.-C. Chang, P. Chang, R. Cheaib, V. Chekelian, Y. Q. Chen, Y.-T. Chen, B. G. Cheon, K. Chilkin, H.-E. Cho, K. Cho, S. Choudhury, D. Cinabro, L. Corona, L. M. Cremaldi, S. Cunliffe, T. Czank, F. Dattola, E. De La Cruz-Burelo, G. De Nardo, M. De Nuccio, G. De Pietro, R. de Sangro, M. Destefanis, S. Dey, A. De Yta-Hernandez, F. Di Capua, S. Di Carlo, J. Dingfelder, Z. Doležal, I. Domínguez Jiménez, T. V. Dong, K. Dort, S. Dubey, S. Duell, S. Eidelman, M. Eliachevitch, T. Ferber, D. Ferlewicz, G. Finocchiaro, S. Fiore, A. Fodor, F. Forti, A. Frey, B. G. Fulsom, M. Gabriel, E. Ganiev, M. Garcia-Hernandez, A. Garmash, V. Gaur et al. (299 additional authors not shown)

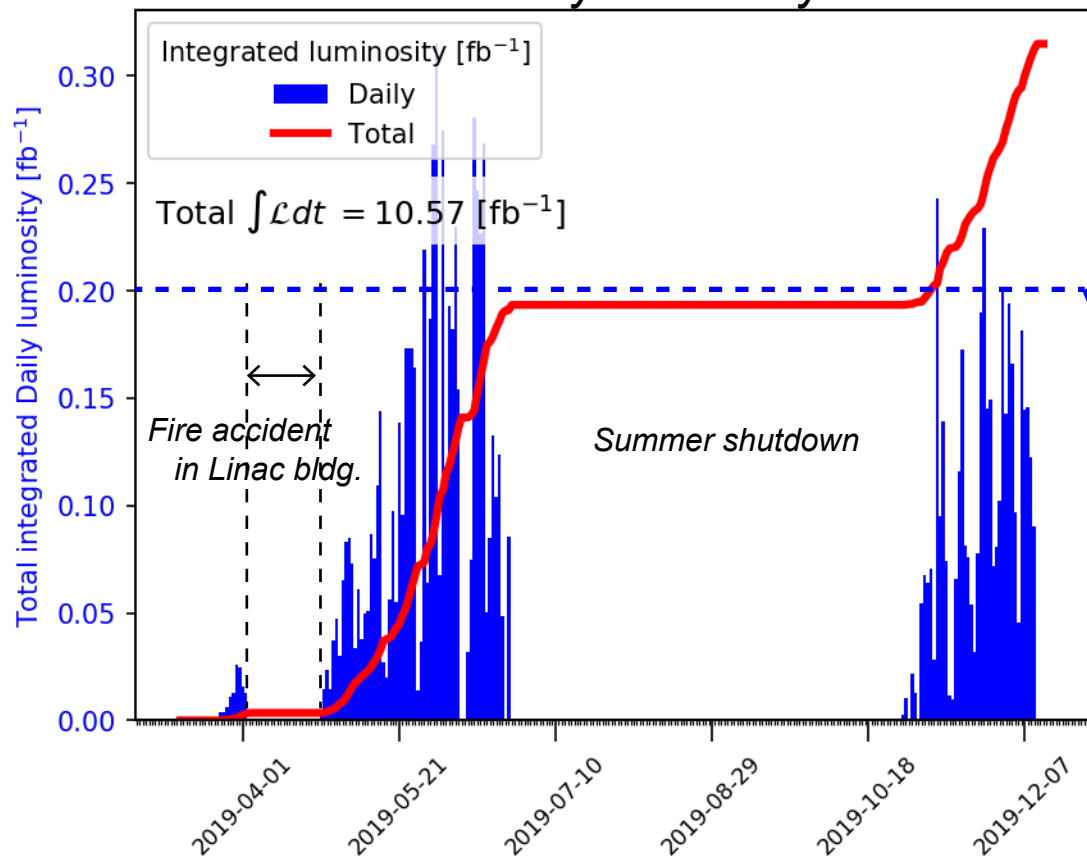
(Submitted on 11 Oct 2019)

From April to July 2018, a data sample at the peak energy of the $\Upsilon(4S)$ resonance was collected with the Belle-II detector at the SuperKEKB electron-positron collider. This is the first data sample of the Belle-II experiment. Using Bhabha scattering events, we measure the integrated luminosity of the data sample to be $(496.7 \pm 0.3 \pm 3.5) \text{ pb}^{-1}$, where the first uncertainty is statistical and the second is systematic. A measurement with digamma events is performed as a cross-check, and the obtained result is in agreement with the nominal result. This work provides a basis for future luminosity measurements at Belle-II.

more physics analysis papers will come...

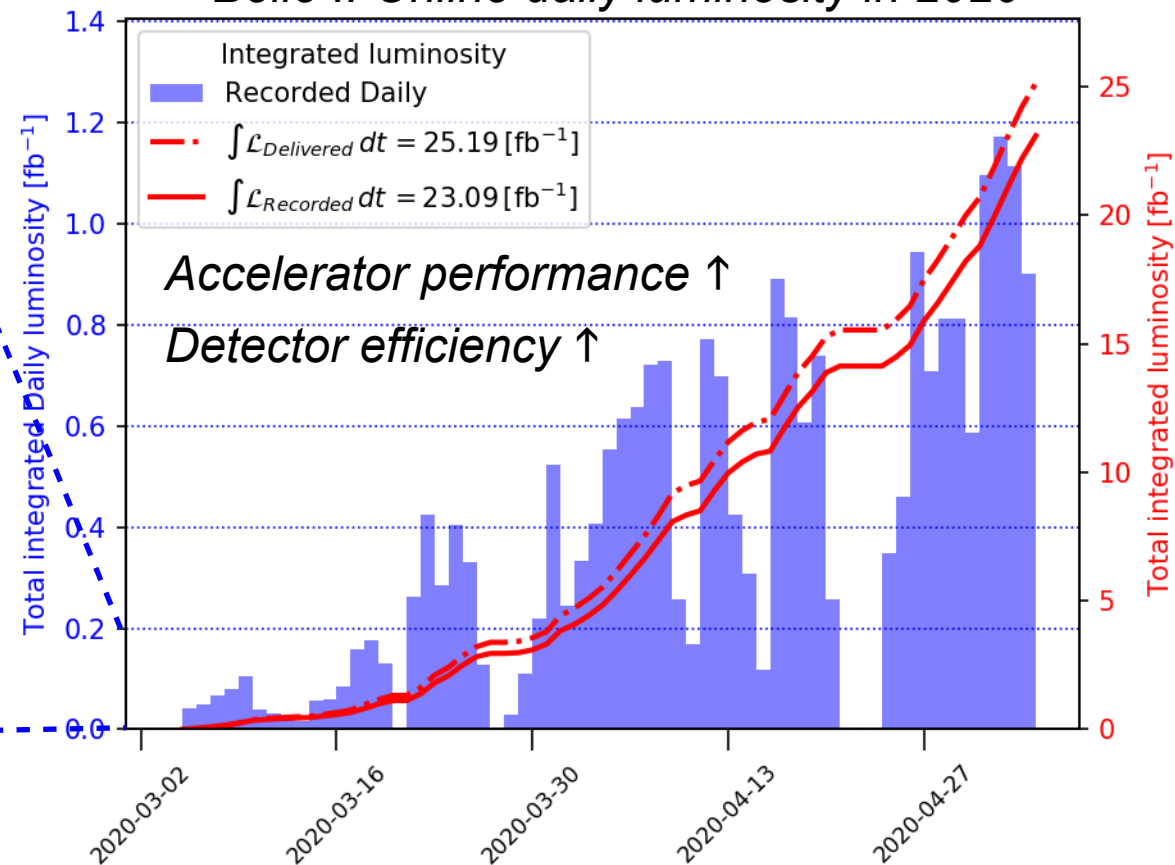
Run operation status

Belle II Online daily luminosity in 2019



c.f. Belle $Y(4S) : 711 \text{ fb}^{-1}$

Belle II Online daily luminosity in 2020



Belle II Computing Update

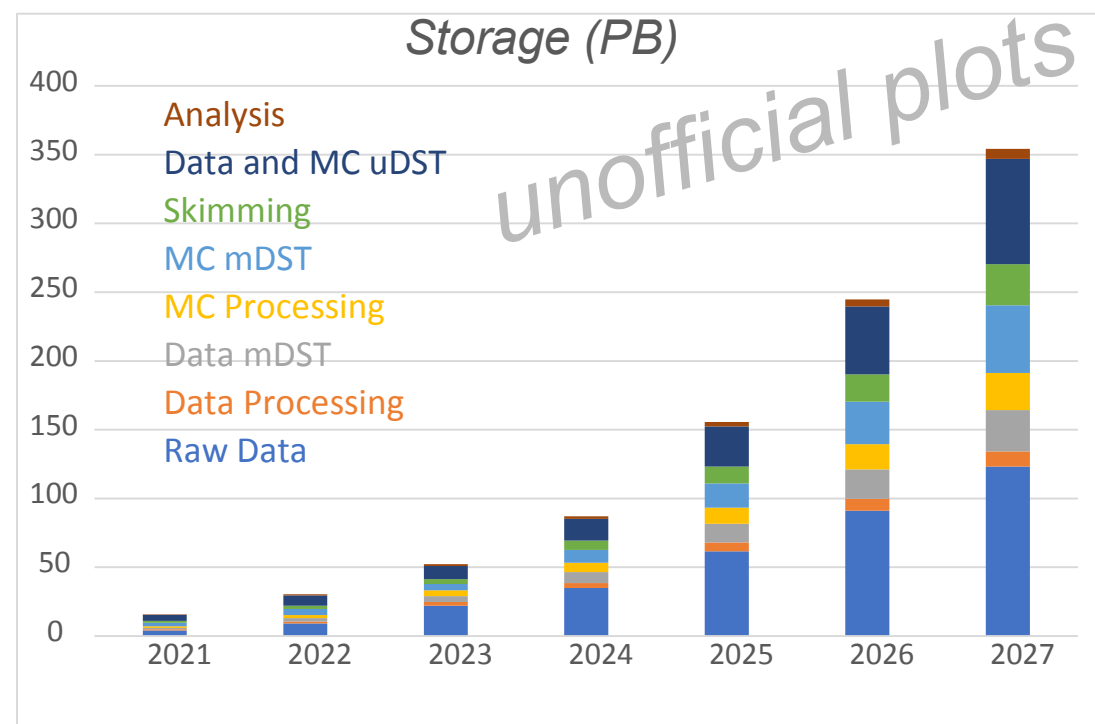
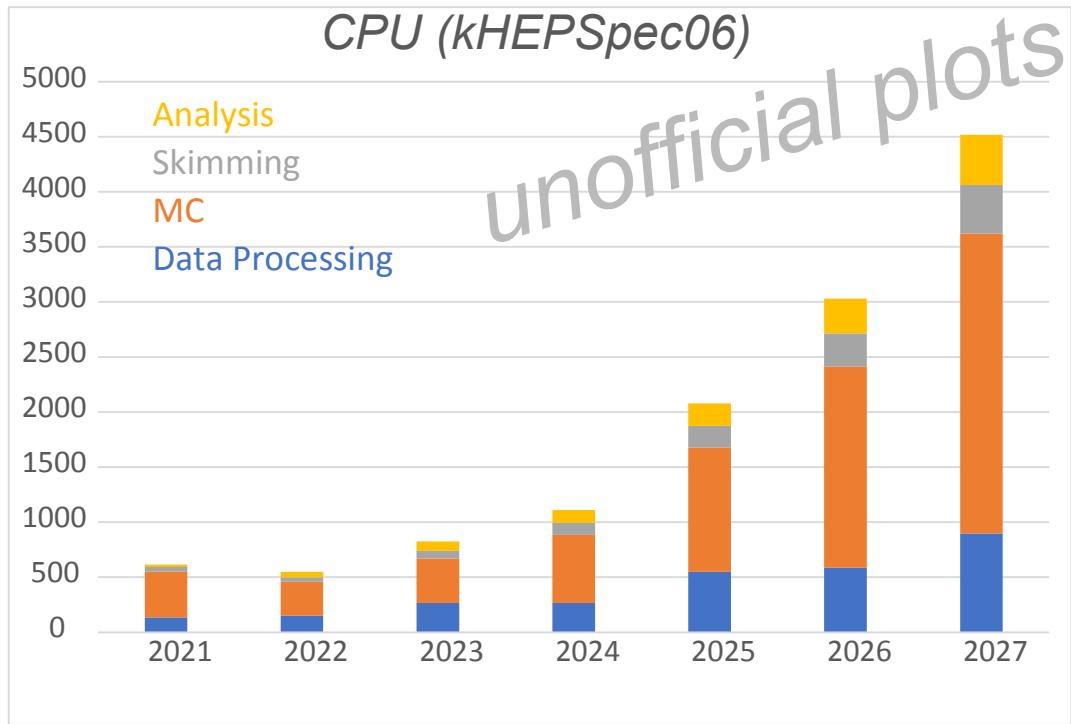
- 1. Experiment status*
- 2. Belle II Computing*

*May 6, 2020
@ 2020 May GDB
Takanori HARA (KEK IPNS)*

Computing resource requirements

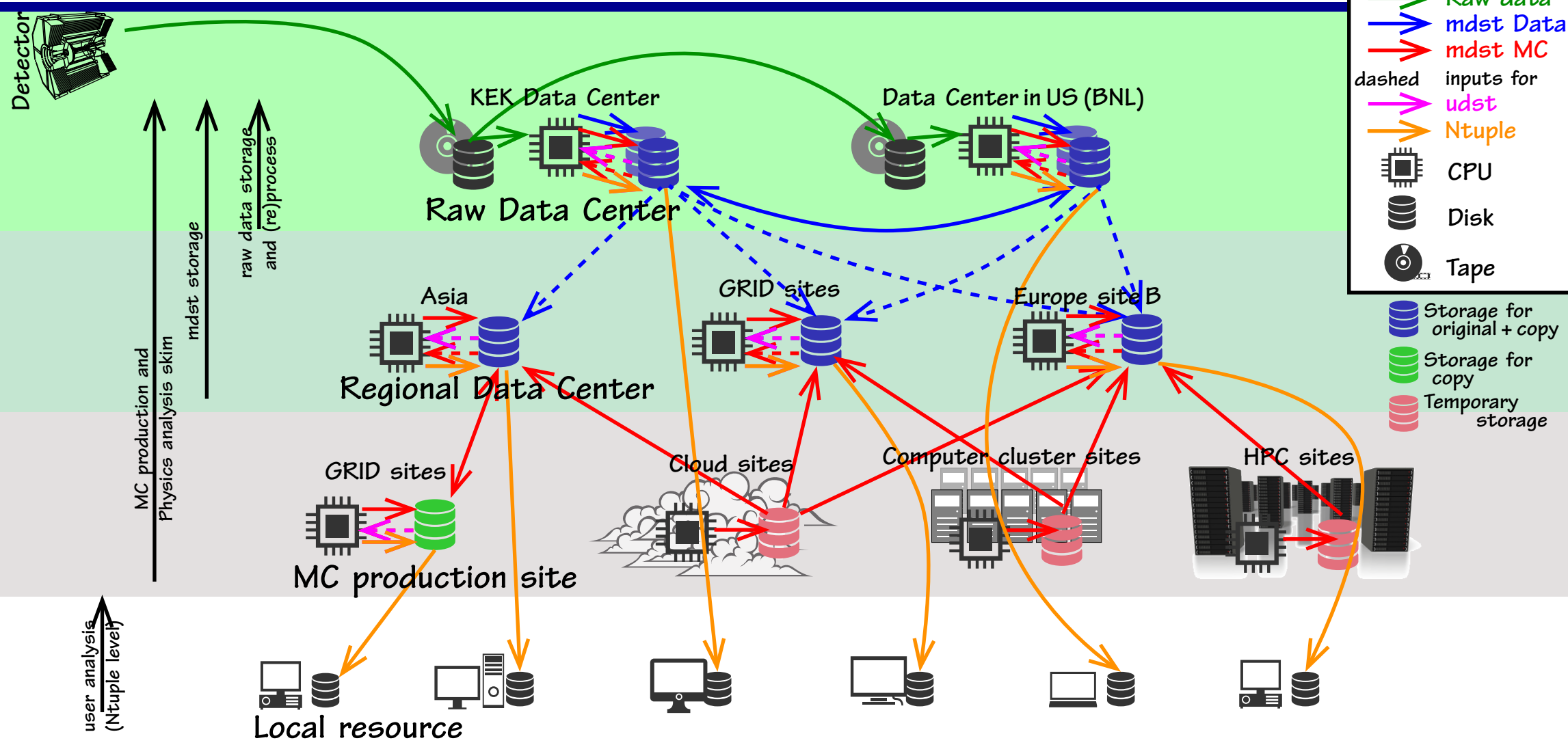
Resource estimation is being realistic thanks to the matured software and run operation experience

Resource estimation for 2021CY (and succeeding three years) will be scrutinized by an external committee soon



*Need to increase resources more by one order of magnitude than the current level
with flat budget...*

Belle II Computing Model for first three years



Belle II Distributed Computing Structure

Human

Software interface
+ Interware extension
+ Analysis user interface

BelleDIRAC

Production Manager

Data Manager

End Users

Interware
+ management system

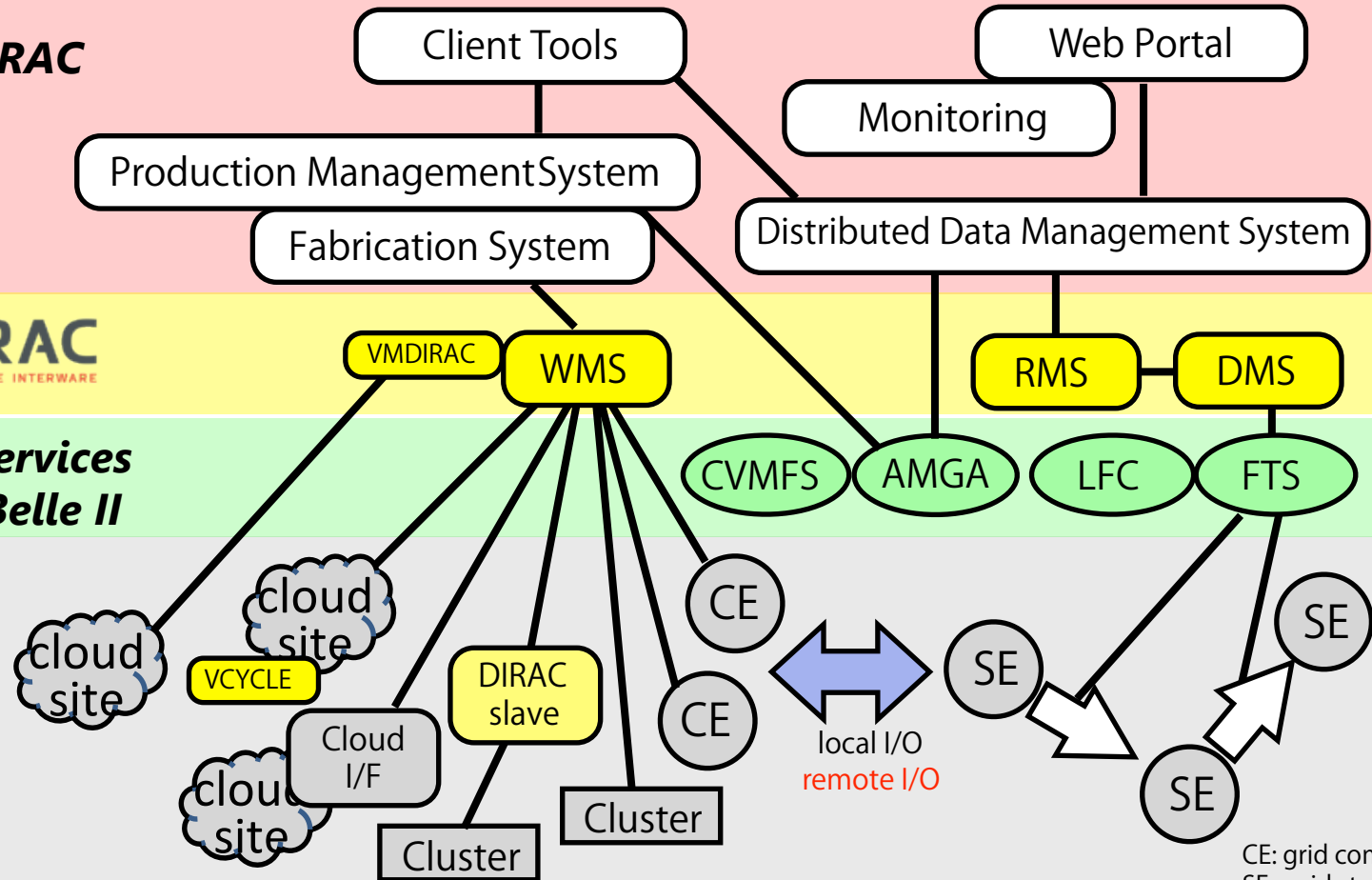


Cyberinfrastructure
+ Services

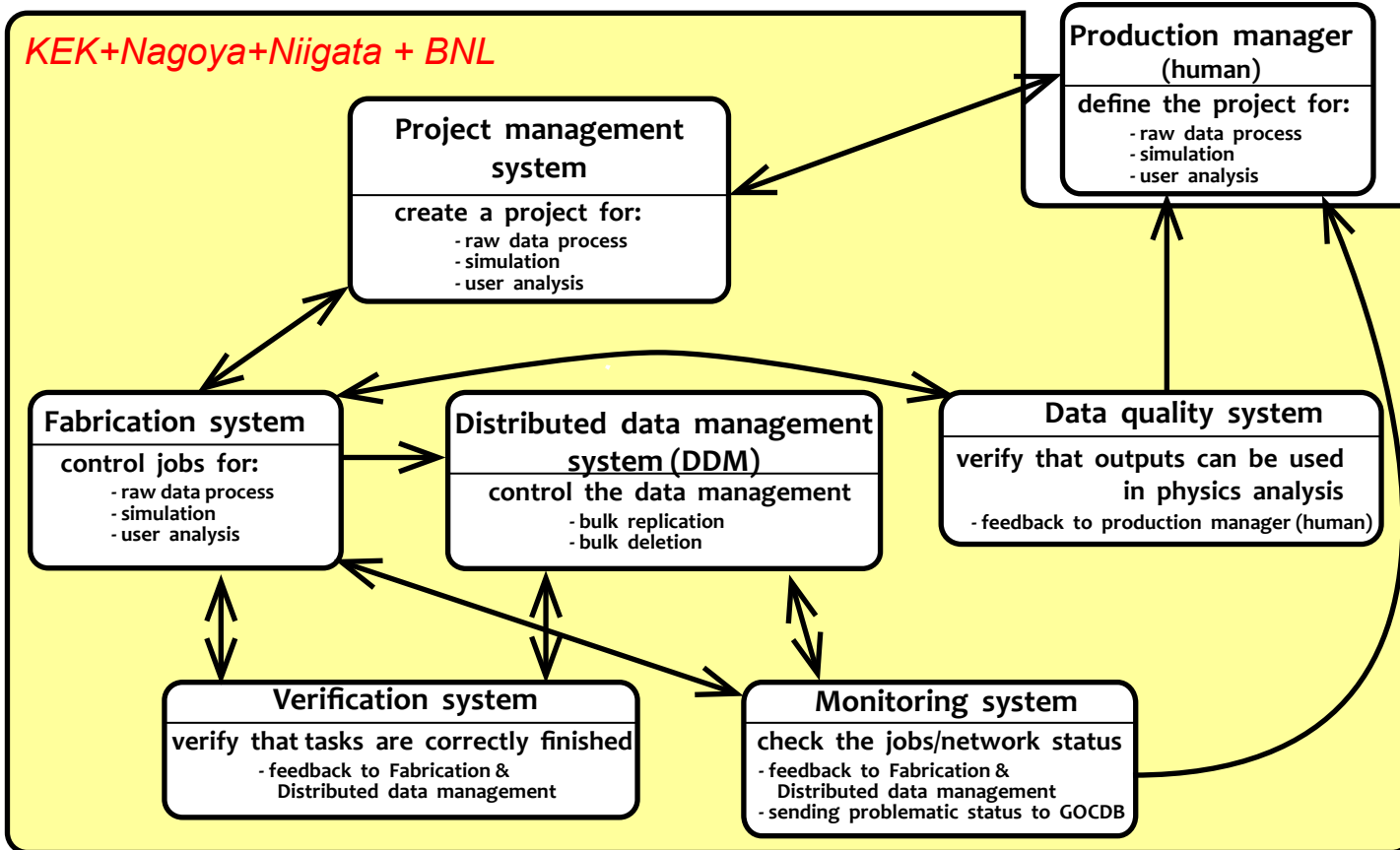
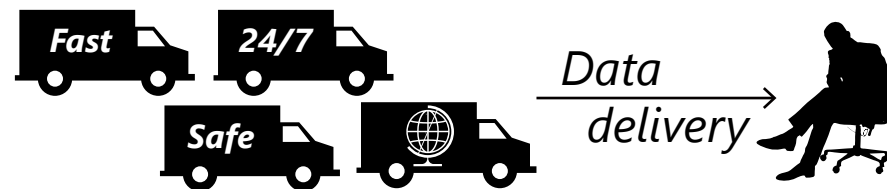
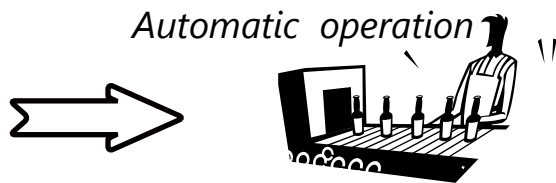
**GRID services
for Belle II**

Platform
+ GRID Middleware
+ OS
+ Hardware
+ Network } Infrastructure

Sites



Automatized Production System



Different types of production

MC production (w/ or w/o BG)

Skim production

RAW data process

Huge variety of modes

BB, udsc, signal, background

many physics skims

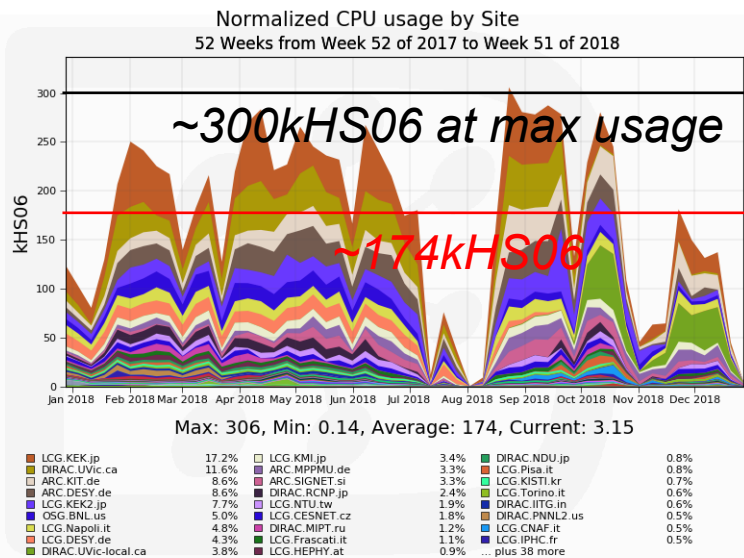
Complicated data management over world-distributed sites

Reduce human error and perform effective operation

CPU usage

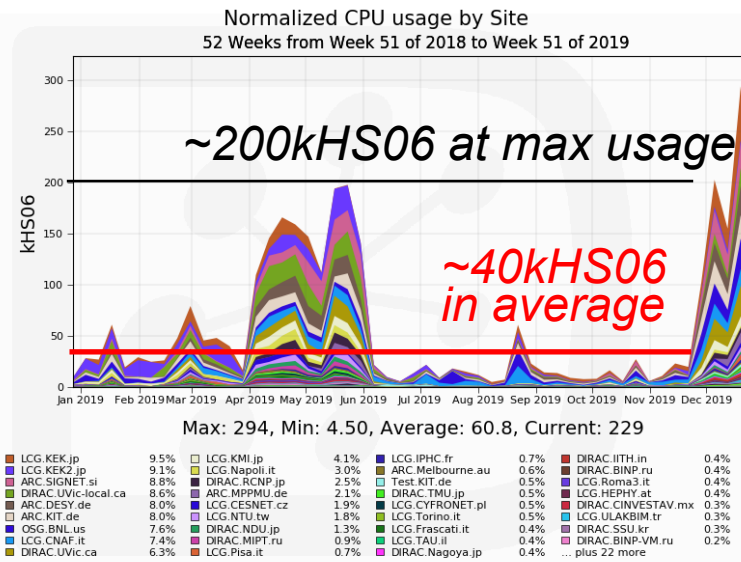
Computing sites in the world provide CPU and Storage for Belle II ...

Belle II has not utilized the provided CPU resources in 2019



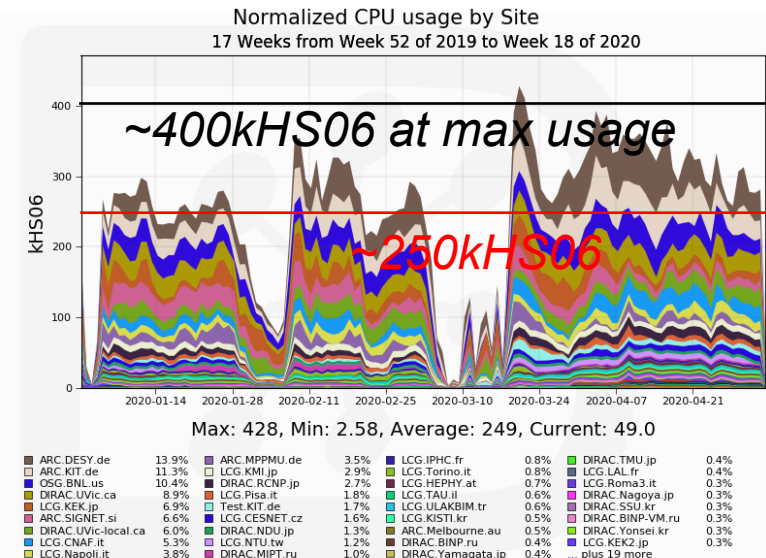
2018 CY

Generated on 2019-10-20 05:36:55 UTC



2019 CY

Generated on 2020-05-04 17:20:50 UTC



2020 CY (up to May)

Generated on 2020-05-04 17:19:45 UTC

~25% CPU usage of 2018
and not used continuously

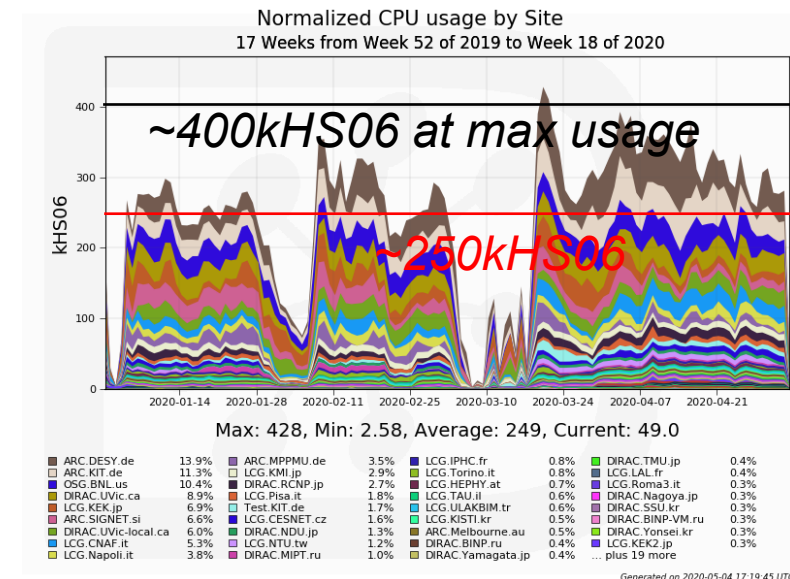
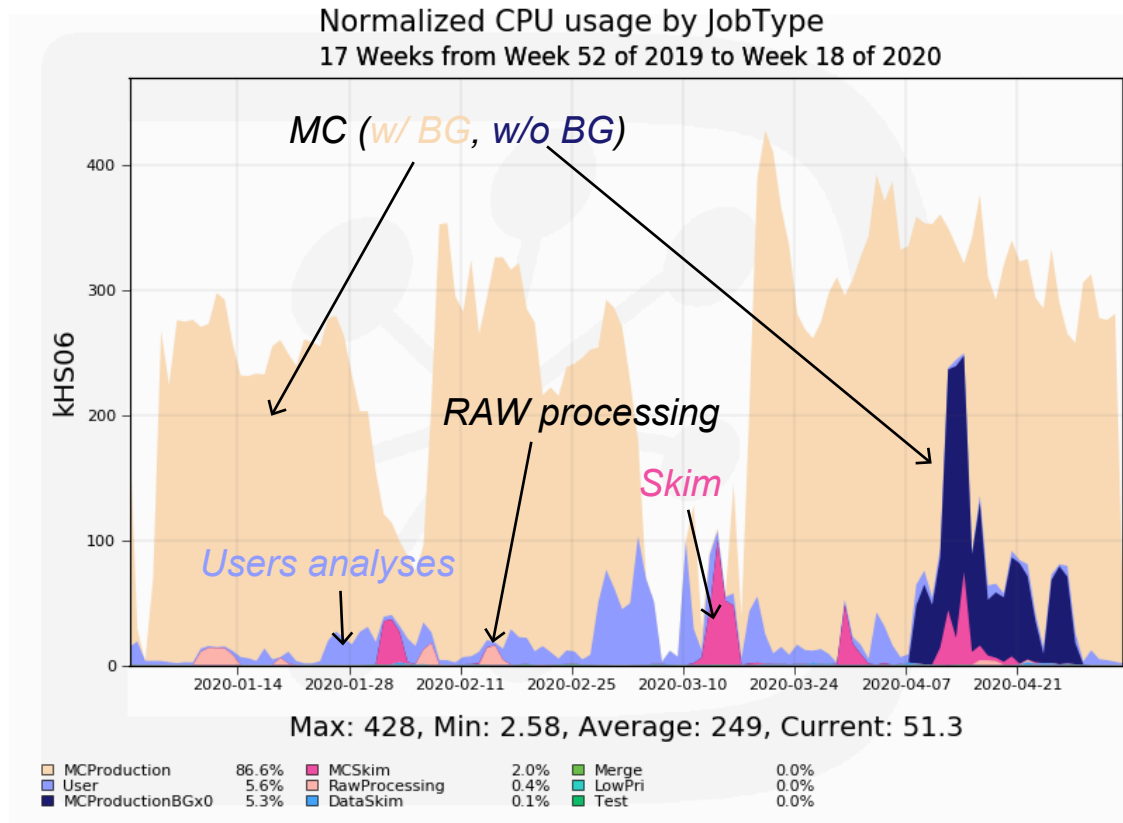
Luminosity was not high as expected
Software was not ready in timely way

→ will be improved for the next year

CPU usage

Computing sites in the world provide CPU and Storage for Belle II ...

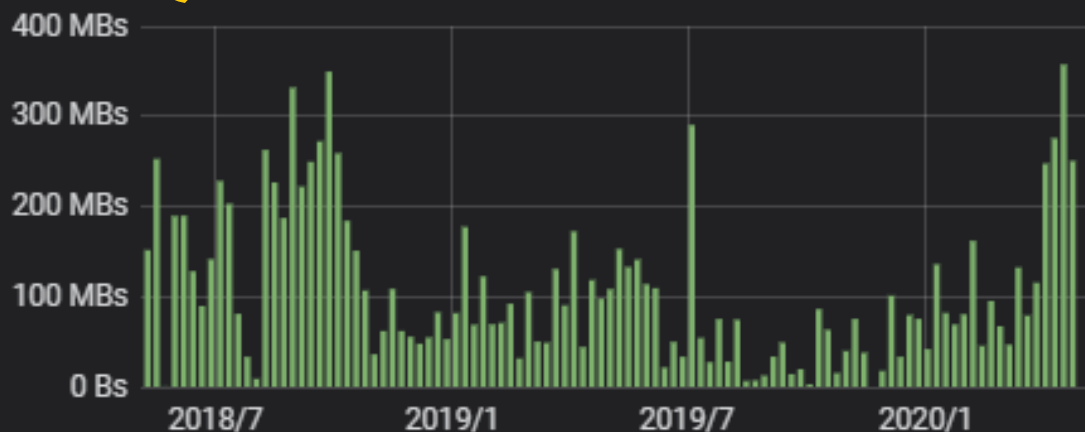
Belle II has not utilized the provided CPU resources in 2019



Data transfer

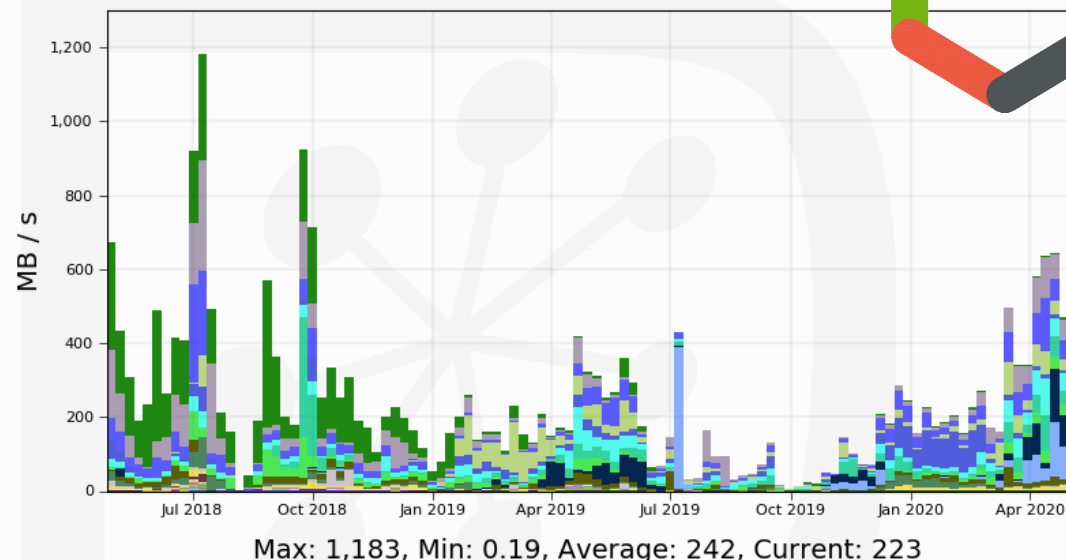


Transfer Throughput



https://monit-grafana.cern.ch/d/000000421/fts-transfers-5-years?orgId=20&from=now-2y&to=now&var-group_by=vo&var-vo=belle&var-src_country=All&var-dst_country=All&var-src_site=All&var-dst_site=All&var-fts_server=All&var-bin=7d&var-rp=five_years_1w

Throughput by Destination 105 Weeks from Week 16 of 2018 to Week 17 of 2020



KEK2-TMP-SE	20.7%	BNL-TAPE-SE	4.8%	KISTI-TMP-SE	0.3%	CYFRONET-TMP-SE	0.1%
BNL-TMP-SE	13.6%	KEK-RAW-SE	4.2%	Frascati-TMP-SE	0.3%	NTUCC-TMP-SE	0.0%
DESY-TMP-SE	11.6%	UVic-TMP-SE	4.0%	Torino-TMP-SE	0.1%	Adelaide-TMP-SE	0.0%
KEK-DISK-TMP-SE	7.8%	KMI-TMP-SE	2.9%	Roma3-TMP-SE	0.1%	Melbourne-TMP-SE	0.0%
SINET-TMP-SE	7.4%	CESNET-TMP-SE	1.0%	TAU-TMP-SE	0.1%	LAL-TMP-SE	0.0%
KIT-TMP-SE	7.3%	MPPMU-TMP-SE	0.8%	IPHC-TMP-SE	0.1%	CINVESTAV-TMP-SE	0.0%
CNAF-TMP-SE	6.9%	Pisa-TMP-SE	0.4%	HEPHY-TMP-SE	0.1%	KEK-TMP-SE	0.0%
Napoli-TMP-SE	4.9%	NTU-TMP-SE	0.3%	ULAKBIM-TMP-SE	0.1%	PNNL-TMP-SE	0.0%

Generated on 2020-05-04 15:19:42 UTC

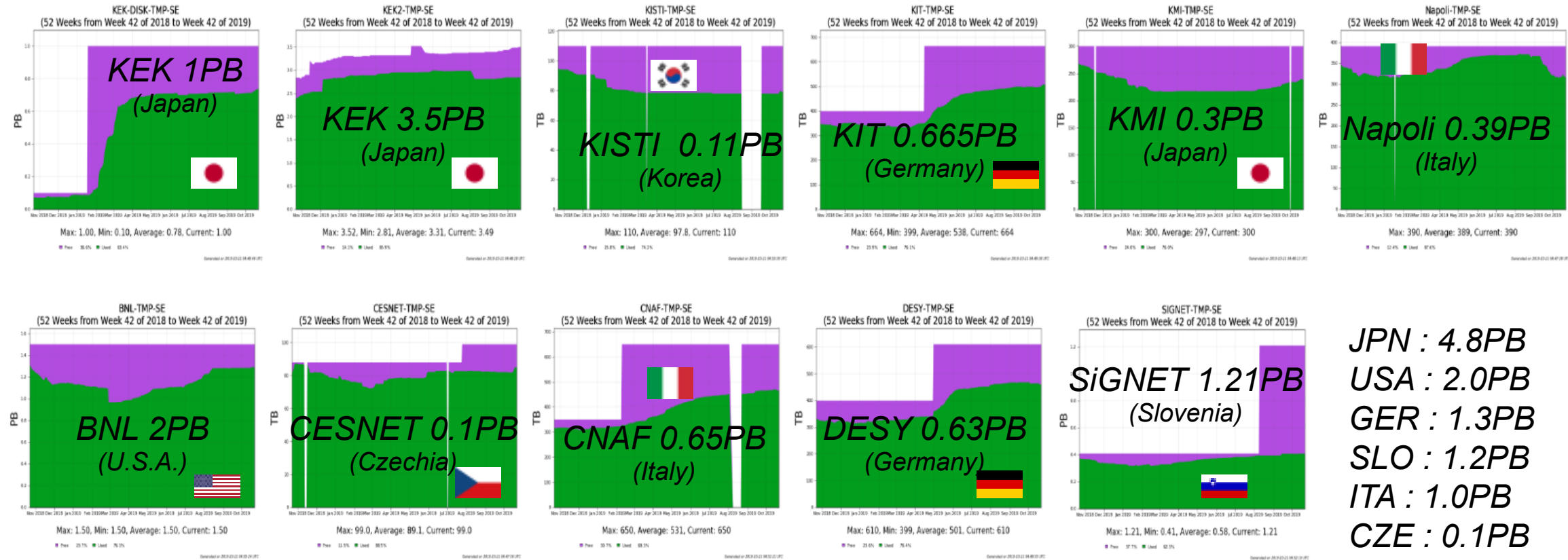
Produced MC / data files are collected in “Primary SEs”
10 primary SEs (Asia:3, US:1, Europe:6) among ~30 SEs

Replica creation : Automated DDM with replica policy

Deletion : still in the manual operation

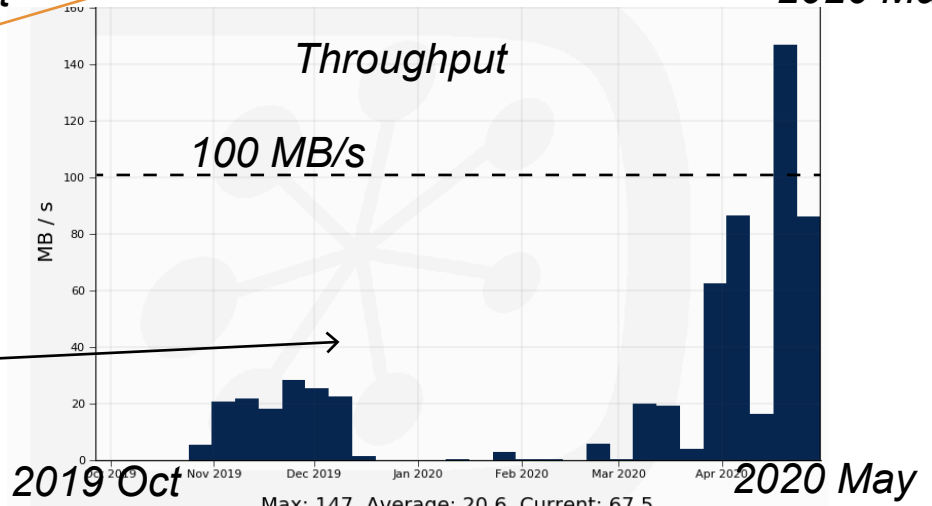
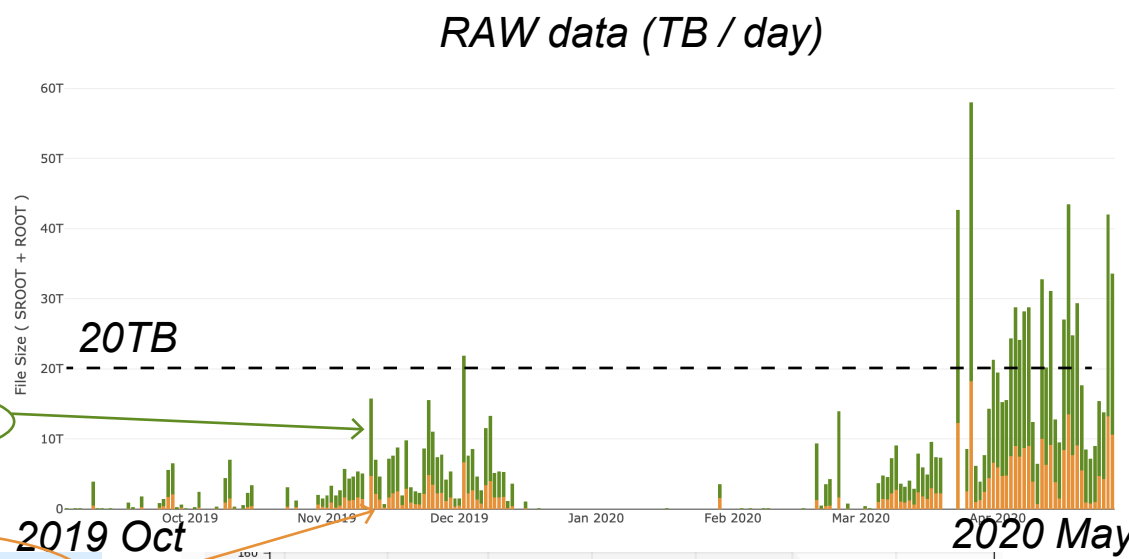
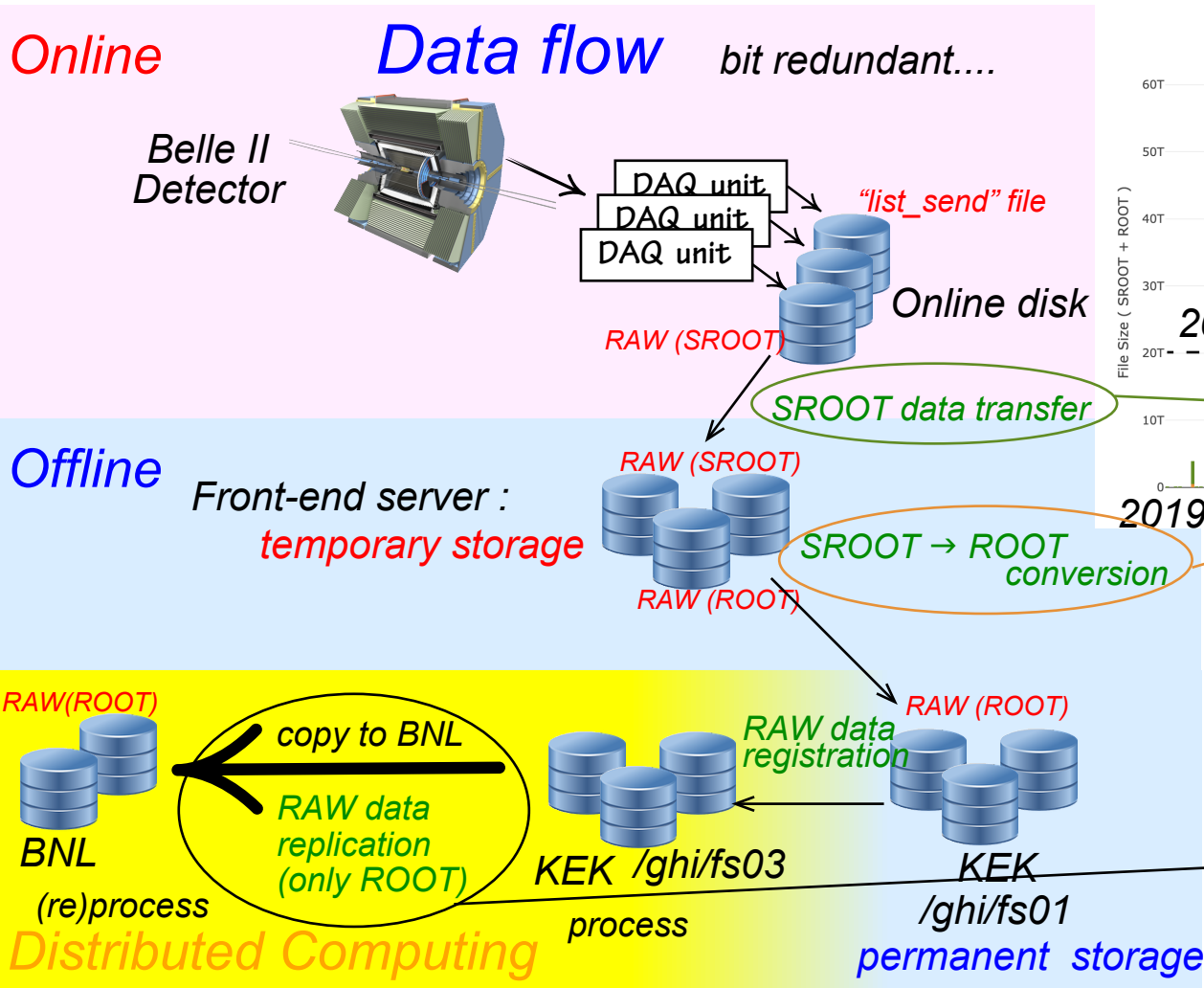
→ Improve BelleDIRAC DDM
→ implement Rucio in DIRAC

Primary Storage providers

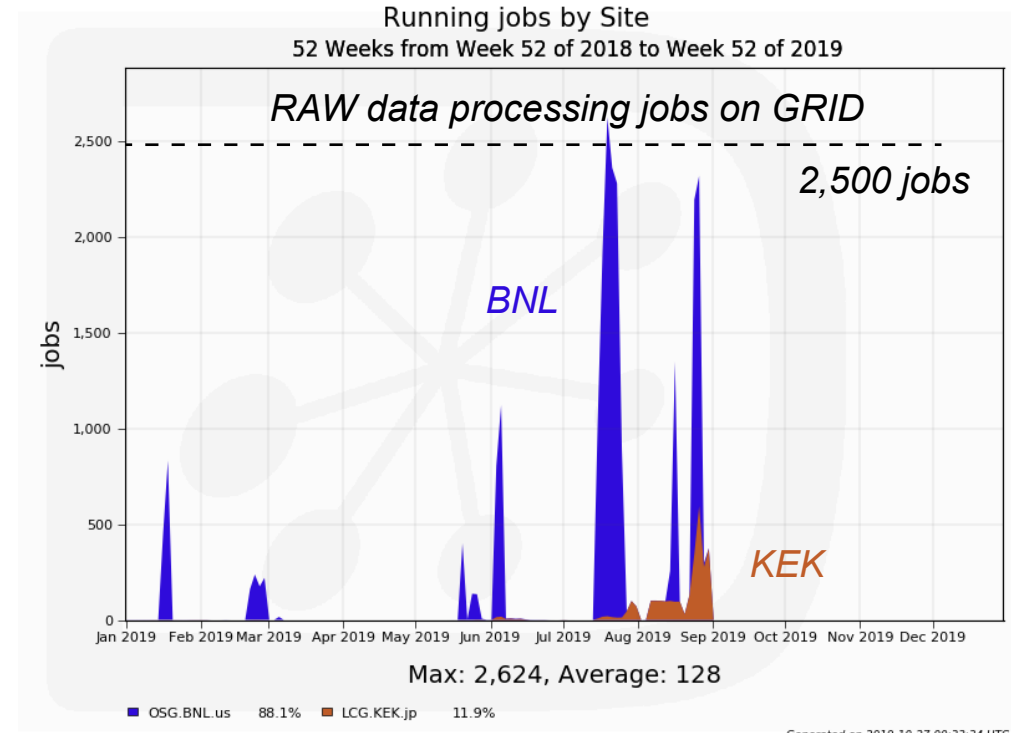
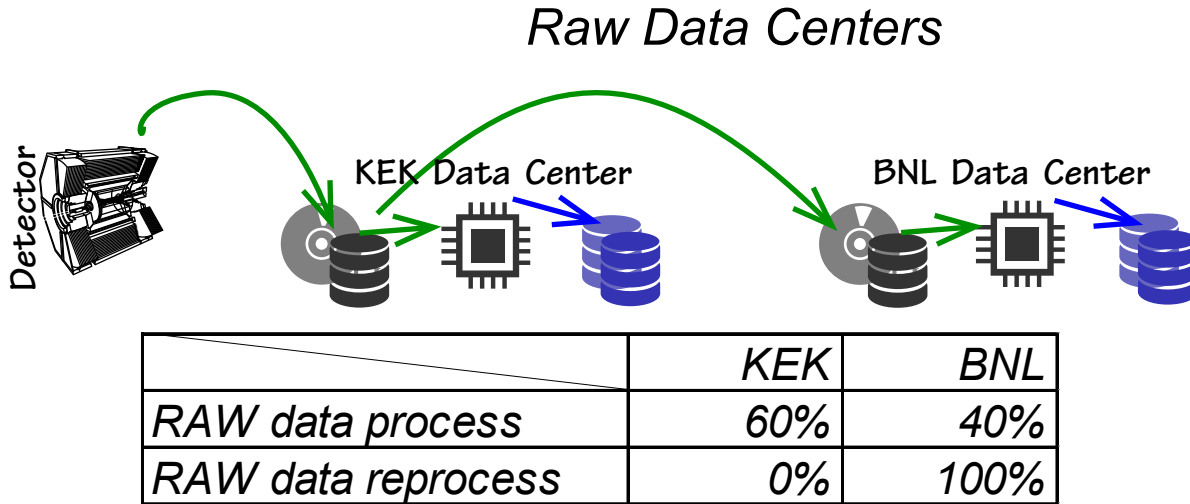


JPN : 4.8PB
USA : 2.0PB
GER : 1.3PB
SLO : 1.2PB
ITA : 1.0PB
CZE : 0.1PB
KOR : 0.1PB

Automatic RAW data copy from Online → Grid



RAW data (re)process



We artificially throttled RAW data (re)process jobs at KEK to confirm if RAW data reprocess jobs can run at BNL

*MC production (mDST)
MC / Data skim production for physics analysis (uDST)* } *are being done on Grid
with production system*

User jobs

Data searcher through DIRAC web portal

David DOSSETT (Melbourne)

Dataset Searcher

Dataset Searcher

Metadata Searcher Tree Browser

Data Type: MC Data

Campaigns: MC13a Beam Energies: 45

Skim Types: Data Levels:

Releases: Global Tags:

Experiment Low: Experiment High:

Run Low: Run High:

MC Event Types: General Skim Names:

Clear Search Help

LPN

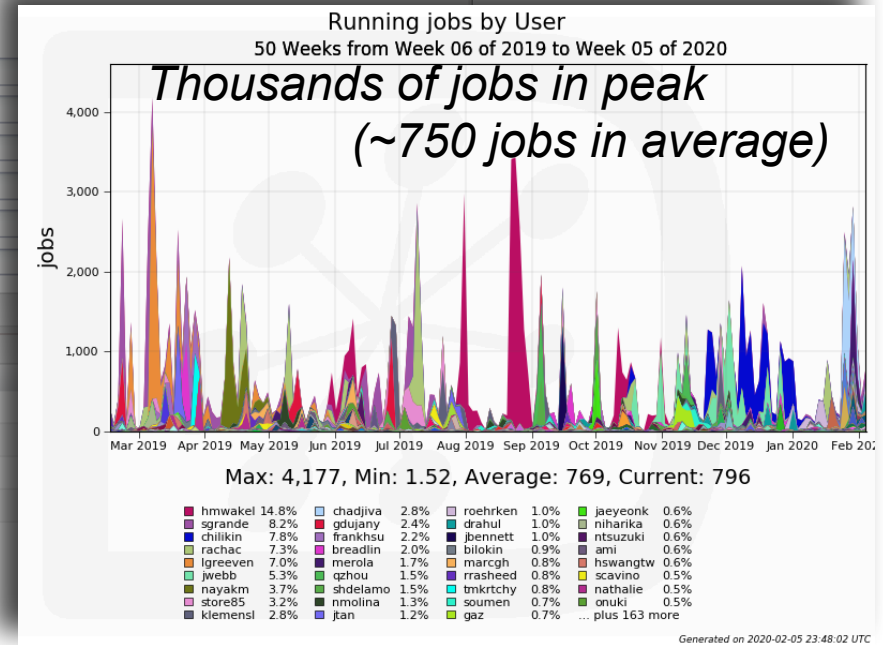
/belle/MC/release-04-00-03/DB00000757/MC13a/prod00009434/s00/e1003/45/r00000/mixed/mdst

/belle/MC/release-04-00-03/DB00000757/MC13a/prod00009435/s00/e1003/45/r00000/charged/mdst

/belle/MC/release-04-00-03/DB00000757/MC13a/prod00009436/s00/e1003/45/r00000/uubar/mdst

/belle/MC/release-04-00-03/DB00000757/MC13a/prod00009437/s00/e1003/45/r00000/ccbar/mdst

/belle/MC/release-04-00-03/DB00000757/MC13a/prod00009438/s00/e1003/45/r00000/ddbar/mdst



User jobs

Data searcher through DIRAC web portal

David DOSSETT (Melbourne)

Dataset Searcher

Dataset Searcher

Metadata Searcher Tree Browser

Data Type: MC Data

Campaigns: MC13a Beam Energies: 45

Skim Types: Data Levels:

Releases: Global Tags:

Experiment Low: Experiment High:

Run Low: Run High:

MC Event Types: General Skim Names:

Clear Search Help

LPN

/belle/MC/release-04-00-03/DB00000757/MC13a/prod00009434/s00/e1003/45/r00000/mixed/mdst

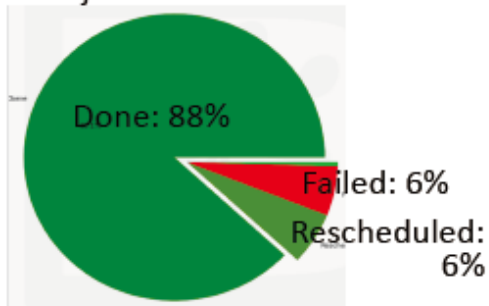
/belle/MC/release-04-00-03/DB00000757/MC13a/prod00009435/s00/e1003/45/r00000/charged/mdst

/belle/MC/release-04-00-03/DB00000757/MC13a/prod00009436/s00/e1003/45/r00000/uubar/mdst

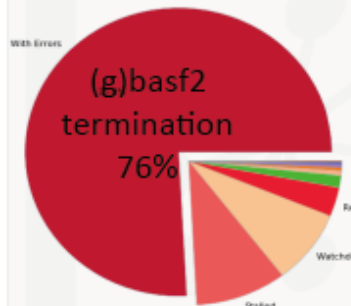
/belle/MC/release-04-00-03/DB00000757/MC13a/prod00009437/s00/e1003/45/r00000/ccbar/mdst

/belle/MC/release-04-00-03/DB00000757/MC13a/prod00009438/s00/e1003/45/r00000/ddbar/mdst

Final status of executed jobs in 2019

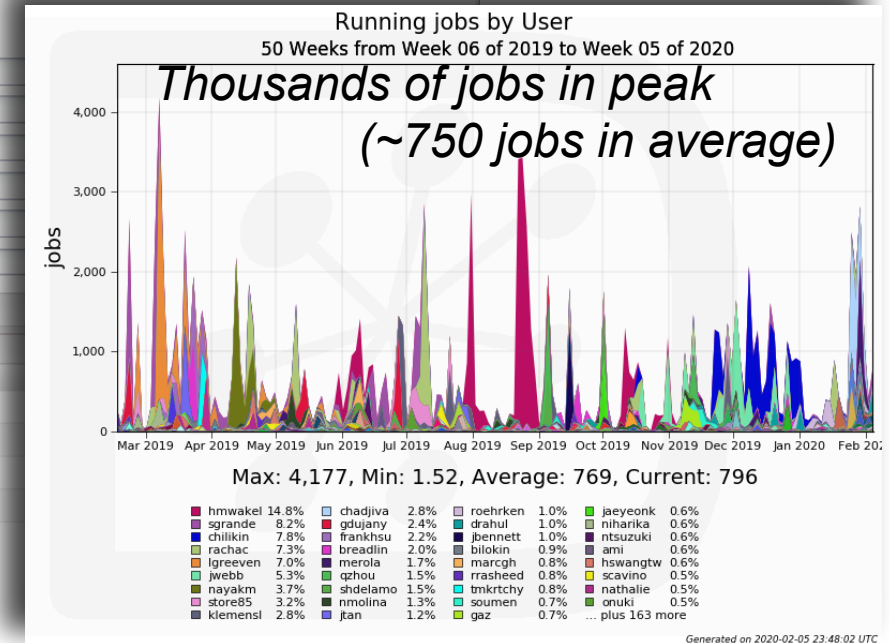
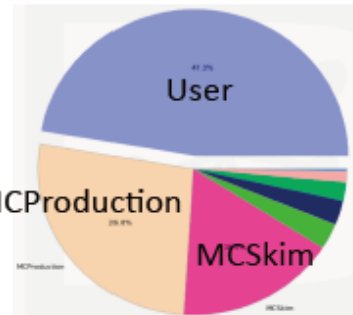


failure reasons



Hirata Hikari (Nagoya)

Type of jobs with (g)basf2 termination



Failure rate relatively high for user jobs

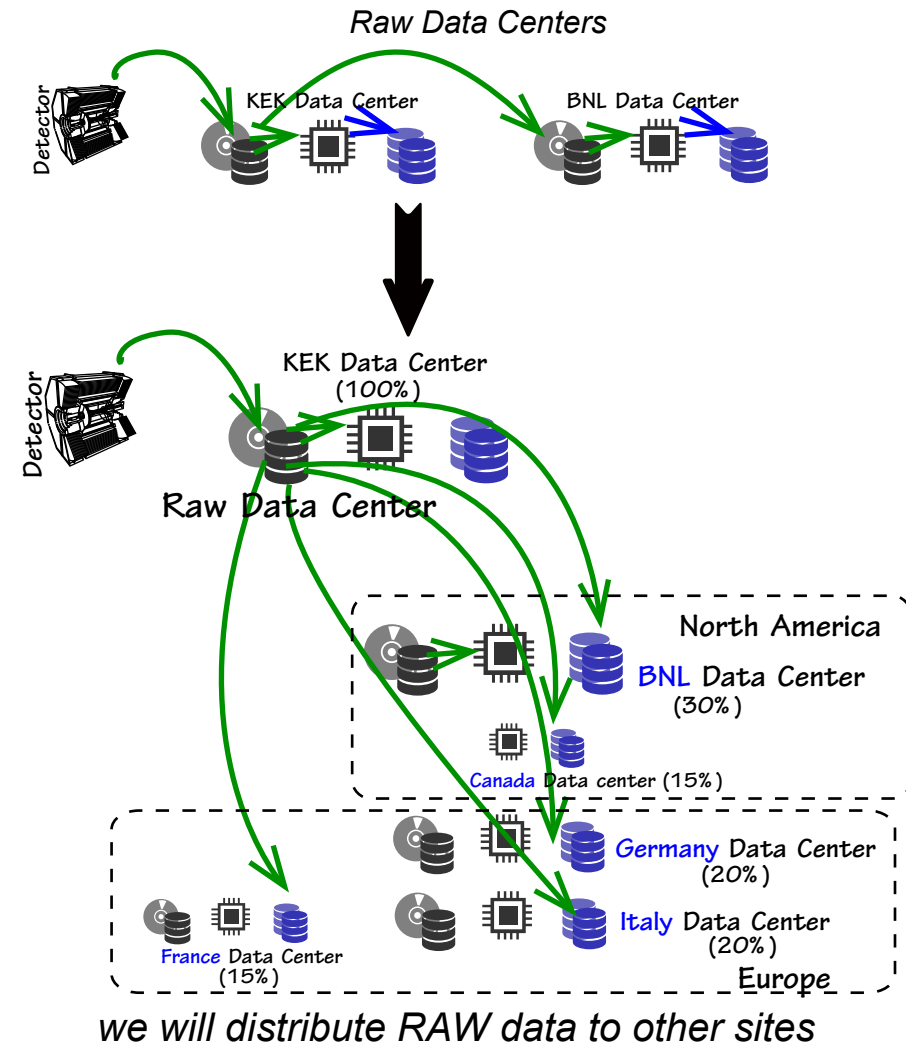
- Majority coming from simple mistakes (syntax error)
- Waste of resources. Additional load on operation

“Scout job” is being implemented

- Send small number of jobs first
- if those finish successfully, full jobs will be submitted

RAW data centers

Silvio PARDI (Napoli)+network providers, site members



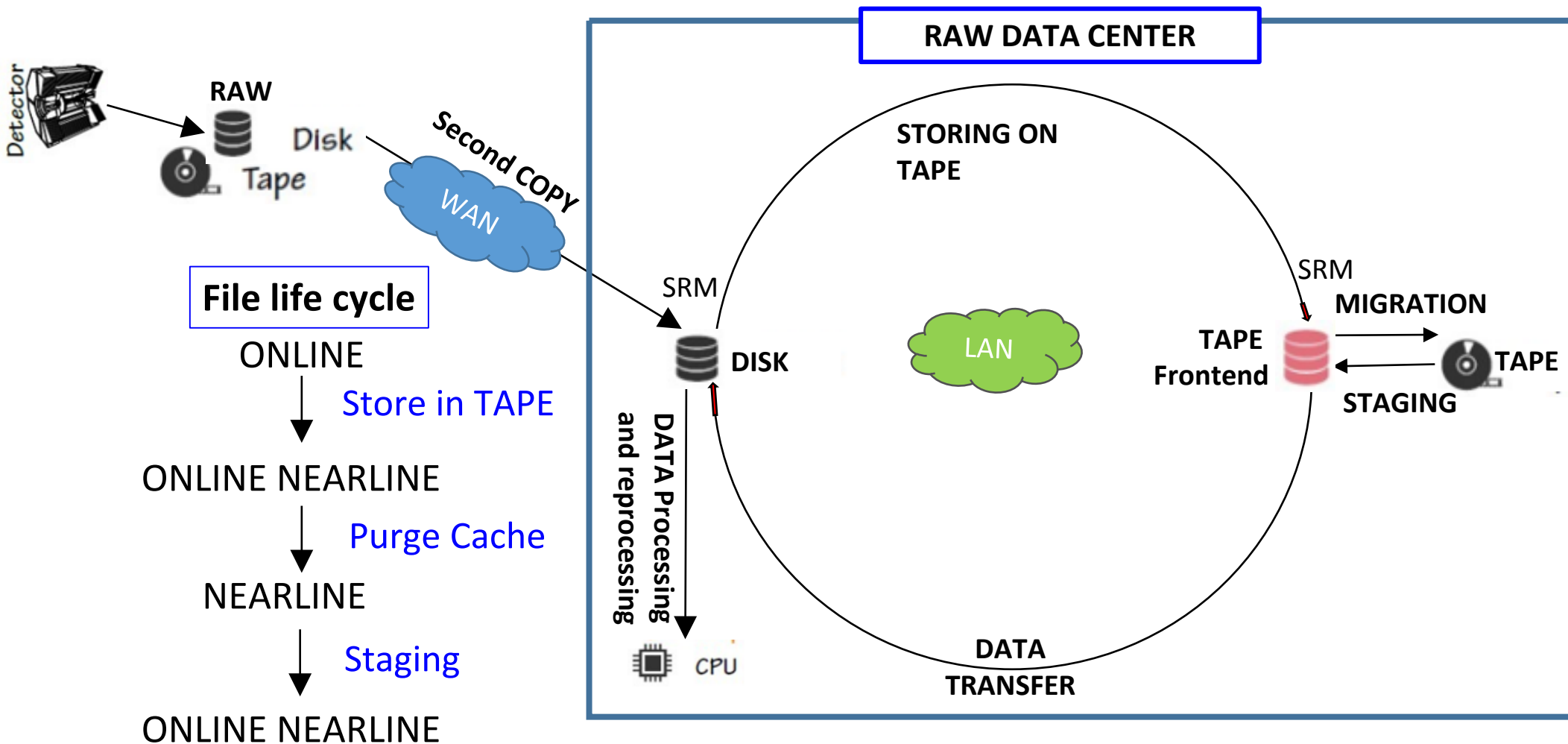
MAXIMUM Performance reached during Network Data Challenge KEKCC → RAW-DC (up to 16 streams per file in FTS Jobs). To be confirmed/improve during the test vs TAPE systems

LINK	Peak (Gbps)	Average (Gbps)	Data per Day (TB)	Site Connection (Gbps)	Peak/Site Connect.	Average/Site Connect.	Security Factor TBperDay /42TB
KEK -BNL	35.0	15.5	167	200	18%	8%	x 4
KEK -CNAF	20.0	15.0	162	200	10%	8%	x 3.8
KEK -DESY	16.0	10.0	108	100	16%	10%	x 2.5
KEK -IN2P3	15.7	14.7	158	100	16%	15%	x 3.7
KEK -KIT	20.0	13.0	140	100	20%	13%	x 3.3
KEK -UVIC	14.0	10.0	108	100	14%	10%	x 2.5

*BNL, DESY, GridKa, MPP, IPP, CNAF, KEK already signed MoU
France will do it, soon*

RAW data centers

Silvio PARDI (Napoli)



Test of RAW data centers

Silvio PARDI (Napoli)

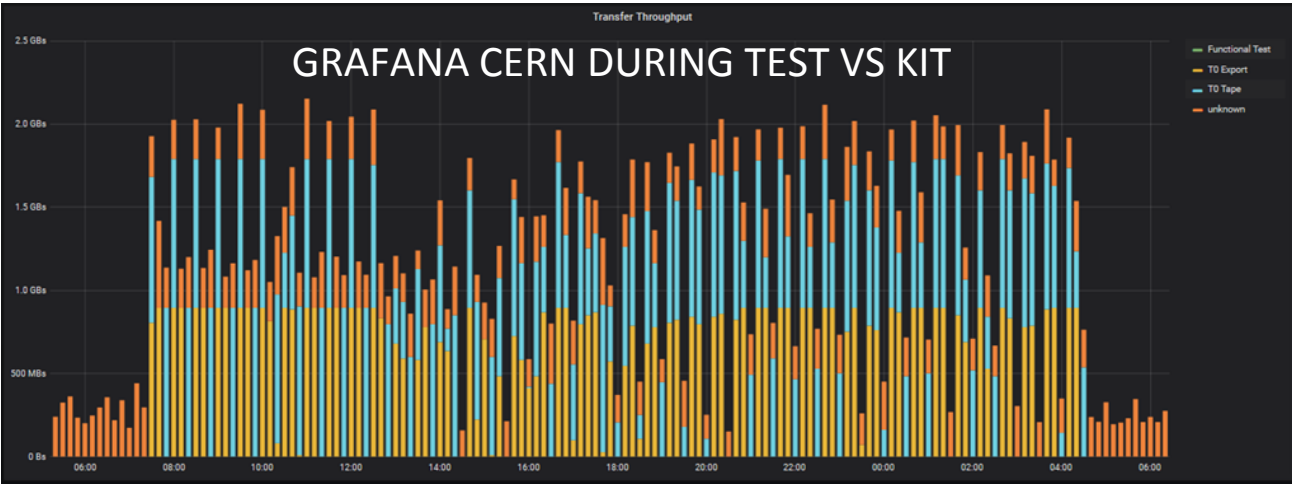
Test of Goals

- *Check the configuration in each RAWDC*
- *Understand TAPE System performance of all RAW Data Centres that will host the second copy of RAW Data starting from 2021.*
- *Test tools for data distribution and activity monitor.*

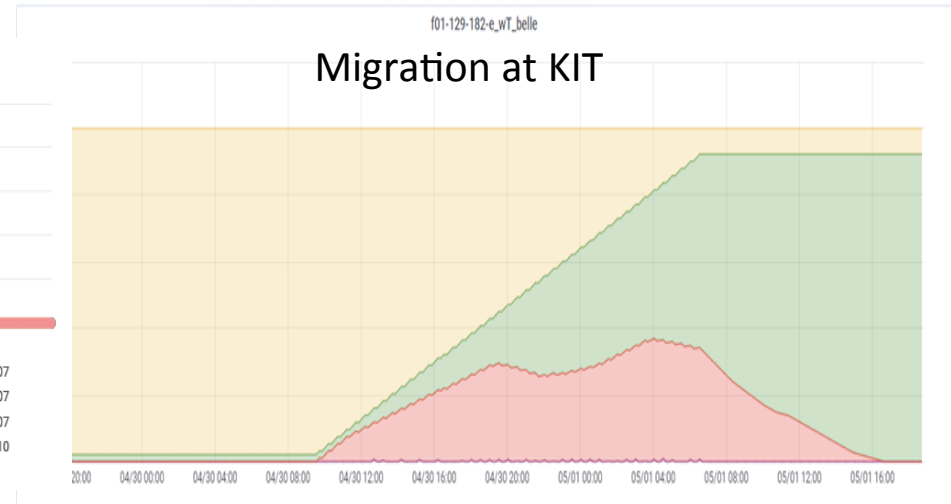
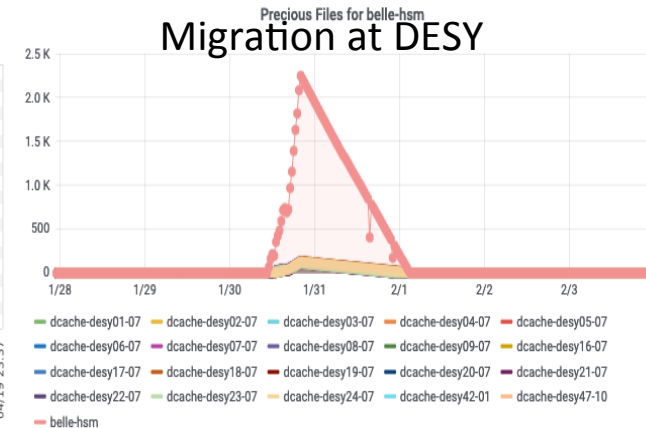
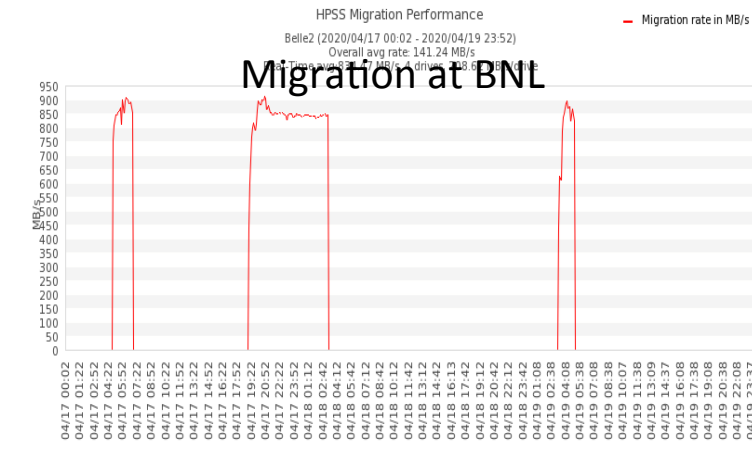
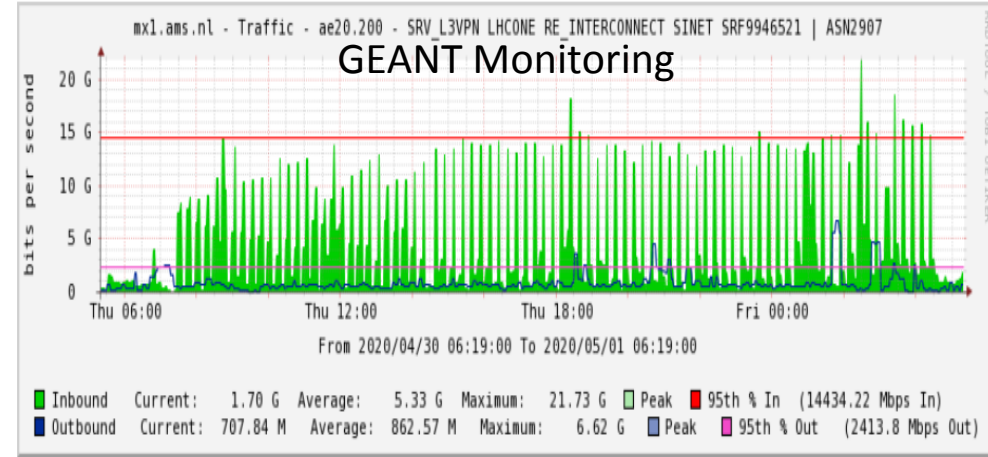
Test to be perform

- *Write in bulk with FTS into SE path with tape backend*
- *Read back again in bulk with FTS from the SE path with tape backend into another SE path with disk-only backend.*

Monitoring tools in action during the test



Silvio PARDI (Napoli)+network providers, site members



Rucio integration in Belle II computing

Cedric SERFON (BNL), Paul LAYCOCK (BNL)

Migration plan

Step 1

Keep using LFC as master catalogue
 Rucio hidden behind current DDM API
 + Ease of migration
 - Restricted functionality

Step 1

Move to Rucio FC as master catalogue
 Change DDM API+
 + Access more Rucio functionality
 - More intrusive

Cedric's risk assessment

In Step 1, Belle II does not use Rucio as the file catalogue, but if the Rucio file catalogue and LFC are out of sync then problem can occur.



Belle II decided to test Step 1 and Step2 in parallel and perform only one migration of Belle II DDM directly to Step 2

Status - DIRAC-Rucio development servers setup at BNL
 - Rucio-as-file-catalogue close to final

Integration - Timescale for migration to Rucio would be early September, after the KEKCC replacement work is finished

New KEKCC

	2016	2020	Upgrade Factor
CPU	Xeon E5-2697v3 (2.6GHz, 14cores) <small>Haswell</small>	Xeon Gold 6230 (2.1 GHz, 20 cores) <small>Cascade Lake</small>	
CPU cores	10,024	15,200	x1.5
HS06	236k	480k (est.)	x2
OS	SL 6.10	CentOS 7.X (7?)	
Disk Capacity	10 + 3 PB (HSM)	17 + 8.5 PB (HSM)	x2
Tape Drive	IBM TS1150 x54	IBM TS1160 x72	
Tape Media	7 TB/vol (JC) 10 TB/vol (JD), 360 MB/s	7 TB /vol (JC) 15 TB/vol (JD-Gen6) 20 TB/vol (JE), 400 MB/s	
Tape max capacity	70 PB	100 PB	x1.4



Time scale:

KEKCC 2016 will be terminated on August 31st.
 cooperation for 1-2 weeks in Sep. (plan)
 KECC 2020 will start on September 1st.
 limited test usage in Aug. (plan)

HSM migration will be executed from the middle of August. (plan)

*Data in home/group directory will be migrated to the new system in background by the system side
 (But final sync might be done by users)*

KEKCCRC tries to minimize the downtime from the user side point of view.

Summary

Belle II has started “phase 3” physics run with the full detector successfully

SuperKEKB accelerator aims

peak luminosity : $2\sim 3 \times 10^{34} / \text{cm}^2 / \text{s}$

integrated luminosity : 200 /fb by 2020 Summer

Belle II detector is well under control for data-taking

RAW data flow from Online storage to Offline storage, then to GRID was established

each process is automatized and all processes work as a system

RAW data has been distributed to outside KEK (currently to BNL)

→ start distributing to Italy, Germany, Canada, and France, too (from 2021)

→ tests of RAD Data Centers including Tape system is being performed

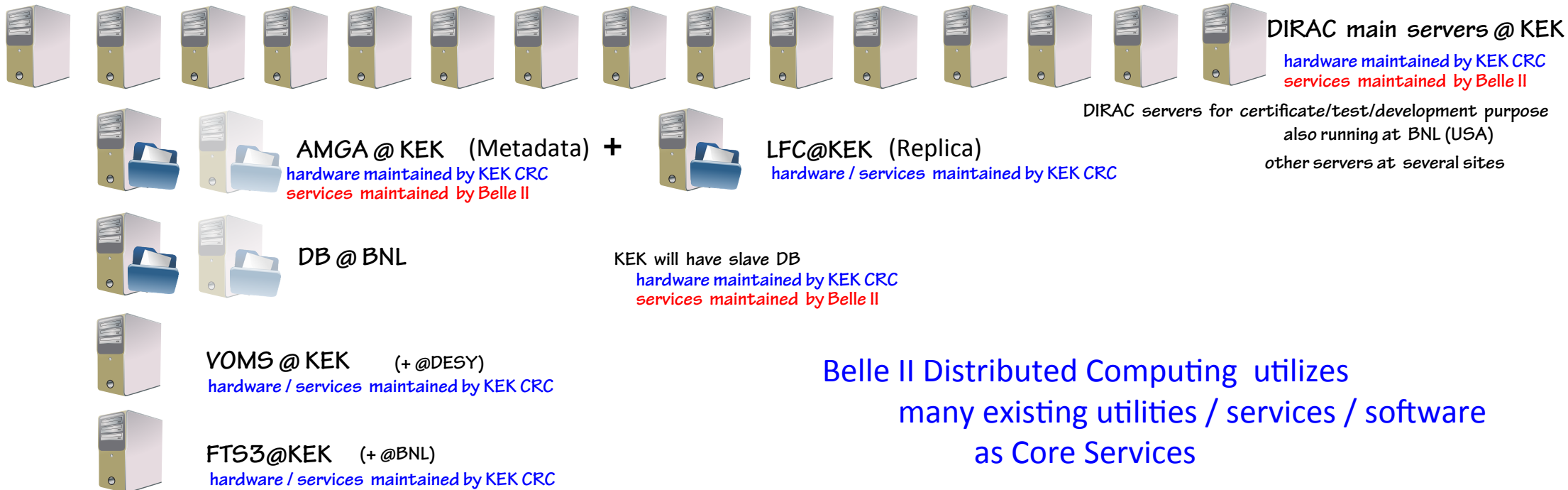
RAW data was (re)processed successfully at remote site (BNL)

Different type of jobs (MC production, Skim production, RAW data process, User analysis) are running on GRID

Improvement is go-going (user scout job, Rucio integration, etc.)

KEK Central Computing system will be replaced in 2020 summer

Belle II Distributed Computing System and services in use



Belle II Distributed Computing utilizes
many existing utilities / services / software
as Core Services



cvmfs /cvmfs/belle.cern.ch/{releases, externals} is used for software distribution
stratum0 @ KEK (under configuration) stratum1 @ KEK, BNL, DESY

DIRAC client installation via cvmfs

Access GOCDB

to get downtime information



ticketing system

non-LCG site: JIRA @ DESY



Grafana +



network monitoring

PerfSONAR mesh is also used

Other Storage providers

