Belle II Computing

Ueda I.

2024.Mar.27 - GDB (co-located at ISGC)

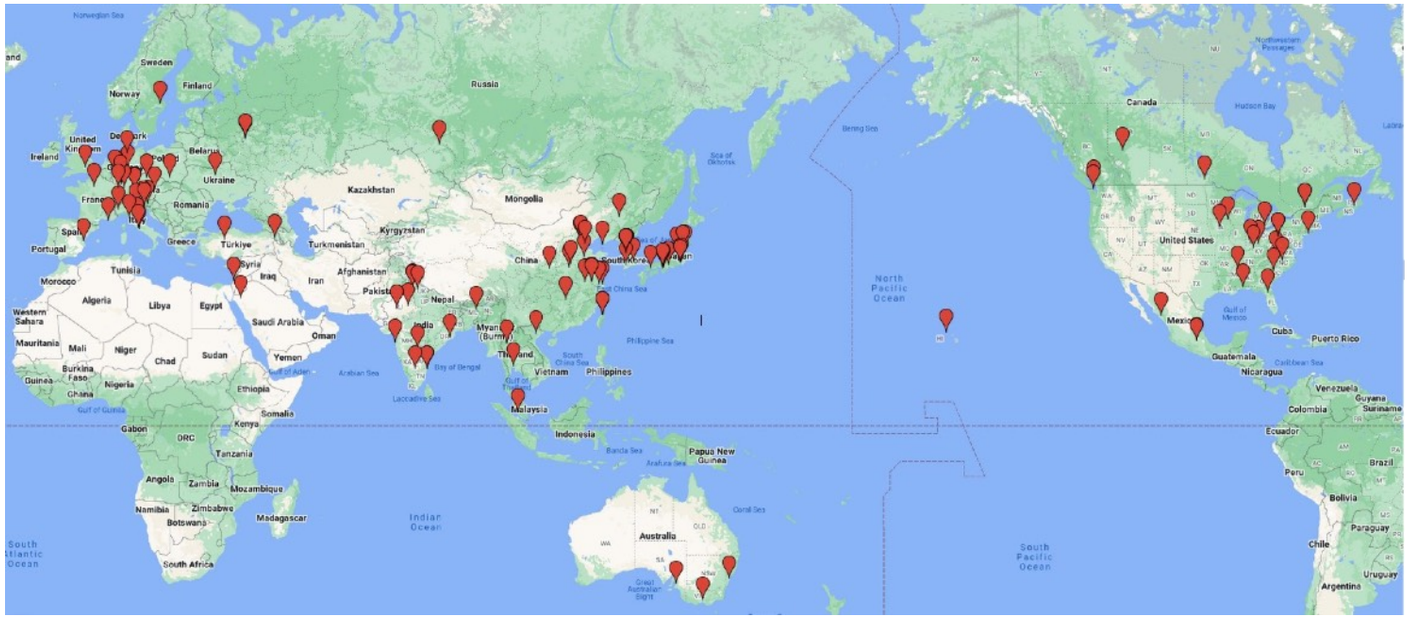
Belle II

Collaboration

- ~1200 members
- 28 countries/regions

Data Taking

- Started in 2019
- Restarted in late Jan. 2024
  after a long shutdown (LS1)



Belle II Computing Coordinators

- T. Hara => I. Ueda since June 2023
- Deputy: M. Hernandez Villanueva since early 2022
- Deputy: C. Serfon since June 2023

Representative for the matters of infrastructure (sites, network)

- S. Pardi (LHCONE, DOMA, ...)

Distributed Computing System

- **DIRAC** as the main system with configuration, workload management + data I/O by jobs
- **Rucio** for data management (distribution, deletion, ...)

Belle II Computing Model

Similar to WLCG, but without using the term "Tier-N"

KEK = Host laboratory (cf. T0)

- To register detector data onto Grid
- Permanent store of 100% raw data
- Prompt processing
- incl. the activities below

Raw Data Centers (cf. T1)

- Permanent store of replica raw data distributed over 6 sites
- Prompt processing and Reprocessing
- incl. the activities below

Regional Data Centers (cf. T1+T2)

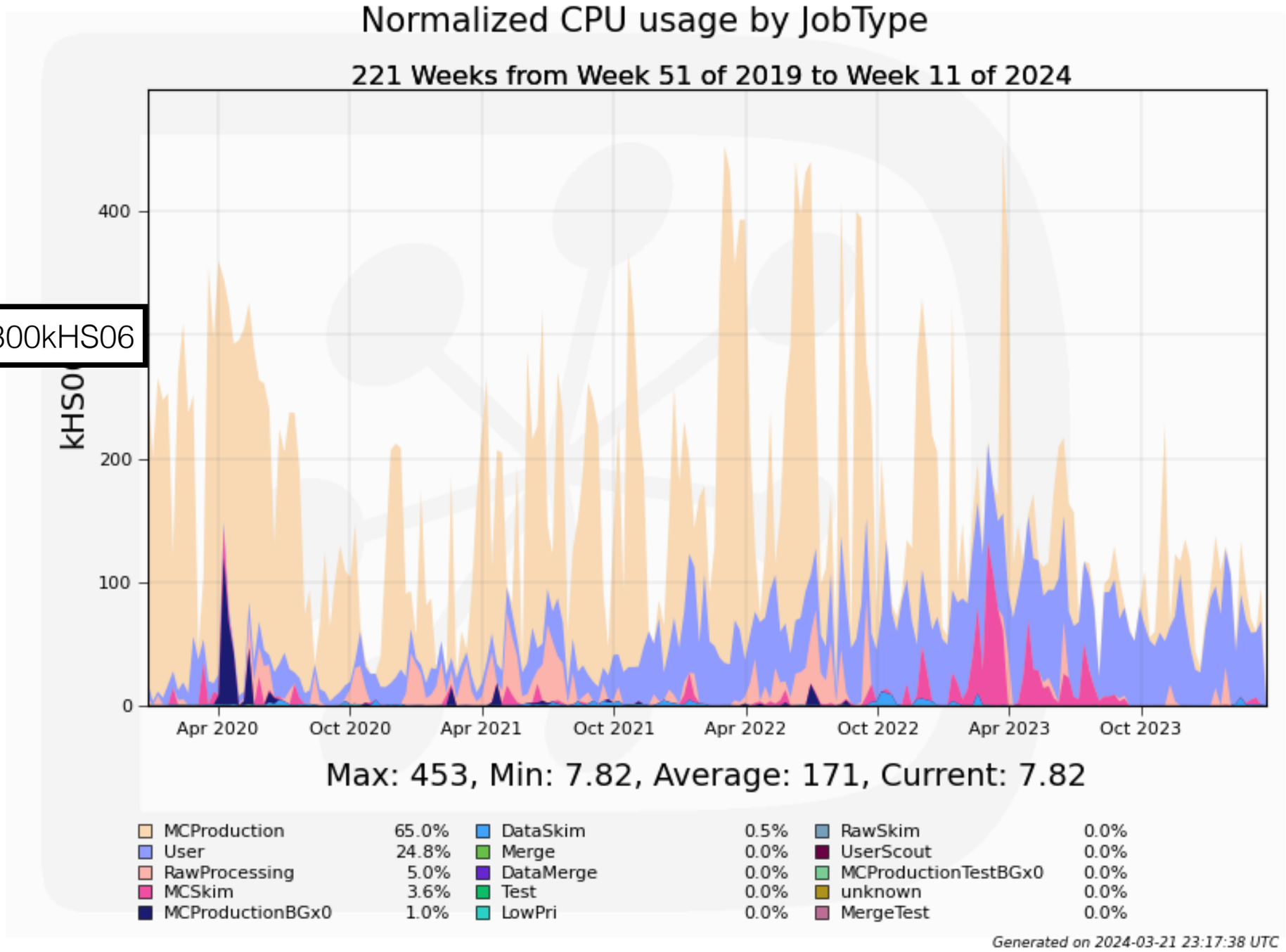
- To host data for analysis
- incl. the activities below

MC Production Centers (cf. T1+T2+T3)

- To run MC production jobs and analysis jobs

Belle II Computing Activities

Jobs



MCProduction

Currently dominant but not constant

Analysis

Increasing on Grid

RawProcessing

Not large while the accumulated detector data are small

MCSkim DataSkim

Evolving to produce more effective data for analysis

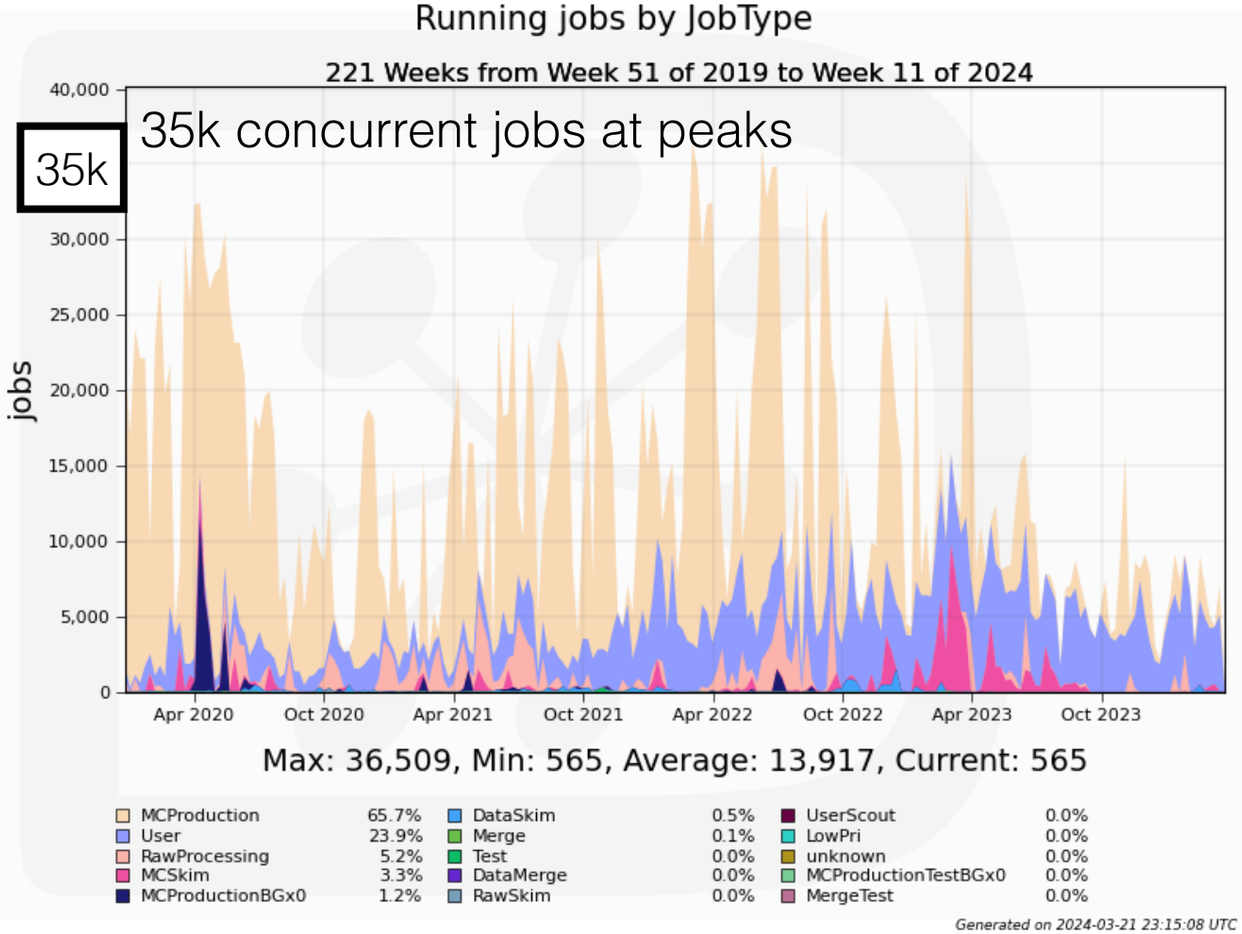
Belle II Computing: Jobs

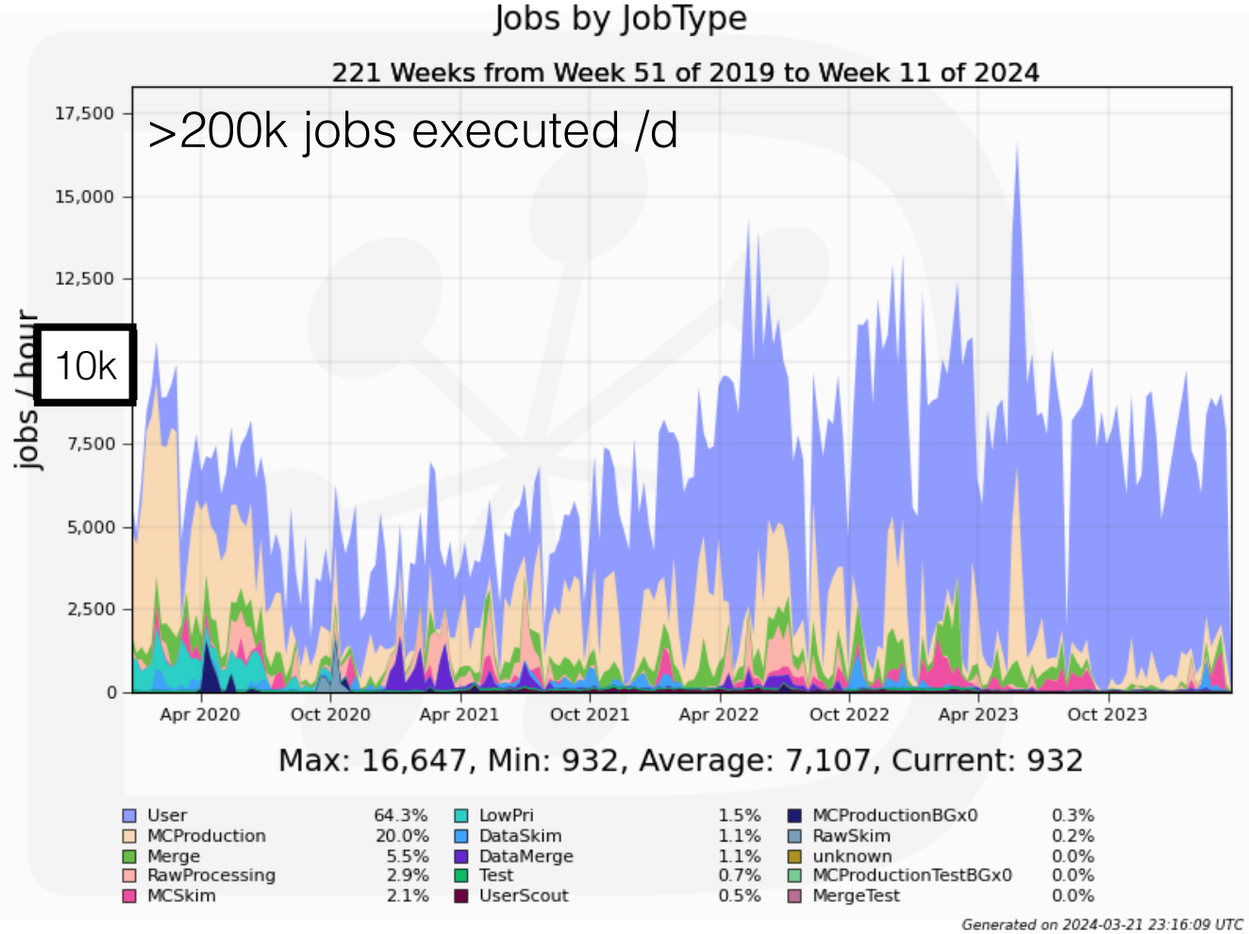
Production jobs = Intermittent

- Usually long jobs with good behaviour, though not constant

Analysis jobs = constant pressure with higher execution rate

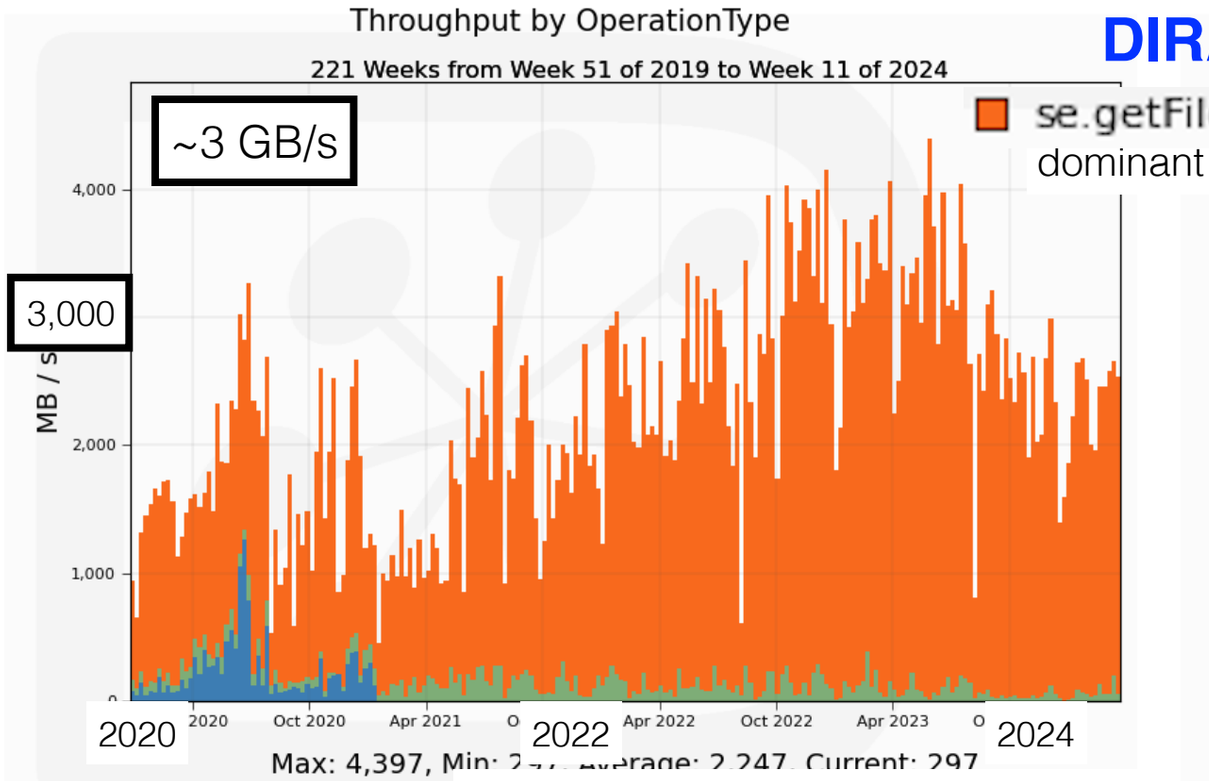
- Large number of short jobs <= Many short "runs"
- Counter measures include merging files beyond runs (redesign of production system)



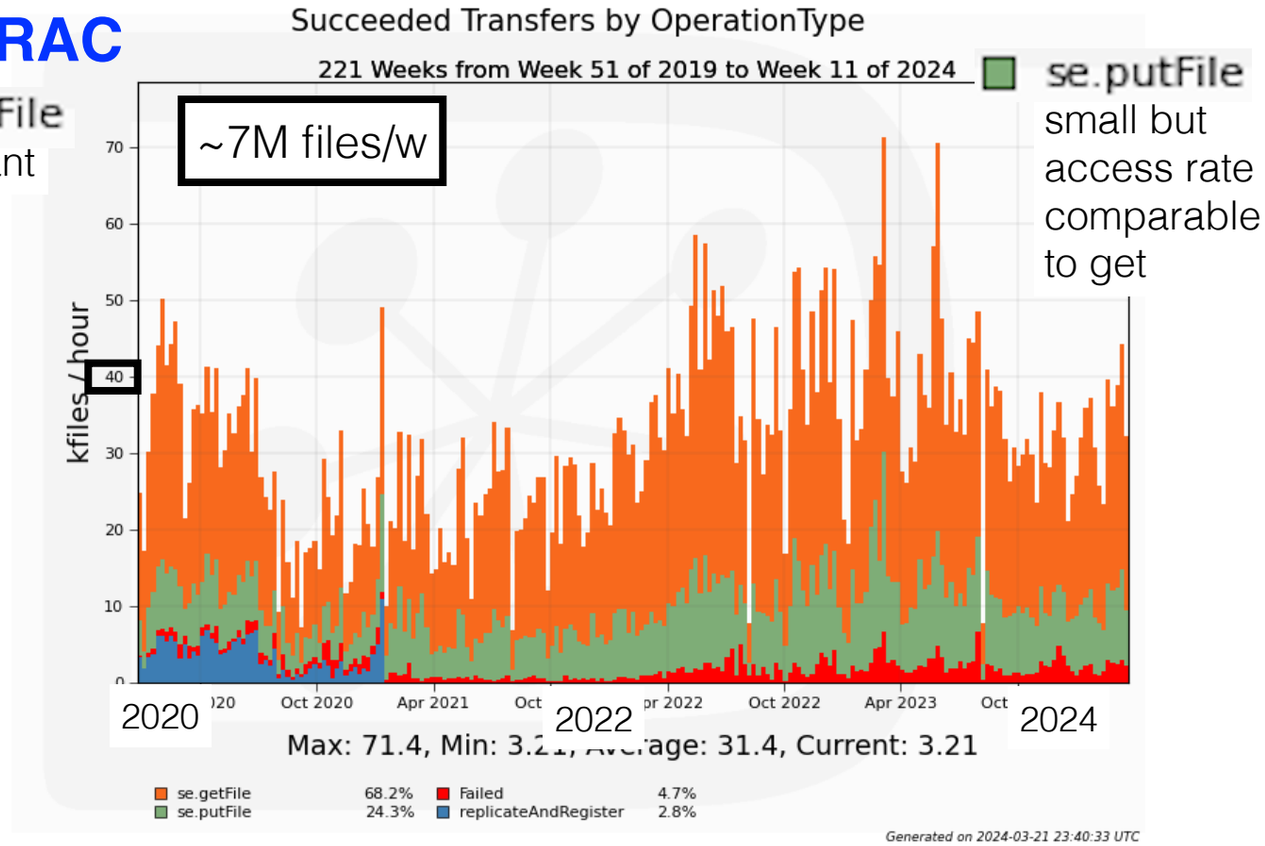


Belle II Computing: Data Access

DIRAC



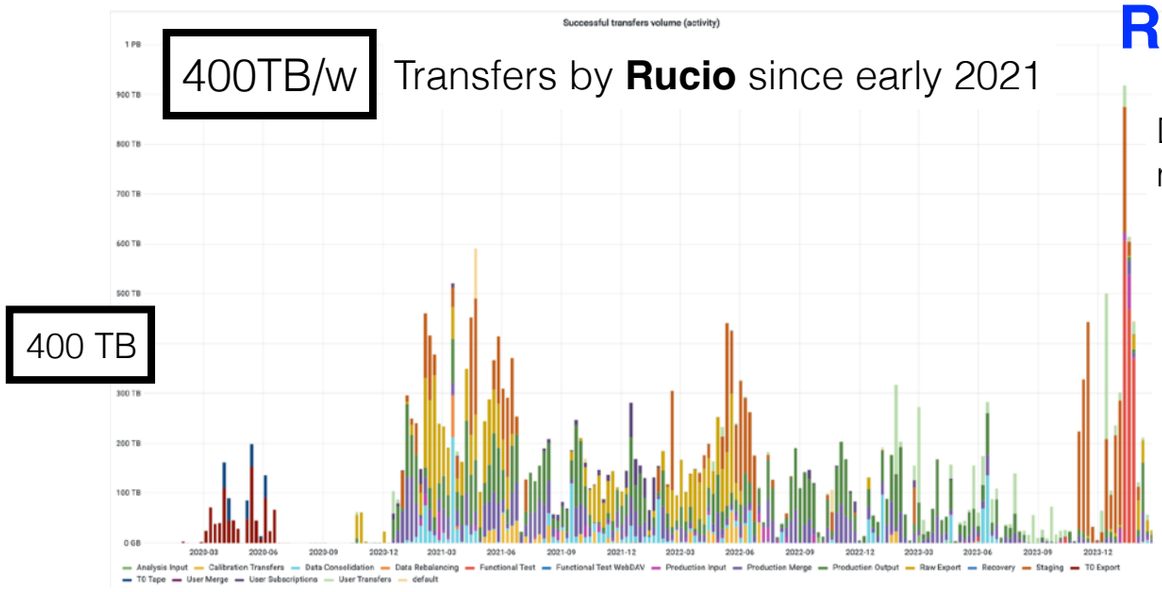
SE-SE transfers by **DIRAC** until early 2021



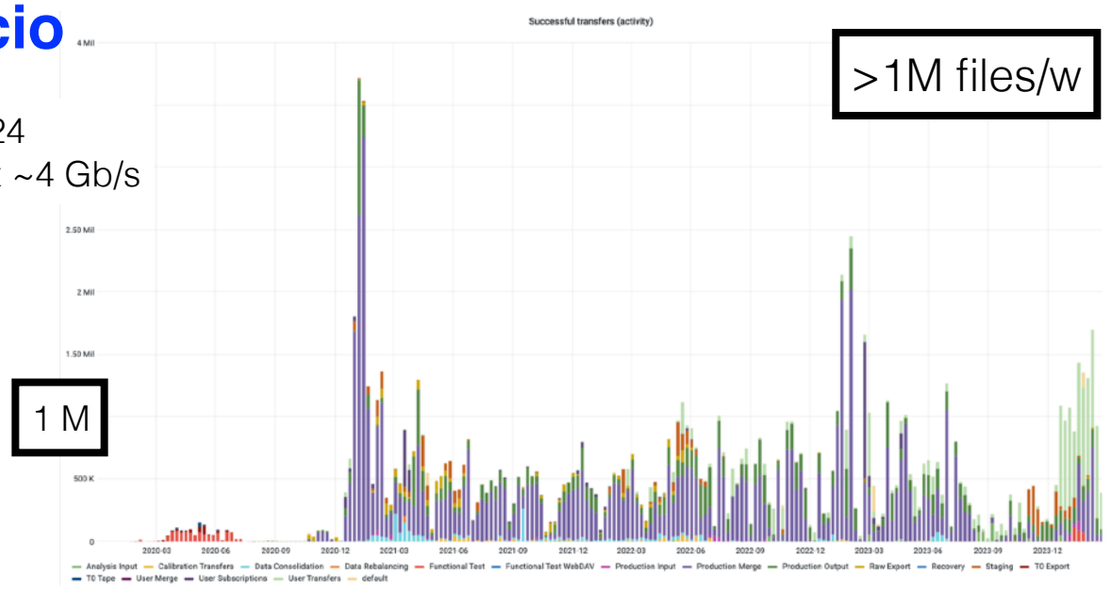
Generated on 2024-03-21 23:39:38 UTC

Generated on 2024-03-21 23:40:33 UTC

Rucio



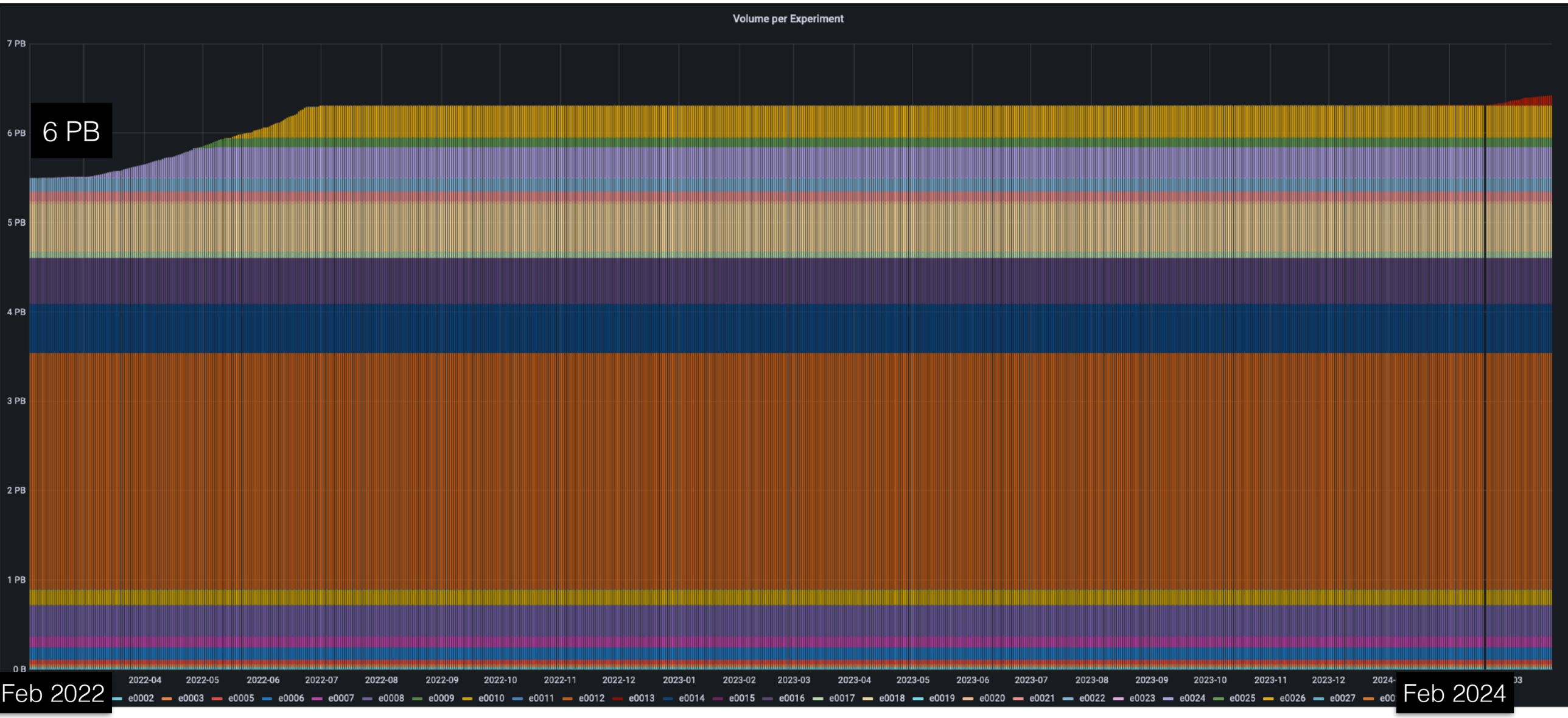
DC24
max ~4 Gb/s



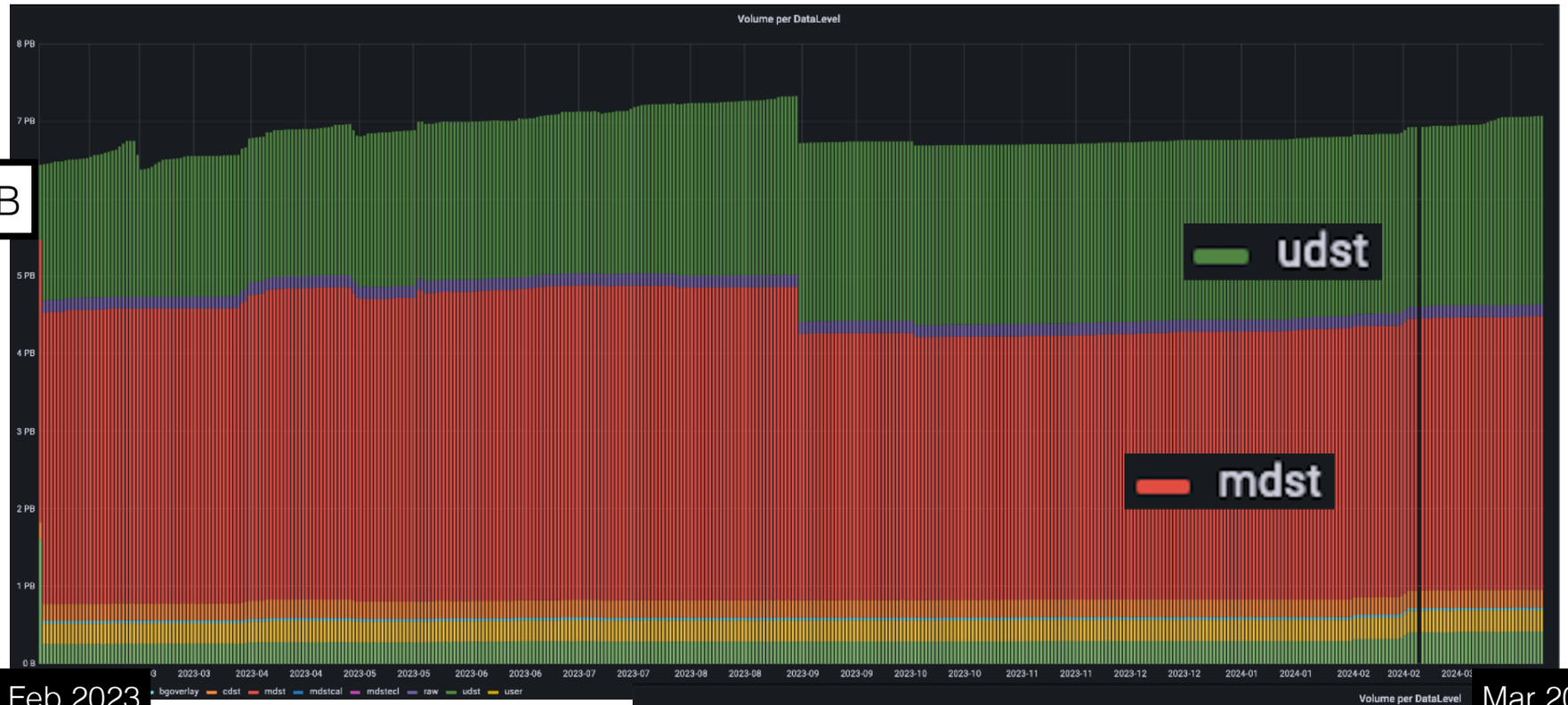
Belle II Data Volume: TAPE

More than 3 PB raw data accumulated

- Incl. skimmed raw data
- Duplicated over Raw Data Centers



Belle II Data Volume: DISK



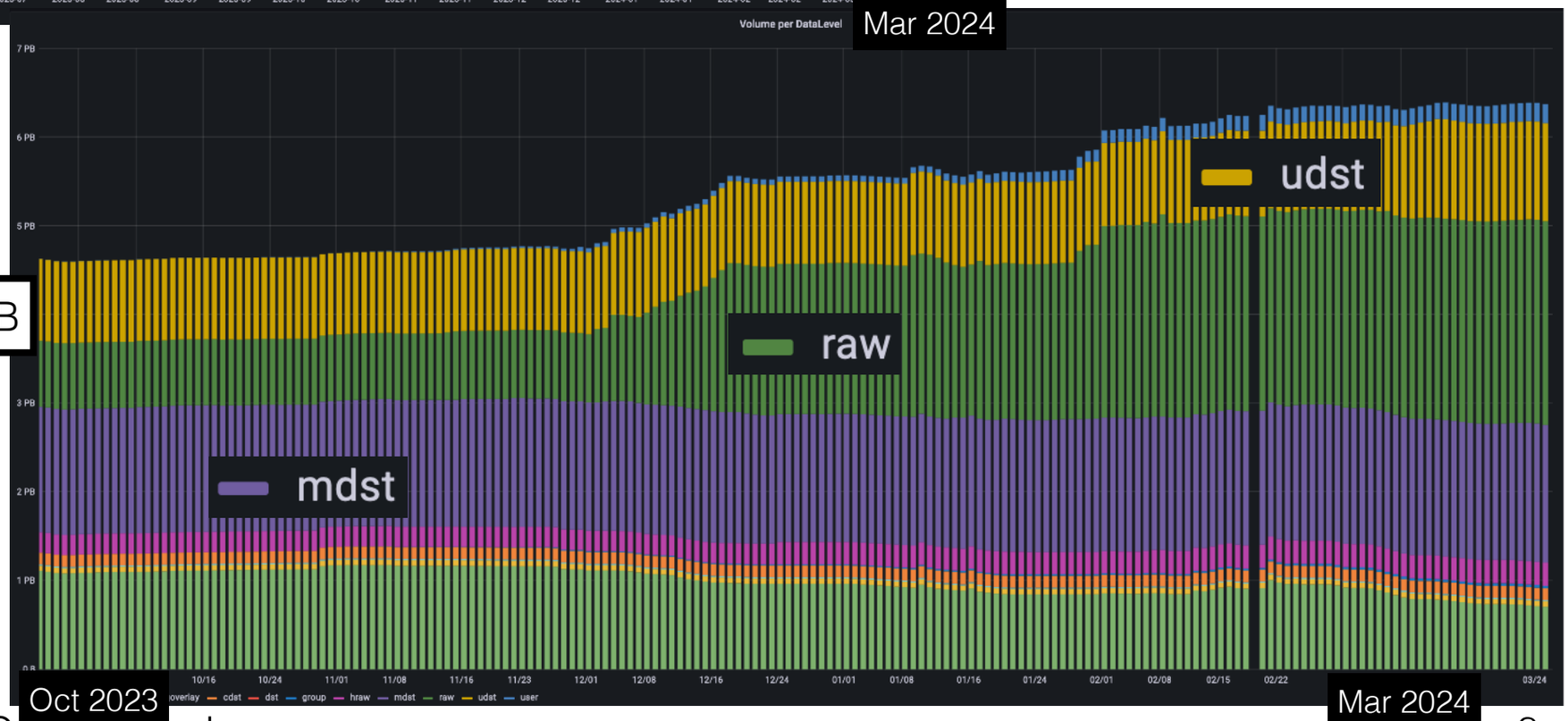
6 PB

Feb 2023

Persistent data on DISK with life cycles based on data production/processing campaigns

mdst: reconstructed data (both real data and MC)

udst: skimmed data produced from mdst, with additional parameters for analysis



4 PB

Mar 2024

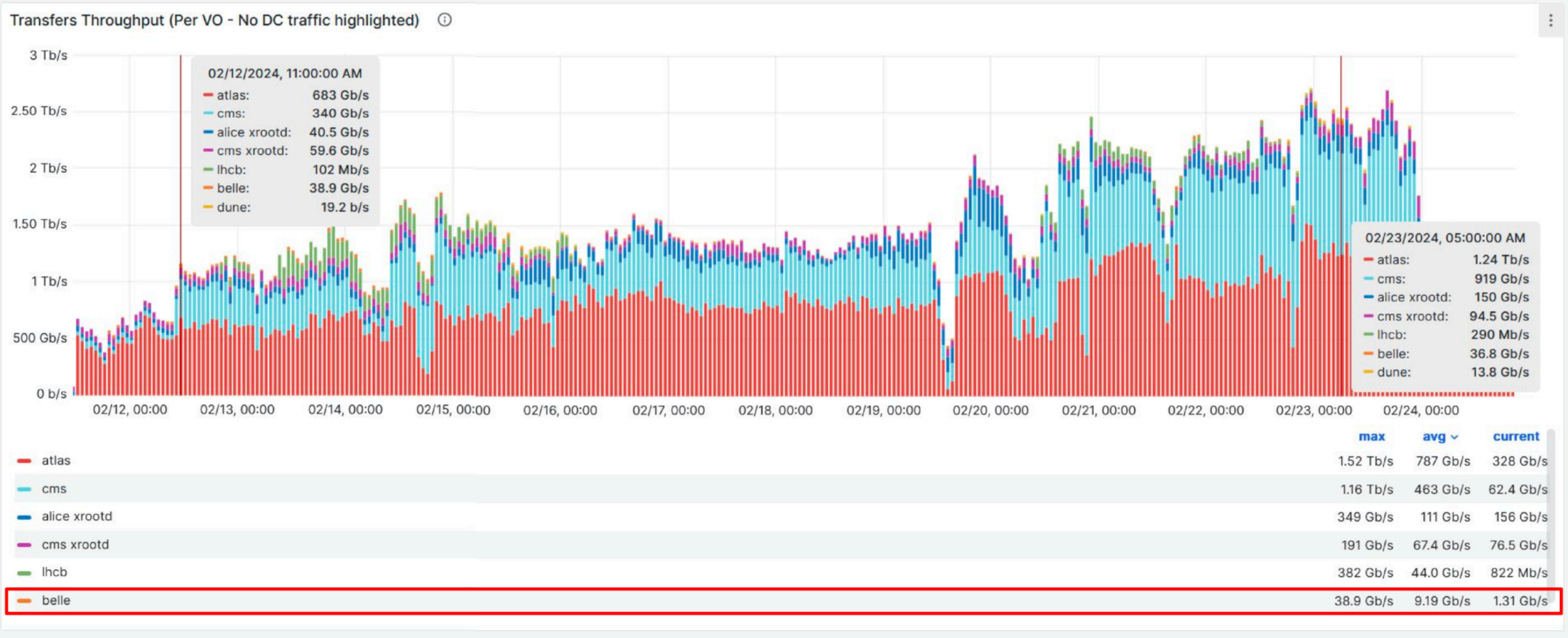
Temporary data/replicas on DISK intermediate output, raw data staged for reprocessing, analysis output, calibration data, ...

Oct 2023

Mar 2024

Belle II in DC24

Belle II on the WLCG Dashboard for DC24



	max	avg	current
atlas	1.52 Tb/s	787 Gb/s	328 Gb/s
cms	1.16 Tb/s	463 Gb/s	62.4 Gb/s
alice xrootd	349 Gb/s	111 Gb/s	156 Gb/s
cms xrootd	191 Gb/s	67.4 Gb/s	76.5 Gb/s
lhcb	382 Gb/s	44.0 Gb/s	822 Mb/s
belle	38.9 Gb/s	9.19 Gb/s	1.31 Gb/s

10

https://indico.cern.ch/event/1388176/contributions/5835937/subcontributions/470515/

Belle II Computing relies on...

DIRAC

- The main system with configuration, workload management + data I/O by jobs
  - We used to have a DIRAC extension for "dataset" distribution and deletion => switched to Rucio
- Hosted at KEK, BNL, and a few other sites

Rucio

- The distributed data management system
- File catalog + Metadata catalog
  - We used to use LFC, migrated to Rucio in 2021
  - We have been using AMGA, supported by KISTI, but now moving to Rucio
- Hosted at BNL

FTS

- To transfer files (protocol = mostly davs, in some cases srm+https, and possibly root)
- Hosted at KEK and BNL

VOMS, IAM

- Hosted at KEK (VOMS replica at DESY for redundancy)
- Test IAM at CNAF, for early phase testing (very helpful)

Belle II Computing relies on...

and of course,

CEs + SEs at the sites, cloud resources

- HTCondor-CE, ARC-CE, dCache, StoRM, some XRoodD/XCache...
- Cloud resources via UVic cloud scheduler, or DIRAC "CloudCE"

CVMFS

- Grid client, Offline software, Conditions data files (Main DB at BNL)

GOCDB

- for downtimes
- DIRAC retrieves downtimes of relevant hosts and disables CEs/SEs
- A DIRAC Agent (made by Belle II) synchronizes downtimes to Rucio

GGUS

- To report issues to the sites
- For some sites to report issues to Belle II

EGI accounting

- To collect CPU/WallclockTime usage for yearly accounting report

Belle II Computing adapting to...

End of GSI

- Transfers are now basically done with davs, with <1% SRM+https for TAPE SEs
- Some Belle II sites kindly staying with the "old" HTCondor-CE, supporting job submission with GSI
- Some sites prepared mapfiles to support job submission with X.509 via SSL (not GSI)
  - https://htcondor-wiki.cs.wisc.edu/index.cgi/search?s=HowToUseProxiesWithSsl&w=1
- Eventually start working with tokens (below)

Tokens

- Supported by DIRAC v8.0 for job submission
  - We have been adapting our DIRAC extension to v8.0, and have been validating it
  - Hoping to upgrade in the coming weeks, but still finding issues... (eg. below)
- Some tests outside of DIRAC have been carried out
  - in cooperation with the volunteer sites

Higher Restriction on Certificate/Proxy

- Some sites upgraded the OS and requires 2048-bit certificates/proxies
- Use of OpenSSL3 in our system would require 2048-bit certificates/proxies
- KEK Grid CA certificate is still SHA-1 (to be renewed in 2024 summer)
- DIRAC v7.3 creates 1024-bit proxies

KEK system renewal (2024 summer)

- New service nodes (DIRAC, etc.) will be RHEL9 (or, should we ask for ALMA9?)

IPv6?

Belle II sites

- No Belle II sites have IPv6-only hosts (yet)
- Some sites suggest possible planning, but no strong requests foreseen yet
  - We may hear some news in the coming "Belle II Site meeting" in April
- Some stats from last year survey:
  - #Sites with Storage in Dual Stack 43%
  - #Sites with WorkerNodes Dual Stack 16%

DIRAC servers at KEK

- Currently IPv4-only
- The next KEK system (after 2024 summer) may have IPv4/IPv6 dual-stack

Rucio servers at BNL

- Currently IPv4-only
- Dual-stack may be tested during the preparation for Rucio upgrade

Outlook

In the coming months

- Switch to 2048-bit proxies before the end of CentOS7
- Upgrade to DIRAC v8.0
- Start using IAM with DIRAC
- Start submit jobs with tokens
  - Or use SSL for those who upgrade HTCondor-CE while we are not ready yet
- Start trying REST API for TAPE SE to replace SRM bringOnline

In the further future

- Much more data to be managed,
- Much more jobs to run