Belle II report

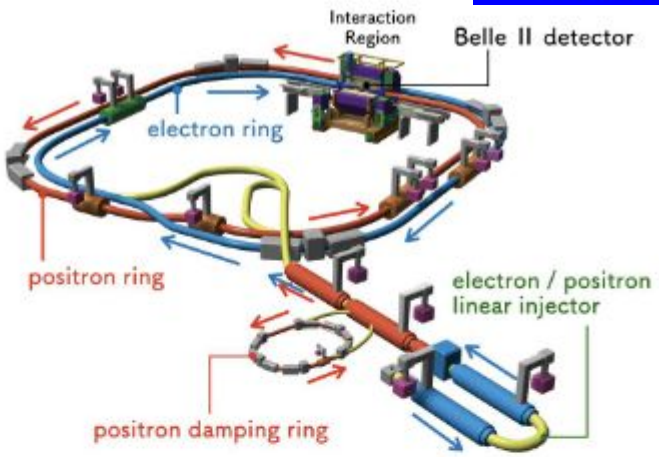
Cédric Serfon (BNL) on behalf of the Belle II Computing Team

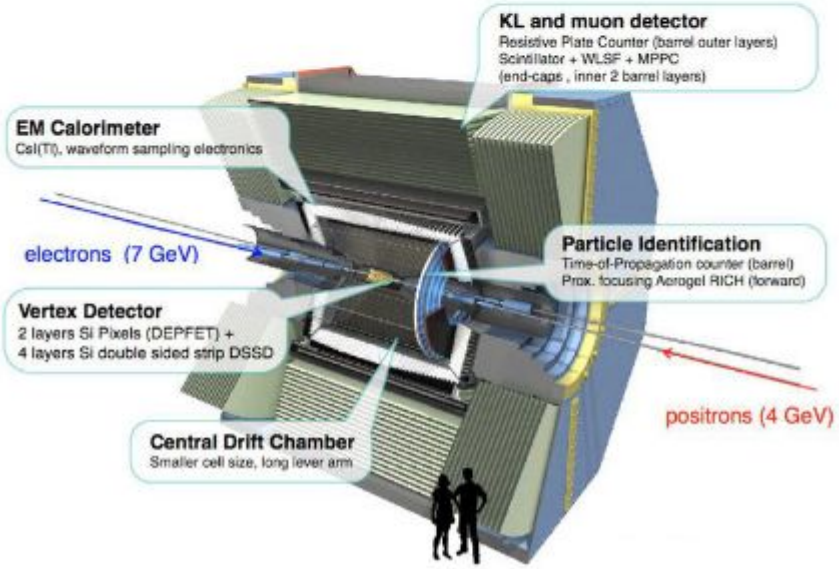
October 17, 2023

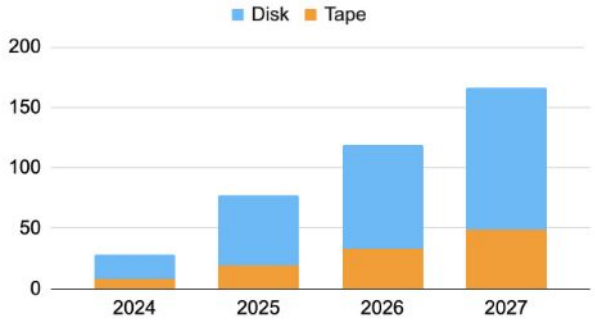
@BrookhavenLab

Introduction

- Asymmetric lepton collider.
- Upgrade from previous Belle experiment.
  - 50 ab^{-1} at the end of the experiment (x50 than the previous B factories)
  - Estimated size of the dataset collected by the experiment is O(10) PB/year.



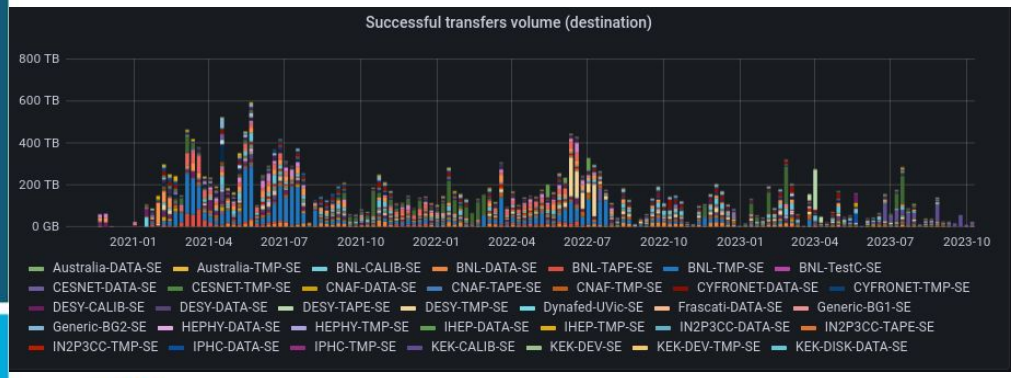




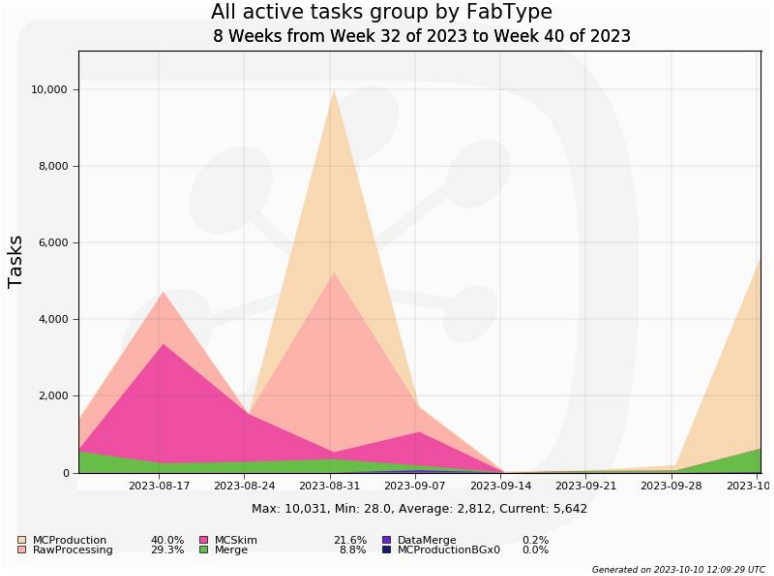
Brookhaven National Laboratory

Belle II computing

- Belle II has been using DIRAC for many years + Rucio (for almost 3 years) in production now :
  - DIRAC as the main framework with user interface for productions and user jobs
  - Rucio used for data management. Direct use of client tools only by the experts
  - Configuration (SEs, accounts, etc.) in DIRAC and sync'ed into Rucio via dedicated Agents



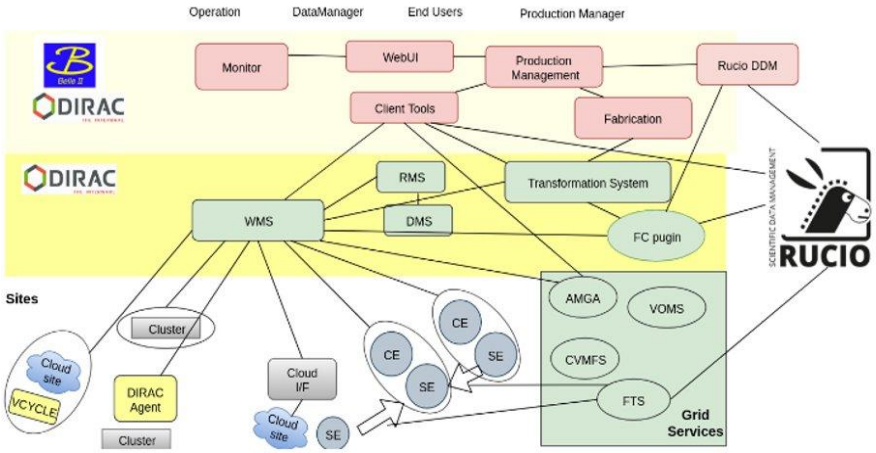
Weekly transfer volume



Production tasks over the last 2 months

Belle II computing

- Belle II has its own extensions :
  - “BelleDIRAC” :
    - Enables a transparent experience using the Belle II Analysis Framework (basf2)
    - Provides an interface to other services used by Belle II (AMGA manager, conditions DB via basf2, etc.)
    - Implements Production System
  - “BelleRAWDirac” :
    - Take care about registering RAW data



Services (DIRAC)

- Production :
  - 11 DIRAC servers + 4 MySQL + 3 Elasticsearch + 2 Web servers (KEK)
  - DIRAC servers to run special SiteDirectors (Niigata, University of Victoria), Vcycle (Napoli)
  - DIRAC server related to Data Management at BNL
  - ReqProxy (KEK, Niigata, Napoli, …)
- Test servers at BNL :
  - Certification: validation of new BelleDIRAC releases.
  - Migration: test of base DIRAC upgrades.
- Development :
  - Multiple instances at KEK, BNL, DESY, Mississippi, etc.

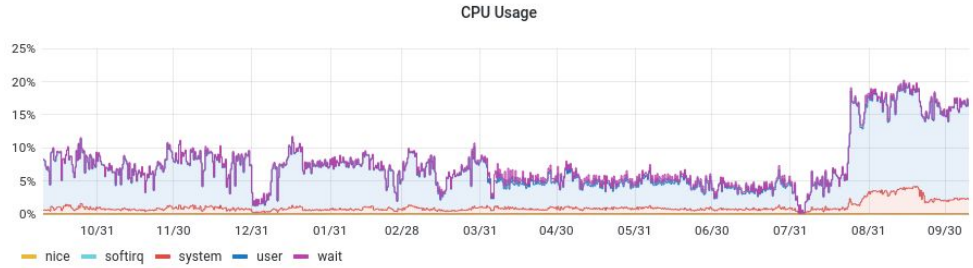
DIRAC installation at Belle II

- Currently in production: v7r3 with python3 (except pilot still python2)
- Migration to v8.0 (then v8.1) hopefully by the end of this year :
  - Needed for token support in particular submission to HTCondor (currently workaround by using SSL (non-GSI) connection)
  - Also needed not to be too delayed in moving to DiracX

Week	Sep 4	Sep 11	Sep 18	Sep 25	Oct 2	Oct 9	Oct 16	Oct 23	Oct 30	Nov 6	Nov 13	Nov 20	Nov 27	Dec 4	Dec 11	Dec 18	Dec 25	Jan 1	Jan 8	Jan 15	Jan 22	Jan 29
Belle DIRAC									**5.8.2**							**5.9 (or 6.0)**						
KEK Dev	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.1	v8.1	v8.1	v8.1
BNL Mig	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0
BNL Cert	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0
KEK Prod	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v7r3 py3	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0	v8.0
Events							DIRAC/ RUCIO	B2GM									Christmas / New year					B2GM

Services (Rucio and other)

- Production (all at BNL) :
  - 1 PostGreSQL server
  - 2 Rucio servers
  - 2 Rucio daemon nodes
  - 2 tracer nodes
  - 1 WebUI node
- Test servers at BNL :
  - 1 test Rucio instance running server + daemon
  - 1 DB node
- Other services :
  - FTS servers (KEK & BNL)
  - CVMFS (KEK) for DIRAC distribution + Rucio and (Belle)DIRAC client
  - VOMS + Test IAM instance
  - ELK stack + Grafana
  - AMGA



DB usage

Rucio installation at Belle II

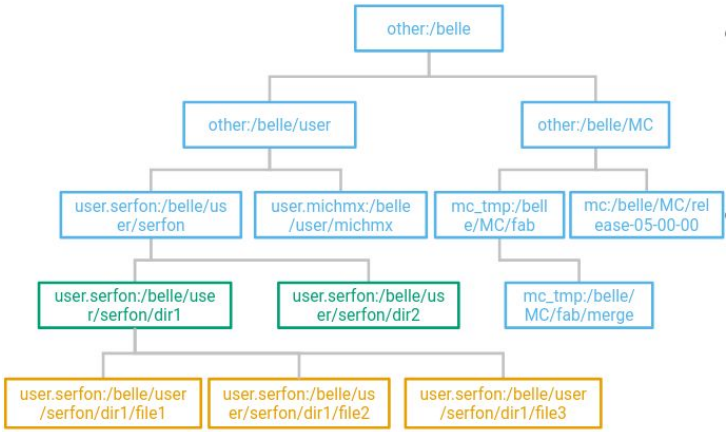
- Current release used in production : 1.28.7
- Migration to latest release was delayed by python3 migration in BelleDIRAC.
  - Plan to move after pilot is python3 ready to the latest release (during winter shutdown ?)
- Since we use an old version, no use of temporary tables in all the daemons

Interplay DIRAC-Rucio

- Integration of Rucio and DIRAC in Belle II based on :
  - Rucio DDM Service that allowed to provide a DDM interface similar to the old DDM Service (BelleDIRAC) to all other components (Production System, etc.), but using Rucio as backend (Belle II specific)
  - Rucio File Catalogue (RFC) plugin that provides the same Catalogue methods as LFC or DFC (not Belle II specific)
- Replication policies based on Rucio subscriptions evaluated every time a new Data IDentifier (Rucio file, dataset or container) is created
- Interaction with Storage Elements :
  - Upload/download still done using the DIRAC DataManager. Recently added the option to use Rucio download client
  - Transfer by Rucio (transfers by DIRAC disabled)

Interplay DIRAC-Rucio

- Namespace :
  - Scope hidden to DIRAC (extracted automatically from LFN)
  - Files/datasets/containers always created by a specific method from the RFC (addFile) that allows to create hierarchical namespace (similar to LFC or DFC)
- Possibility to “break” the hierarchy by using “collections” (aggregation of different datasets into containers similar to what is done in ATLAS)



Latest developments in BelleDIRAC/Rucio

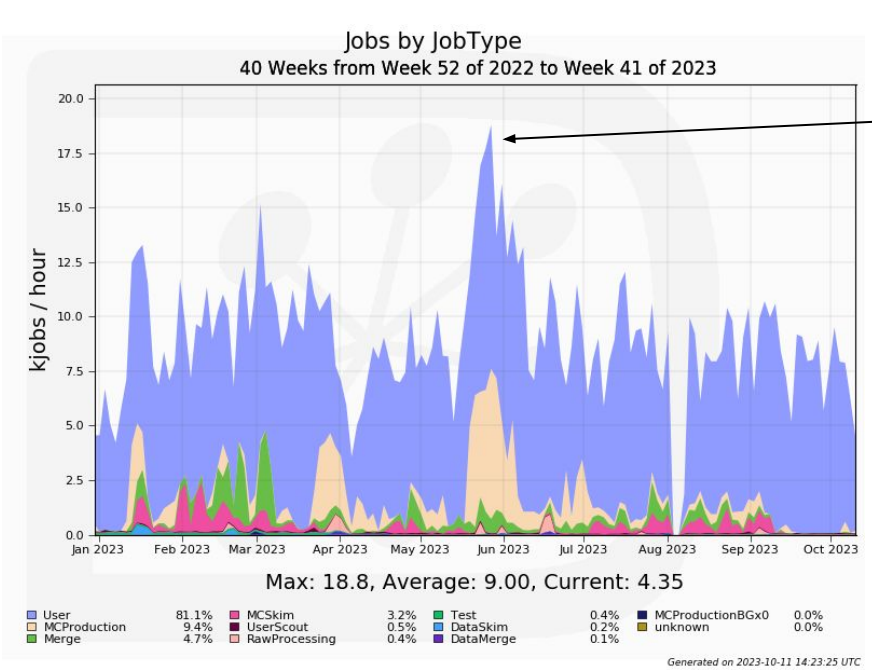
- Many new developments will go into production soon :
  - Staging daemon : Automatically staged data from TAPE. Interfaced with BelleDIRAC Production system
  - New metadata service : Metadata will be registered into Rucio
  - Popularity report : Record all the files/datablocks accesses into Rucio
  - Simplification of DDM Service layer that was done to ease the transition to Rucio
  - New pilot command "BelleSetupDIRAC" to sets up DIRAC env. using the client installation on cvmfs instead of installing DIRAC on-the-fly
    - We started working on this before PR:205 so might need to adjust our implementation accordingly

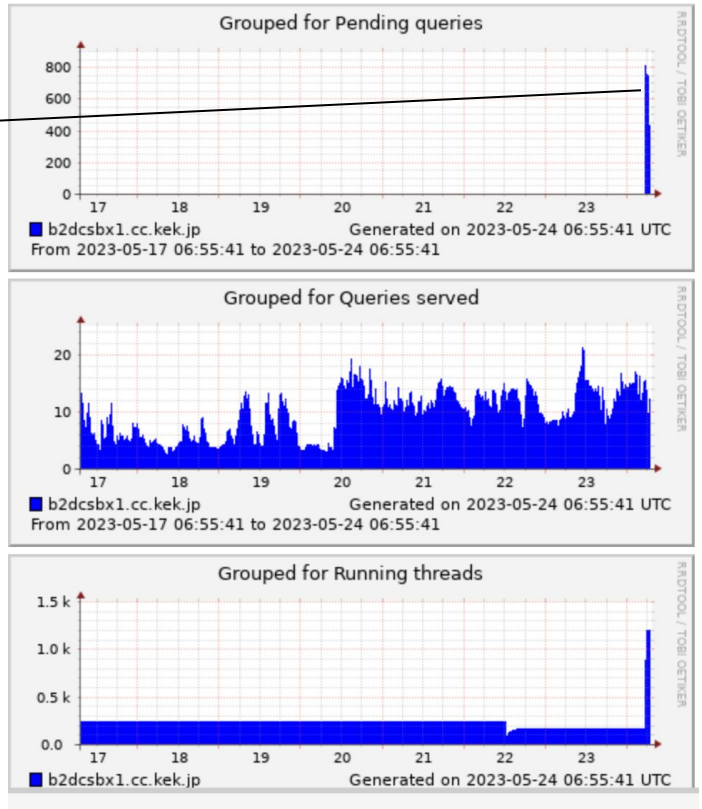
Future plans

- Move to token
- Participation to DC24
- Work on integration of DiracX
- Continue to be involved into Rucio development, as well as increase participation in DiracX development

DIRAC Questions

- Any notable operations incident in the last year ?
  - Input sandbox overload with high execution rates (~ 12K jobs/hour or higher), triggering timeouts.



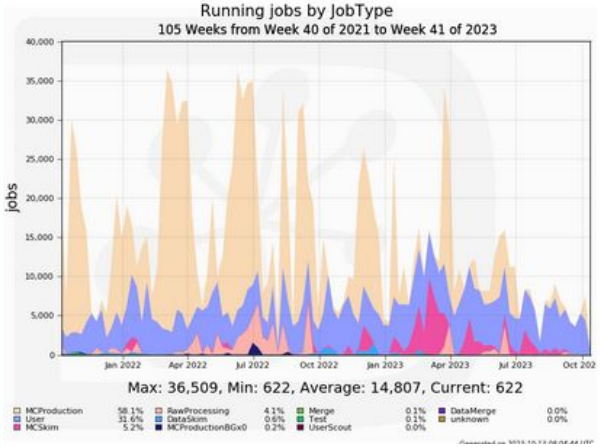


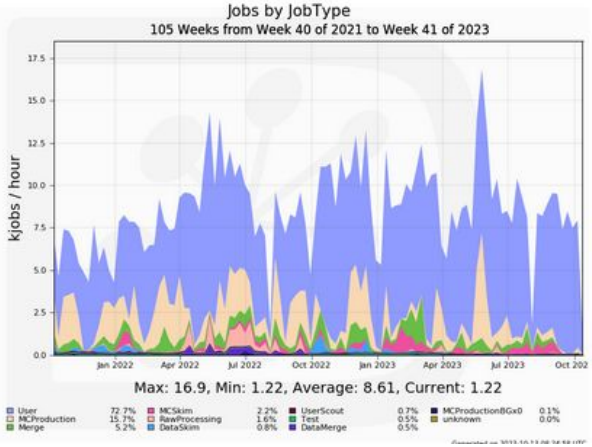
DIRAC Questions

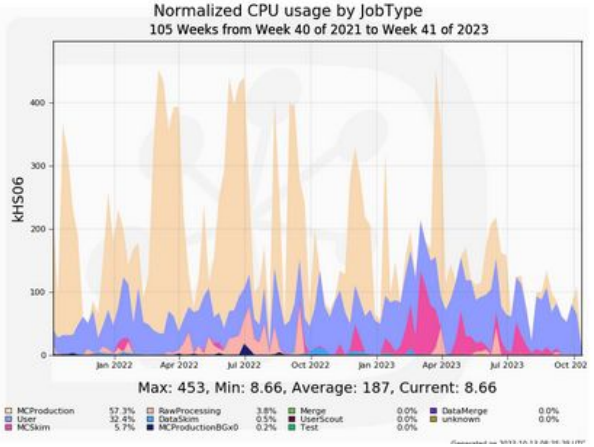
- What is your biggest frustration with DIRAC ?
  - Once jobs are matched to resources, it is not possible to reassign them to other site with a rescheduling. Would be desirable when failure rates are too high in a site.
  - Complex, steep learning curve for newcomers. Maybe too naive and everyone knows this, but makes very hard to get new developers.
- What can be improved for communication in the project ?
  - Nothing

DIRAC Questions

- In the last two years, what has been the DIRAC usage in terms of jobs ran, CPU (or wall time) used, and data transfers ?

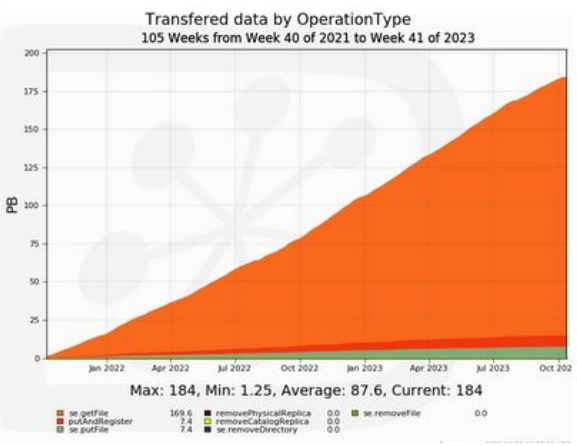


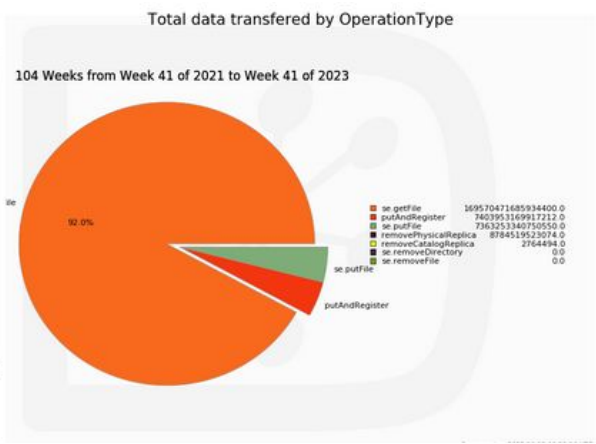


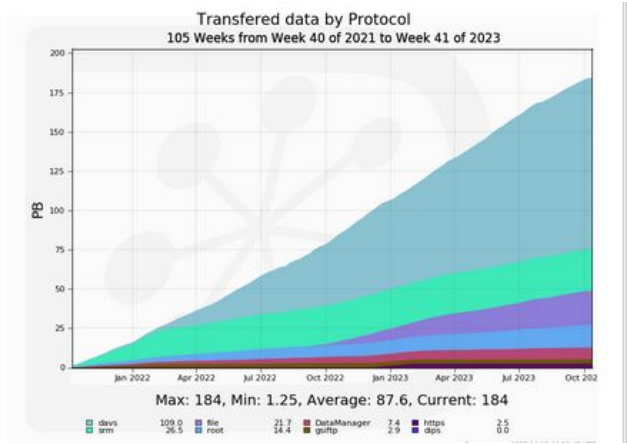


DIRAC Questions

- In the last two years, what has been the DIRAC usage in terms of jobs ran, CPU (or wall time) used, and data transfers ?
  - Data transfer between SEs handled by Rucio
  - Upload/Download handled by DIRAC







DIRAC Questions

- What are your expectations with DiracX ?
  - Pretty much what was already presented on BiLD: easy to deploy, welcoming to newcomers, based on standards.
  - A seamlessly migration procedure from Dirac to DiracX
  - Further integration with Rucio

Rucio Questions

- What is your biggest frustration with Rucio ?
- What features are missing from Rucio ?
  - Nothing as of today
- Any notable operations incident in the last year ?
  - High load on DB from reaper (capped at 300k replicas deleted/day). After changing DB parameters, problem disappeared and we reach 2M replicas deleted/day
- What can be improved for communication in the project ?
  - Nothing

What would you improve in DIRAC & Rucio interoperability ?

- Provide an easy way to set up a DIRAC instance using Rucio (improve documentation)
- Include some Rucio specific unit-tests into DIRAC (or DiracX)
- Combined monitoring (based on Kibana or Grafana)
- Better integration of Rucio into DIRAC dataManager (not in DIRAC, but DiracX)

Conclusion

- Belle II is happily using both DIRAC and Rucio
- Deployment lags a bit behind :
  - DIRAC (v7.3) and Rucio (1.28.7)
  - Slowly catching up
- About development :
  - Continue further on integration of Rucio and (BelleDirac) : Metadata, staging, etc.
  - We are eager to move to DiracX and to take part to its development as well as Rucio