



# Belle II report

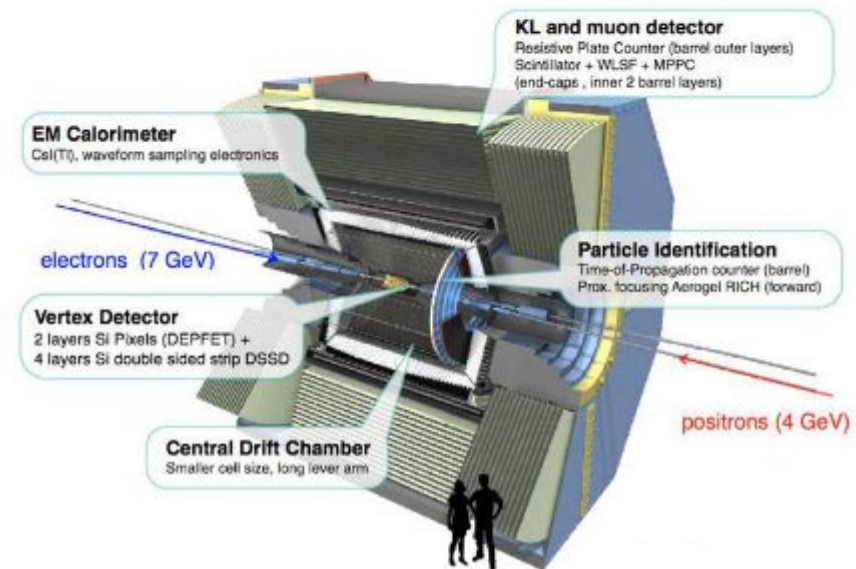
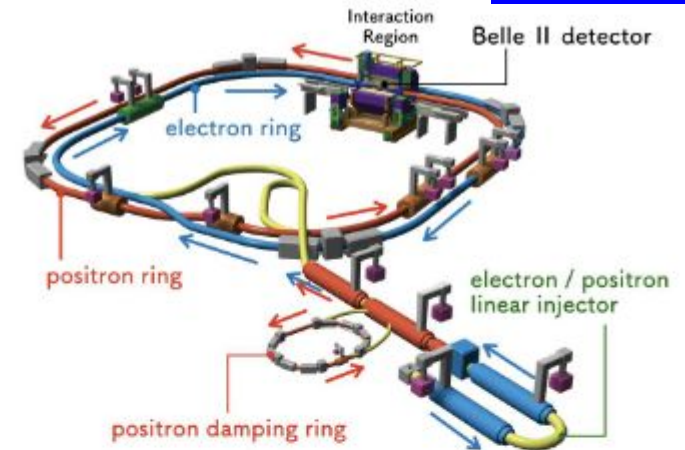
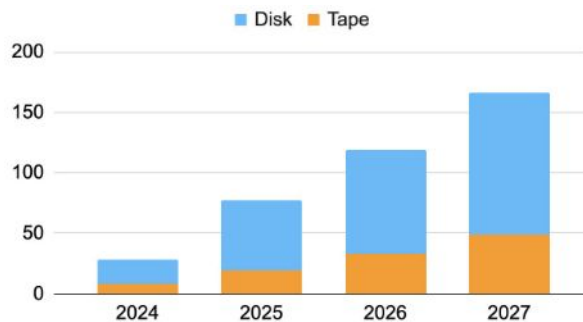
Cédric Serfon (BNL) on behalf of the Belle II Computing Team

October 1, 2024



# Introduction

- Asymmetric lepton collider.
- Upgrade from previous Belle experiment.
  - $50 \text{ ab}^{-1}$  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.



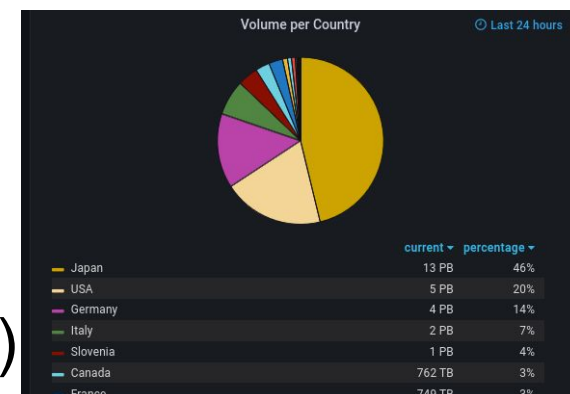
# Rucio in Belle II

- **Statistics**
  - **DIDs :**
    - Containers : 23.3M (high number due to hierarchical namespace, c.f. next slide)
    - Datasets : 9.5M
    - Files : 73.6M
  - **Rules :**
    - 13.5M (85% generated by subscriptions)
  - **Replicas :**
    - 110M (i.e. replication factor : 1.5)
    - 27.7 TB (i.e. 252 MB average file size)
- **Infrastructure running on VMs :**
  - 2 Rucio servers
  - 2 Rucio daemons
  - 1 WebUI
  - 2 tracers
- **Database : PostgreSQL (see Hiro's talk)**

Subscriptions

Show 100 entries Search:

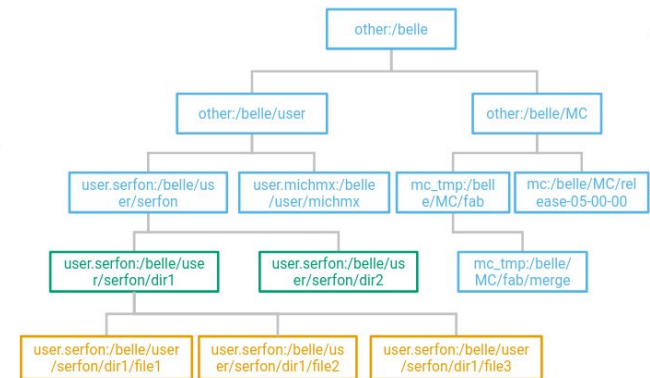
Name	Ok	Replicating	Stuck	Suspended
beambg_distribution_all_data	413	0	0	0
beambg_distribution_primary_data	668	0	0	2
BG_beambg_distribution_all_data	23114	14	0	13
BG_beambg_distribution_primary_data	23303	0	0	0
BG_production_intermediate	1176	0	0	0
Chained export cDST to KEK	45218	0	0	764
Chained export to RAW DC	9638	0	0	0
Data cDSTs final to calibration storage at KEK	0	0	0	0
Data production final to all DATA-SE	60042	2	0	4
Data production final to DATA primarySE	65691	0	0	1
Data production intermediate	159272	10	18	2
Data Raw chained export to RAW DC	15146	0	0	1
Data Raw export to KEK RAW-SE	2540	1	0	2
Data skim final to all DATA-SE	1092011	0	0	49
Data skim final to DATA primarySE	1093255	0	0	40





# Belle II specificities

- Belle II uses a hierarchical namespace :
  - All DIDs have a parent, except the root (other:/belle)
  - The creation of this hierarchy is enforced by a specific method associated the “dirac” REST endpoint.
    - Strictly speaking this is not linked to dirac and could be used by communities using hierarchical namespace
  - The fact that containers cannot contain files impose some restriction
- Rucio CLI not directly exposed to the end-users :
  - All methods to create rules, download/upload files are part of (Belle)Dirac that provides the interface to Rucio → No impact expected in case of CLI redesign (move to SOV structure)



# Current status

- Use a pre-historic version of Rucio (1.28.7) for various reasons :
  - Latest python2 compatible version and late migration of our last component (pilot) to python3
  - Focus on getting metadata to Rucio first
  - Now difficult jump from 1.28.7 to latest version expected. Need careful tests and validation. Will probably wait for winter shutdown
- Token status, no tests of TPC with tokens yet
  - IAM instance not validated for production
  - Need to upgrade to a newer Rucio version supporting token submission



# Current status

- Participation to DC24

- Belle II took part to DC24 organized by WLCG
- Goal was to test that ability to sustain the transfer rate expected at the end of the decade from KEK to our RAW Data Center (40 TB/day)



- Successfully reached the target (up to 5 time the expected throughput)
- No bottlenecks observed during the test (but no tokens...)

# Using Rucio as metadata service

- Belle II has been using a metadata service called AMGA for many years
  - AMGA was developed at the same time as LFC by LCG/EGEE (see [CHEP2006 paper](#))
  - AFAIK, Belle II has been the only user of AMGA since many years
  - AMGA is not able to handle recent changes in computing (OS changes, move to tokens, etc.)
- Decision in January 2024 to move to use Rucio as metadata service
  - Use the json metadata feature of Rucio
  - Decision made after running series of test to demonstrate that it will scale for Belle II (see [CHEP2023 paper](#))

## Storing JSON in a SQL Database



# Future plans

- Currently use “flat” json with values as string or int to preserve compatibility with AMGA

```
-bash-4.2$ rucio get-metadata /belle/Data/build-light-2401o/DB00003049/bucket31/prod00041893/e0024/4S/r00000/all/18520300/udst/sub00/udst_000027_prod00041893_task230000027.root --plugin JSON
checksum:          cea90872
checksumType:     Adler32
date:             2024-03-02 13:31:18
eventHigh:       78681509
eventLow:        92
experimentHigh:  24
experimentLow:   24
jobId:           394866591
lfn:             /belle/Data/build-light-2401o/DB00003049/bucket31/prod00041893/e0024/4S/r00000/all/18520300/udst/sub00/udst_000027_prod00041893_task230000027.root
nEvents:         700000
parentGuids:     13d4ce01-d897-f3b8-dcb8-95da4478c1f5,33dab5f8-9e9b-1475-aae3-896f9c3162d7,e4d1de1c-5ebc-cb87-6eb6-2cafd2cb6c01,7a48ac0a-4173-279d-75cc-08afa11b57e8,11ece644-2e25-6cd5-765a-3d815b1498f0,624f16fb-9019-b756-7d83-55fbbb0c95f3,08cb8252-2fc4-2f4b-8f34-8b412fec798e,0297dcf7-f1e8-e4c1-3e8e-77fabd0a88ed,59a34435-714b-5741-30d9-856e86328335,615db85a-3c93-7a02-8a31-c559306e855b,a4ca5bc1-fd19-d5ad-188e-438d4f961838,8aa47bcc-ceb9-2eb2-e84e-678586aac390,6629dea4-1ac2-4ed6-fdd7-3db75d0e40a2,c5ccb920-6bb9-45f3-272e-9175fcb98609,fc6b8bde-4392-8a2a-3d6c-66716fa927c5,a7a93e32-a6a2-a97b-5503-8927bd818959,2a7e6a9c-a899-41d9-d88d-7b5ae068f8a1,bd131d56-13f4-afa3-301a-566bdf3a3270,474f8183-cd53-3242-c3a3-a9fdc192ba25,fe41602e-9b9a-73a5-f968-c6c1610cc3e1,a591cd0c-cfe6-066f-f83a-0d03876fab0d,cd41d75d-f17a-b3ef-3da5-a7dca8c1631a,bc8917bc-963a-3e2a-d29b-764a701ed4bf,6a76c7f8-9b9f-4edc-3650-afc1f6ebe529,ae9c936f-353d-eb8f-b573-7be11fd4fc5f,185449bb-eb4d-b1cd-9fe6-d6c6b23cf94a,31976a5d-6897-6f49-8f7e-f38ef2f2dad5,3ba2ce48-47dc-0bfe-f69e-a06bdbbffb3a,a9de11c6-b294-8537-0489-f205cafc122e,5da7ced0-7ea5-6941-9c0e-0021009242c
runHigh:         1447
runLow:         1148
site:           LCG.KEK2.jp
size:           1760542864
status:         good
```

- Didn't check the possibility to use array or nested objects



# Using Rucio as metadata service

- Since March 2024, all the metadata are registered both in Rucio and AMGA to prepare the transition
- Historical data (prior 2024) import :
  - Fully imported for files (65M) and containers (8M) which are the most important ones
  - Datasets metadata importation is being finalized (about 7M)
- Belle II software changed to use Rucio as primary metadata service with possibility to fallback to AMGA
  - Switches in DIRAC Configuration Service to enable/disable writing/fallback to AMGA
  - Code still under test but should go into production soon
  - If everything runs smoothly, AMGA will be turned off after some time

# Benefits of using Rucio metadata

- No initialization when new container/dataset is created, metadata can be created on-the-fly
- Reduce the number of critical services
- Scalability : Tests show the write/read supported by Rucio can scale up-to 1-2 orders of magnitude higher than Belle II's needs
- Possible improvement :
  - Metadata registration at the creation (already supported by did resources, but not dirac)
  - Index some metadata to try metadata queries

# Future plans

- Turn off AMGA, i.e. Rucio will become the official metadata service of Belle II (hopefully before EOY)
- Move to latest Rucio
- Start testing tokens
- Getting rid of SRM :
  - All DISK already have WebDAV enabled and all transfers done via WebDAV (staging still use SRM with WebDAV as transfer protocol)
  - Start testing/enabling into production Tape REST API

# Conclusion

- Still happy with Rucio
- DC24 allowed to test and validate Rucio transfer functionality up to the throughput expected at the end of the decade
- Rucio will soon be the official metadata service from Belle II. It should bring benefits wrt current metadata service
- Slowly catching up with the latest developments (tokens, tape REST API, etc.) and latest Rucio release