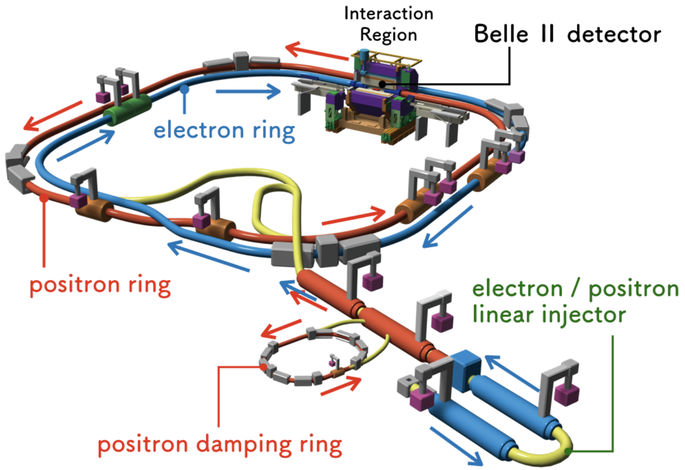
Evaluation of Rucio as a Metadata Service for the Belle II

Panta, Anil (University of Mississippi)
Serfon, Cedric ; Ito, Hiro ; De Stefano Jr, John Steven ; Mashinistov, Ruslan; Laycock, Paul (BrookHaven National Laborotary)
Hernandez Villanueva, Michel (DESY)
Miyake, Hideki; Ueda, Ikuo (KEK/IPNS)

CHEP, 2023

Introduction



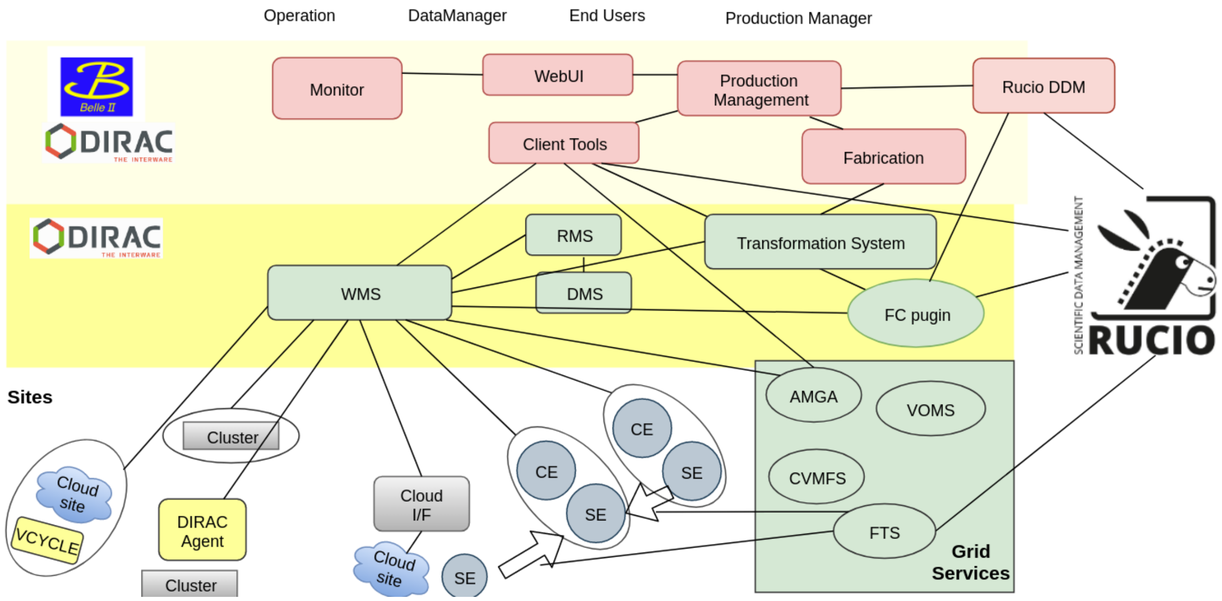
- Belle II is a B-factory based at KEK (Tsukuba Japan).
- International collaboration of institutes all over the world that started data taking in 2019.
- Belle II uses a distributed computing model based on standard tools in HEP:
  - DIRAC for workload/workflow management.
  - Rucio for Data Management.
  - FTS for file transfers.
- For metadata Belle II uses the AMGA service initially developed for LCG :
  - Service not used outside Belle II (i.e. limited support)
  - Rucio provides also metadata functionality and is supported by a wide community
    → Decision to evaluate it.

DIRAC
THE INTERWARE

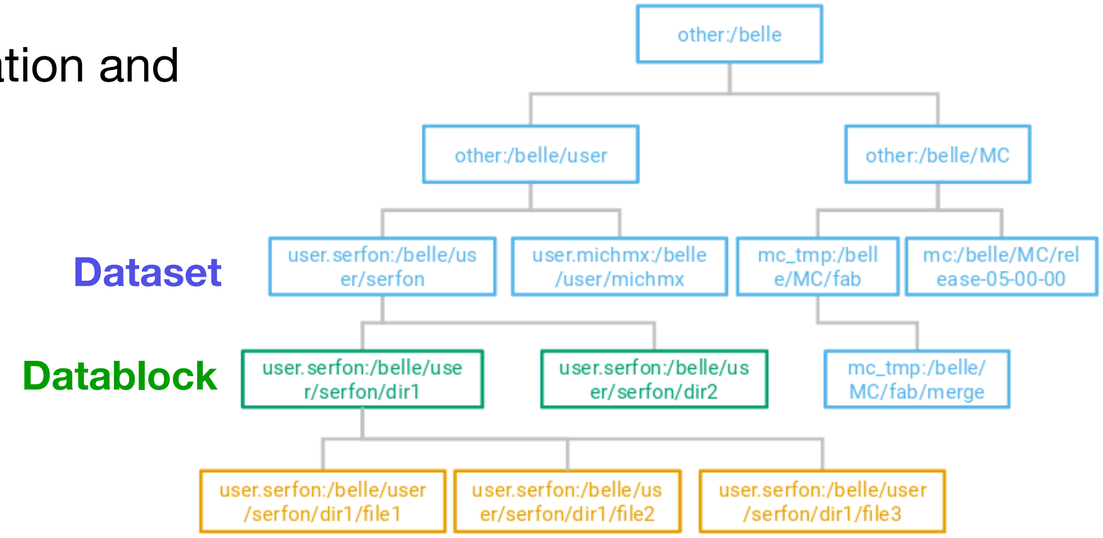
RUCIO
SCIENTIFIC DATA MANAGEMENT

Belle II Computing

- Belle II uses DIRAC with a specific extension called BelleDIRAC
- Rucio:
  - For the data management part as service
  - As a catalog that is used by BelleDIRAC. Contains the full namespace of Belle II data.



- Belle II uses a hierarchical namespace
  - Files are at the deepest level.
  - Datablocks contain files and are the unit of replication and processing
  - Datasets are an aggregation of datablocks



Metadata in Belle II

- As mention earlier, Belle II uses Hierarchical namespace.
- Metadata are stored in AMGA.
- Belle II uses different metadata depending of the level in the namespace hierarchy
  - Files : Number of events, site where the file was produced, etc.
  - Datablock : Number of file in the datablock, creationDate, etc.
  - Dataset : beamEnergy, dataLevel, productionId, etc.
- Metadata can also be classified by their use cases .
  - Use for processing : status, nEvents, checksum, etc.
  - Use for monitoring/accounting : size, dataLevel, etc.
  - Use for traceability : steeringFile, productionId, etc.
- A metadata service should be able to support these different type of metadata and use-cases.

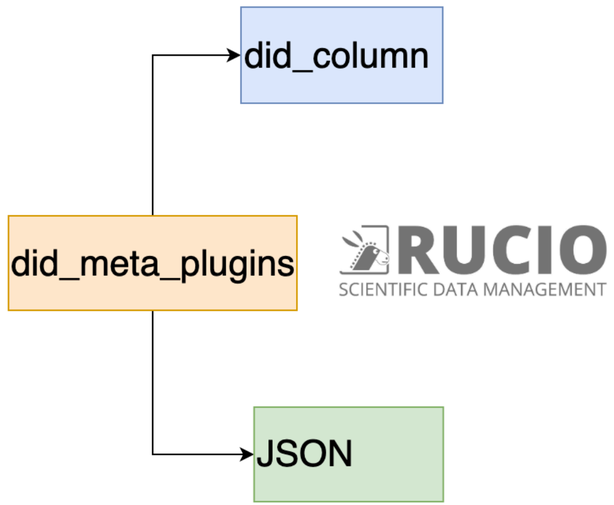
Metadata in Rucio

- Rucio is a Data Management advanced tool that provides many Data Management advanced features

- Rucio can be used as a (hierarchical) file catalog to register the namespace and the metadata associated to all the component of the namespace

- Different type of metadata are supported by Rucio .
  - Fixed set of metadata stored as column of specific type used in the main table in Rucio database (aka column metadata)
  - Any type of key:value pair stored using the json type in Rucio database (aka json metadata)
  - External metadata services (not considered here)

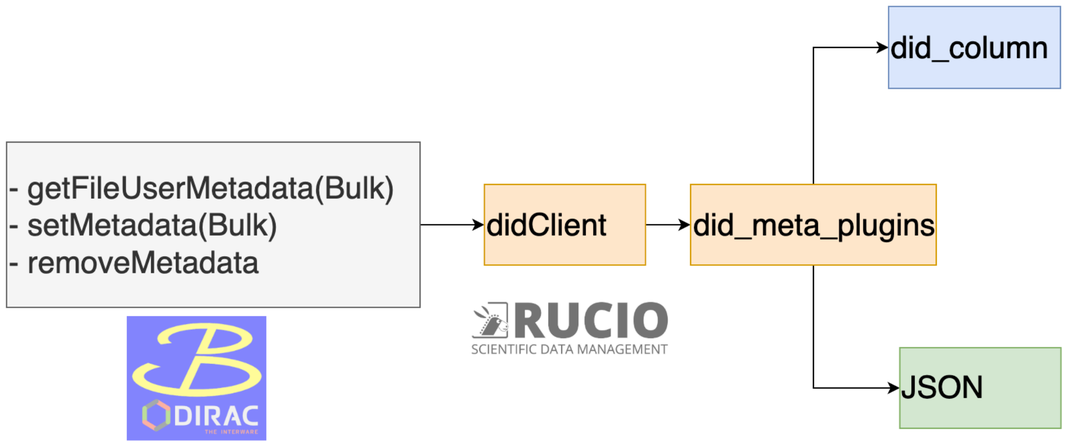
- **For Belle2 we chose to store:**
  - **Metadata used for accounting in the column metadata.**
  - **all the rest are stored in json metadata.**



Metadata Related Development

- File or the directories in the namespace hierarchy can inherit the metadata of the parent.
  [get_metadata_bulk(dids, inherit=False)]

- Provide new bulk methods in Rucio to register metadata.
  [set_dids_metadata_bulk]



- Implement methods in RucioFileCatalog interface in BelleDIRAC similar to the metadata methods of the DIRAC File Catalogue
  - getFileUserMetadata (getFileUserMetadataBulk)
  - setMetadata (setMetadataBulk)
  - removeMetadata

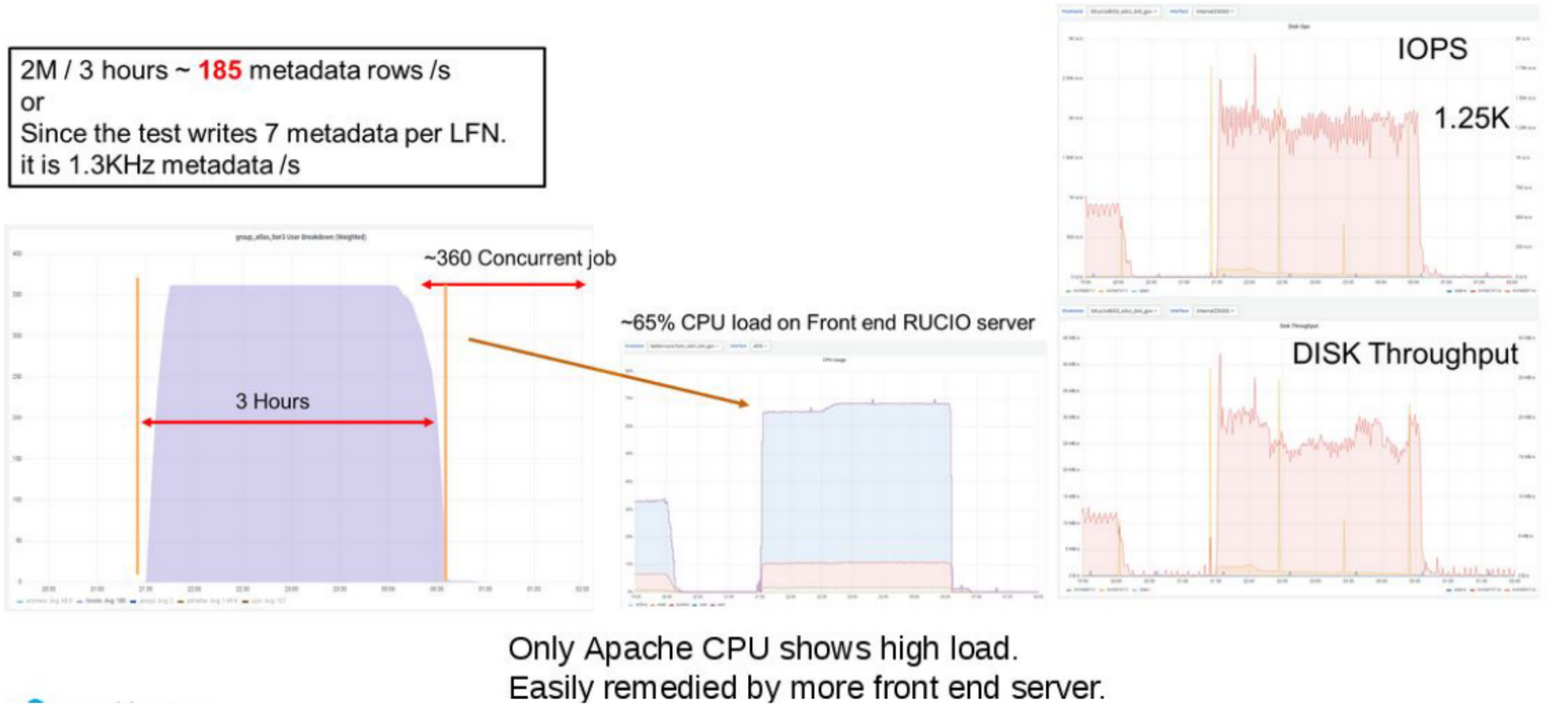
- Changes Done in Pilot and Production Client to report metadata to Rucio (see latter slide).

- Introduce new tool to create accounting summaries based on metadata and populate influxDB

- Adapt the end-user tools to allow using Rucio as metadata backend

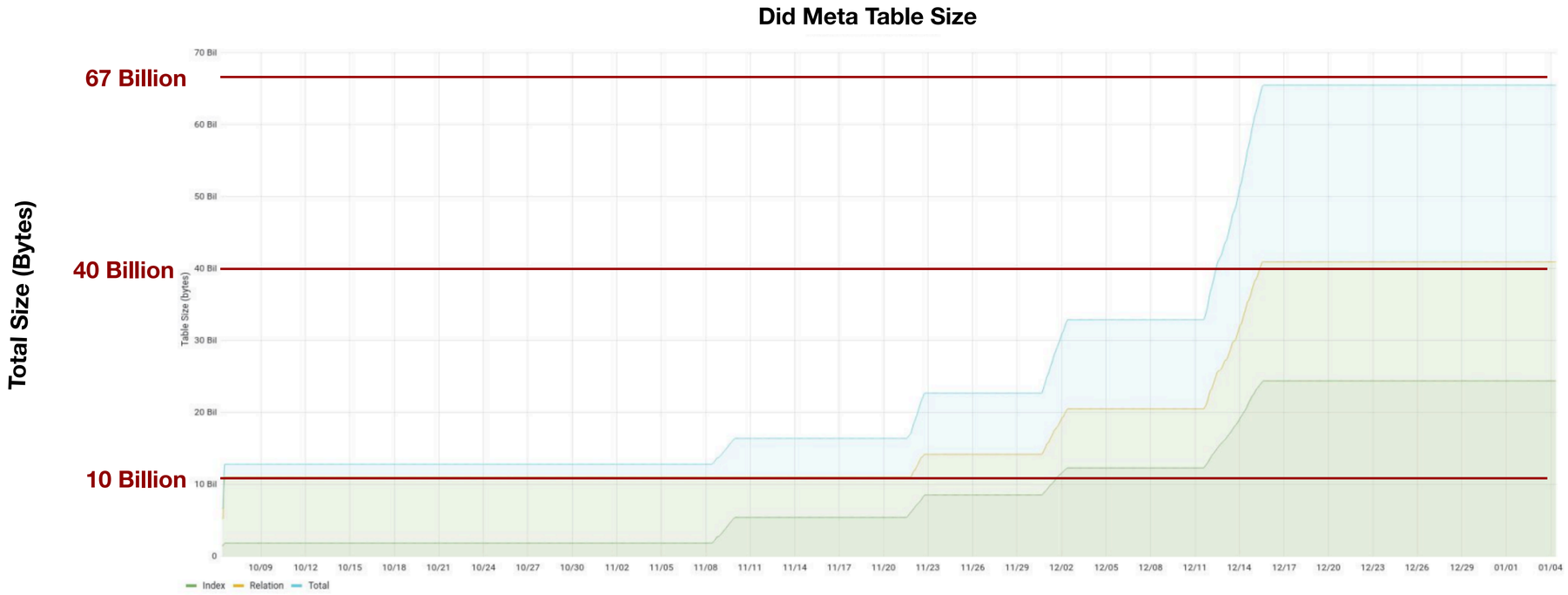
Metadata tests

- A stress test was conducted using a snapshot of Belle II production instance (~100M files imported) and deployed on a test instance :
  - Similar DB backend as the production one, but only one Apache front-end
  - Test is querying a list of files in Rucio and set a few metadata for each file
  - Multiple tests are run in parallel using a batch system to increase the load on Rucio (up to 360 jobs)
- **No bottleneck observed on the DB side.**
- **Limitation comes from the single front-end used for the test, but can be scaled horizontally**



Metadata Import

- Metadata were then gradually imported to the production instance of Rucio in a background mode. (i.e. no service downtime required).
  - First import of “accounting” metadata
  - Then import of generic metadata
- No issue was observed during or after the metadata import
- Space occupied on the database by table and indices scales linearly with the number files registered in Rucio
  - 1kB/file
  - Allows to provision DB hardware for coming years.



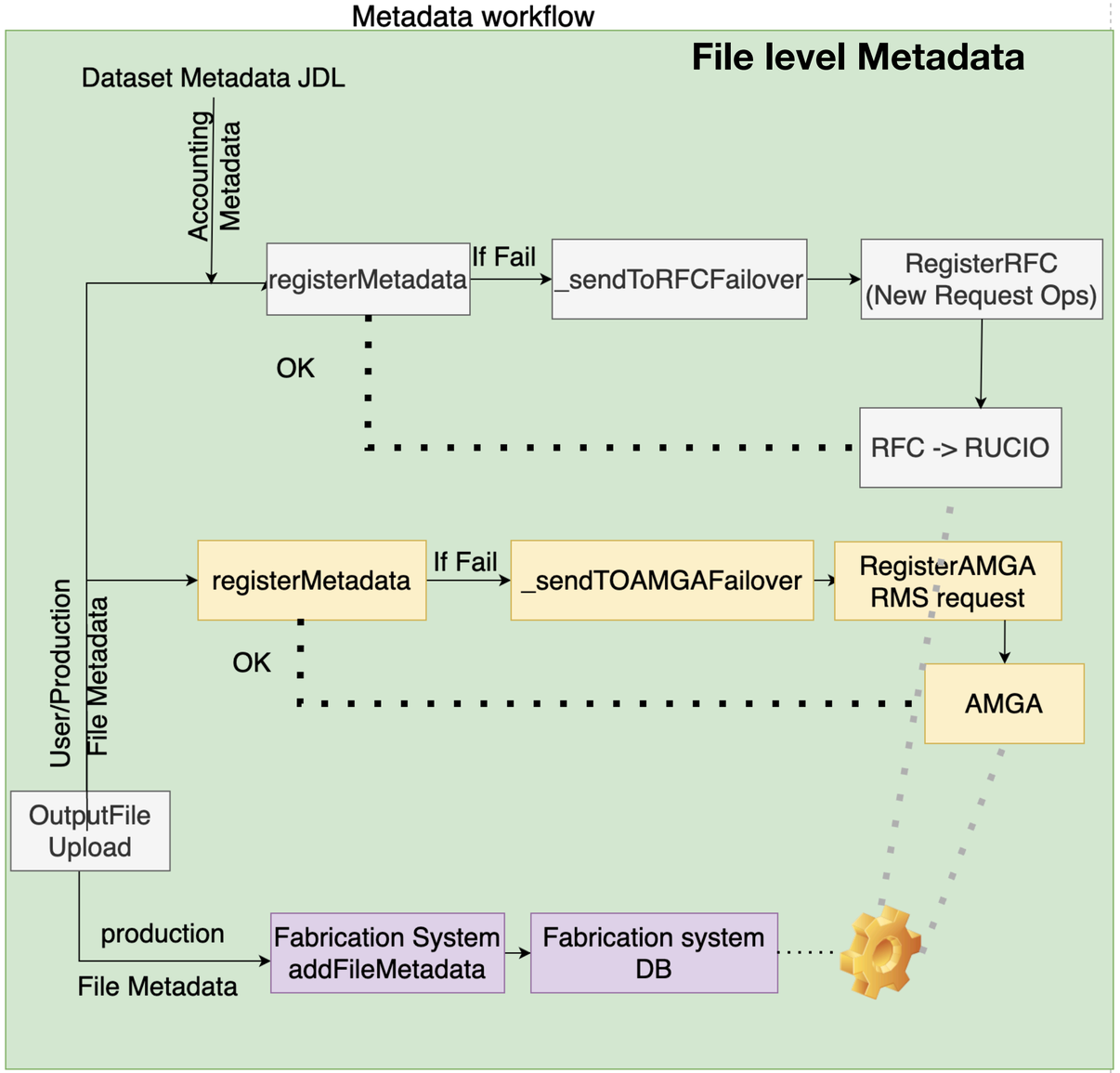
Metadata registration workflow:

File level Metadata

- File metadata registration is done from pilot jobs.
- Follows the same workflow as AMGA registration.
- New request operation in DIRAC for Rucio metadata registration.
- Accounting metadata are registered at this point to file metadata.

DataBlock/Dataset Metadata

- Dataset and Datablock Metadata are registered via a sub-system in BelleDIRAC for AMGA.
- We follow the same procedure of Dataset metadata in Rucio.
- Datablock metadata is registered from pilot in Rucio.



Metadata Service workflow:

- Choice of service to use for metadata registration is configurable.
- Configuration parameter are set in DIRAC configuration system.
- Choices:
  A. Only AMGA (Rucio off).
  B. AMGA and Rucio . (Request operation is done by AMGA)

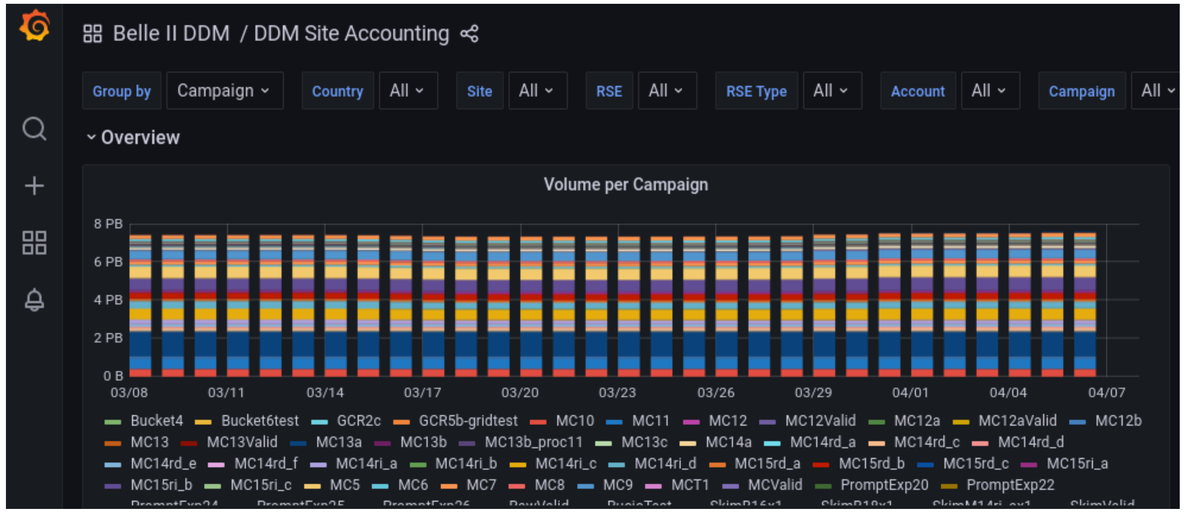
     - Results in some inconsistency and will be handled by manual check machinery.

  C. Rucio Only. (Turn on registerRucio request operation.)

We are going with option B in our initial Phase.

Benefit of Rucio Metadata

- Have accurate accounting based on the file metadata (already in production)
- Same API from RucioFileCatalogClient for metadata.
- Single system to maintain. (If only Rucio is used)
- Scalability of Rucio is already proven and can handle high luminosity era of Belle II.
- Migration from AMGA to Rucio is shown.
  (Gradual Migration with both and then only Rucio possible.)
- Scalability of user project submission.



Conclusion:

- We presented here the work done to integrate Rucio metadata into Belle II's computing framework:
  - New developments allow to cover all Belle2 workflows already supported by the current metadata service
  - The different tests performed show that Rucio is able to handle Belle II's need

- Workflow for Rucio metadata at Belle II is developed and tested.

- Rucio as a additional metadata service is expected to happen in the coming days.