Improvement in User Experience with RUCIO Integration at the Belle II

Panta, Anil (University of Mississippi)
Serfon, Cedric (BrookHaven National Laboratory)
Hernandez Villanueva, Michel (DESY)
Miyake, Hideki ; Ueda, Ikuo (KEK/IPNS)

CHEP, 2023
Belle II Experiment

- Asymmetric lepton collider at Tsukuba, Japan.
- Upgrade from previous Belle experiment.
  - Target Integrated Luminosity: 50 /ab
    (x 50 more than previous B-Factories.)
- Main Physics Goals:
  - Source of CP violation.
  - High precision measurement.
  - Many more.

Global Collaboration
Belle II Computing:

- Grid Computing model
  - i.e. Distributed Computing Architecture.
- Raw Data:
  - One copy at KEK
  - Second copy in Multiple center DESY, BNL, KIT, IN2P3CC, CNAF
- Data reprocessing, MC production and User Analysis
  - All done on the grid.
Computing Architecture

- (Belle) **DIRAC** as
  - WorkLoad Management (WMS)
  - Production Management (PMS)
  - Request Management (RMS)
  - Data Management (DMS)
- **Rucio** as Distributed Data Management (DDM)
- Rucio as File Catalog (RFC)
- AMGA as metadata Catalog
- CVMFS as software distribution service
- VOMS as Virtual Organizations and attribute-based authorization
**User Analysis Pre-workflow at Belle II**

- Belle II uses gbasf2 for grid-based user analysis.
  - Input for jobs: path or list of grid path (LFN/LPN).
- Users search for input datasets via dataset-searcher (metadata-based).
  - Dataset-searcher is an extension system that stores paths and corresponding metadata.
- The list of datasets can be very large (order of 10K)
  - changes frequently due to data reproduction, configuration errors, etc.
- **It is difficult to keep track of large numbers of files and errors can occur.**
- Most of the time, users run over the same list of datasets (reprocessed data).
- Pre-analysis/pre-job submission can be time-consuming due to dataset management.
Collection

- Collection comes from container concept in Rucio.
  - Belle II itself has hierarchical namespace.
  - Rucio by default provides non-hierarchical name space.
  - We can have orthogonal namespace to current Belle II namespace.
  - Container is just a collection of dataset (in rucio term) or container itself.

- Single path for collection of dataset of interest
  `/belle/collection/MC(Data)/<collection_name>`

- Collection is centrally defined by Data-Production team.

- Advantages:
  - Single path for analysis -> intuitive for user.
  - Collection is Immutable -> analysis reproducibility is ensured.
  - Decrease in pre-work for gbasf2 job submission.
  - Extra info on luminosity and description is provided, so no need for user to look elsewhere.
  - Huge decrease in gbasf2 analysis jobs submission time.
    (Order of 10x decrease for 8K files analysis project)
Collection Management/User Tools

- Command line tools are provided for user to get information on collection.
  - `gb2_ds_collection --list_all_collection`
  - `gb2_ds_collection --get_metadata`
    ```
    -> gb2_ds_search collection --get_metadata /belle/collection/test/MC14ri_cubar_labinv_v1
    # Metadata of Collection MC14ri_cubar_labinv_v1
    dataLevel: mdst
    description: Collection MC14ri_cubar for cubar - 4S
    campaign: MC14ri_d,MC14ri_a
    dataType: mc
    skimDecayMode:
    int_luminosity: 1000.0 /fb
    generalSkimName: 
    # Extra metadata info
    ```
  - `gb2_ds_collection --list_datasets`

- Collection management tools for C(reate), U(update), D(lete) operation.
  - `gb2_ds_collection create`: create a collection in Rucio via list of datasets path and add metadata of collection.
  - `gb2_ds_collection delete/update`: only be able update extra metadata info

- Future work (In dev):
  - Tools to search collection by its metadata.
Rucio Download: MultiThread Download.

- We use DIRAC Data Manager for download with synchronous single-thread download.
- **Rucio also provides a download client with built-in multithreading.**
- Three additions/contributions made in Rucio to satisfy our needs:
  - Option not to raise exception in DownloadClient.
  - Option to validate files by file size in DownloadClient.
  - Handling of hierarchical namespace with '/' is done.
- **We provide download using Rucio as an extra option, keeping DIRAC download as default.**
- This has resulted in happy user as download of user analysis output is fast.

```bash
(base) [watts@xenox validation]$ go3_do_get xi_true_xim --new -I download.txt
Do you want to download 3 files:
Please type [Y] or [N]: Y
2023-04-13 18:43:48.922 INFO Processing 3 item(s) for input
2023-04-13 18:43:48.922 INFO No preferred protocol impl in rucio.cfg: No section: 'download'
2023-04-13 18:43:48.922 INFO No preferred protocol impl in rucio.cfg: No section: 'download'
2023-04-13 18:43:48.922 INFO No preferred protocol impl in rucio.cfg: No section: 'download'
2023-04-13 18:43:48.922 INFO Using 3 threads to download 3 files
2023-04-13 18:43:48.922 INFO Thread 0/3: Preparing download of user.anl123/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00000_job679086_00_root
2023-04-13 18:43:48.922 INFO Thread 1/3: Preparing download of user.anl123/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00002_job679086_00_root
2023-04-13 18:43:48.922 INFO Thread 2/3: Preparing download of user.anl123/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00004_job679086_00_root
2023-04-13 18:43:49.712 INFO Thread 0/3: Trying to download with size and timestamp of 360s from CMM-TMP-G: user.anl123/belle/user/anl1123/xi_true_xim
2023-04-13 18:43:49.712 INFO Thread 1/3: Trying to download with size and timestamp of 360s from CMM-TMP-G: user.anl123/belle/user/anl1123/xi_true_xim
2023-04-13 18:43:49.712 INFO Thread 2/3: Trying to download with size and timestamp of 360s from CMM-TMP-G: user.anl123/belle/user/anl1123/xi_true_xim
2023-04-13 18:43:49.712 INFO Thread 0/3: Trying to download with size and timestamp of 360s from CMM-TMP-G: user.anl123/belle/user/anl1123/xi_true_xim
2023-04-13 18:43:49.712 INFO Thread 1/3: Trying to download with size and timestamp of 360s from CMM-TMP-G: user.anl123/belle/user/anl1123/xi_true_xim
2023-04-13 18:43:49.712 INFO Thread 2/3: Trying to download with size and timestamp of 360s from CMM-TMP-G: user.anl123/belle/user/anl1123/xi_true_xim
2023-04-13 18:43:49.800 INFO Thread 0/3: Using PPN: davs://test-archive-arch.archinfo.in.r1/14403/embassy/belle/tmp/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00000_job679086_00_root
2023-04-13 18:43:49.800 INFO Thread 1/3: Using PPN: davs://test-archive-arch.archinfo.in.r1/14403/embassy/belle/tmp/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00002_job679086_00_root
2023-04-13 18:43:49.800 INFO Thread 2/3: Using PPN: davs://test-archive-arch.archinfo.in.r1/14403/embassy/belle/tmp/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00004_job679086_00_root
2023-04-13 18:43:57.223 INFO Thread 0/3: File user.anl123/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00000_job679086_00_root successfully done
2023-04-13 18:43:57.223 INFO Thread 1/3: File user.anl123/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00002_job679086_00_root successfully done
2023-04-13 18:43:57.223 INFO Thread 2/3: File user.anl123/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00004_job679086_00_root successfully done
2023-04-13 18:43:58.490 INFO Thread 0/3: File user.anl123/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00000_job679086_00_root successfully done
2023-04-13 18:44:00.842 INFO Thread 1/3: File user.anl123/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00002_job679086_00_root successfully done
2023-04-13 18:44:00.842 INFO Thread 2/3: File user.anl123/belle/user/anl1123/xi_true_xim/sub00/lambdas2igzpipl_00004_job679086_00_root successfully done
```

# Download Summary

Total Files: 3
Successfully Downloaded Files: 3
Files already found locally: 0
Files that failed to be downloaded: 0
Files with duplicate jobID (not Downloaded): 0

#-----------------------------
Asynchronous Deletion

- Pre-Rucio era:
  - Remove from FileCatalog.
  - Remove each file from Storage Elements.
  - Multiple and synchronous operations

- Deletion happens in Rucio by setting lifetime of rule.
  - Rucio daemon (undertaker) takes care of deletion in background.

- Lifetime of directory is chosen as maximum of child directories lifetimes.

- User can run the command and let Rucio do the deletion without waiting in terminal to complete.
Asynchronous Replication

• Create extra replication of datasets are needed:
  - For User: replicate to closer storage elements (SEs) or specific institution SEs.
  - For operation: replicate user input to another SE in case of issue with some computing elements (CE)/SE.

• Replication in rucio happens by creating replication rules or subscription.
  - Subscription is for new datasets.
  - New Rule for existing datasets.

• User creates a rule using CLI.
  - Checks the status of replication.

• Rucio on backend will do the transfer via FTS
User Space Management using Rucio Quota

• Belle II computing model doesn’t let user to store their files for long.
  - User have to download it in local space for further analysis.
• We have to be make the system automated to handle user space in Grid.
• Rucio provides quota system.

• For each RSE an account can have:
  - 0 bytes quota → No rules can be created by the account on this RSE.
  - ∞ bytes quota → As many rules as possible (Until the RSE is full) can be created.
  - N bytes quota → Rules until the accumulated amount of N bytes can be created.

• Few development has been done on BelleDIRAC side to enable quota.
  - All files should be owned by user to get proper accounting.
  - Addressing problem where the file can successfully be uploaded to the Storage Element but the registration to Rucio fails because of exhausted quota
  - No job submission if quota is filled. (Avoids unnecessary CPU use).
  - CLI for checking quota.

• It will be activated soon, ensuring efficient use of space and resources at Belle II.
Trace and Data popularity

- Rucio provides the functionality of Data popularity via trace.
- Trace is just informations when a physical file is accessed.
- Rucio will provide last accessed date and access count of file.
  - Helps in making Data management decisions.
  - Clear picture of the file usages from different category.
  - Trace info is collected and sent from Pilot.
    - BelleDIRAC: InputDataDownload (plugin of DIRAC)
    - Infrastructure is ready and tested.
    - Development for trace collection and sending is done.
    - Tested for User Jobs.
    - Finishing the development for production jobs.
- Trace will be turned on from next BelleDIRAC release.
Conclusion

- Since the integration of Rucio at Belle II, we are continuously exploiting rucio features in workflow.
- Rucio has shown significant improvement in user experience of using grid at Belle II.
- We are still in process of adding more Rucio features into our workflow.
- Many thanks to Rucio and DIRAC team for the continuous and great support.