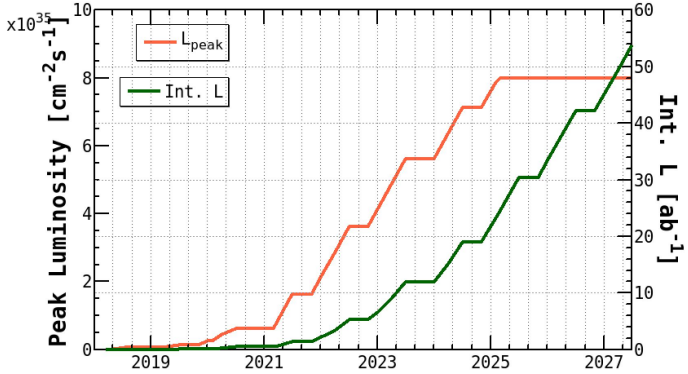
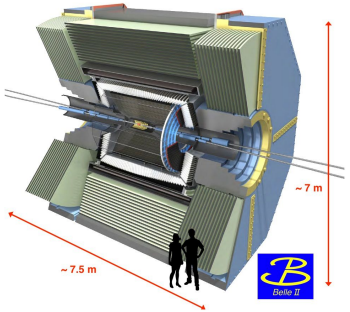
Belle II data management with Rucio

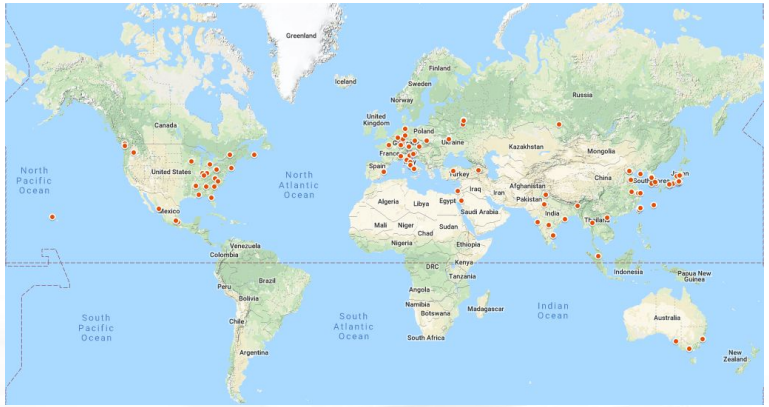
J. De Stefano, H. Ito, P. Laycock, R. Mashinistov, C. Serfon
Brookhaven National Laboratory
H. Miyake, I. Ueda
KEK
Y. Kato
KMI
M. Villanueva
University of Mississippi

Introduction

- Belle II is a B-physics experiment located at KEK (Japan)
- Start of data taking last year
- 1000 members from 26 countries
- > 60 PB data expected by 2023 (disk + tape)



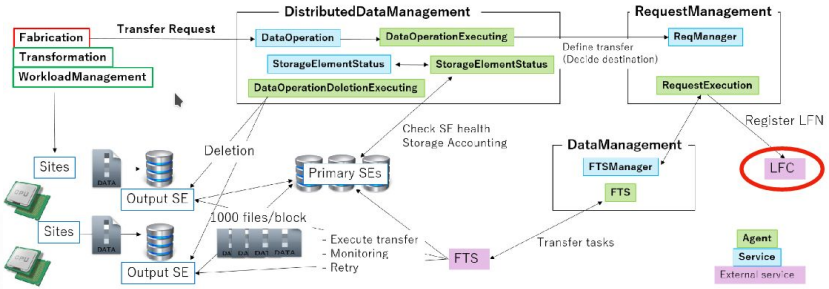




U.S. DEPARTMENT OF ENERGY | KEK Inter-University Research Institute Corporation High Energy Accelerator Research Organization | KMI Kobayashi-Maskawa Institute for the Origin of Particles and the Universe | THE UNIVERSITY of MISSISSIPPI | BROOKHAVEN NATIONAL LABORATORY

Belle II computing

- Belle II uses DIRAC (see Federico's talk) for the Workload and Data Management :
  - Customisation called BelleDirac were done to fit Belle II's needs
- File catalog based on the LCG File Catalog (LFC)
- Current Distributed Data Management (DDM) is part of this BelleDirac :
  - Original design by PNNL group respecting Dirac paradigms, good for Belle II customisation but all development effort must come from Belle II
  - Looking ahead we saw lots of development work, why not use Rucio* instead

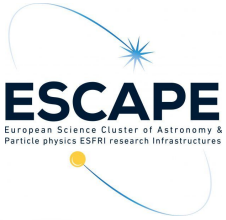


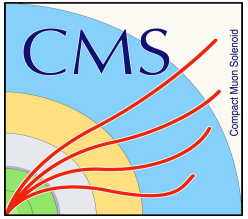
*see [last year talk](#) from M. Barisits

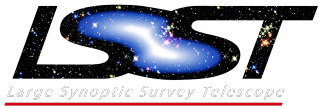
What is Rucio ?

- Rucio is an advanced Distributed Data Management initially developed for the ATLAS experiment
- Its development started in 2012 and it was fully put into production in December 2014, before the start of LHC run 2
- In ATLAS Rucio is replacing a previous DDM system called DQ2 that used LFC as replica catalog. It addresses the issues that were identified in DQ2
  - Scalability
  - Dependency on external services
  - No support of multiple protocols
  - Limited policy replication tools
  - And many more
- Rucio is now evaluated or used by a large community

Rucio community







Kobayashi-Maskawa Institute
for the Origin of Particles and the Universe

How will Rucio fit into Belle II DDM ?

- The move to Rucio should induce no change or very little changes for the applications using DDM or the end-users
- Rucio is not yet used in production by Belle II but it should be during the end of year shutdown
- Rucio will do multiple things :
  - Replaces the file catalog (LFC) used by DIRAC
  - Takes care of interacting with the File Transfers Service (FTS) to move around files and takes care of the physical deletion of the file replicas
  - Enforces the replications policies specified in Belle II computing model
  - Provides tools to monitor efficiently the data movements and to identify the data stored on the sites

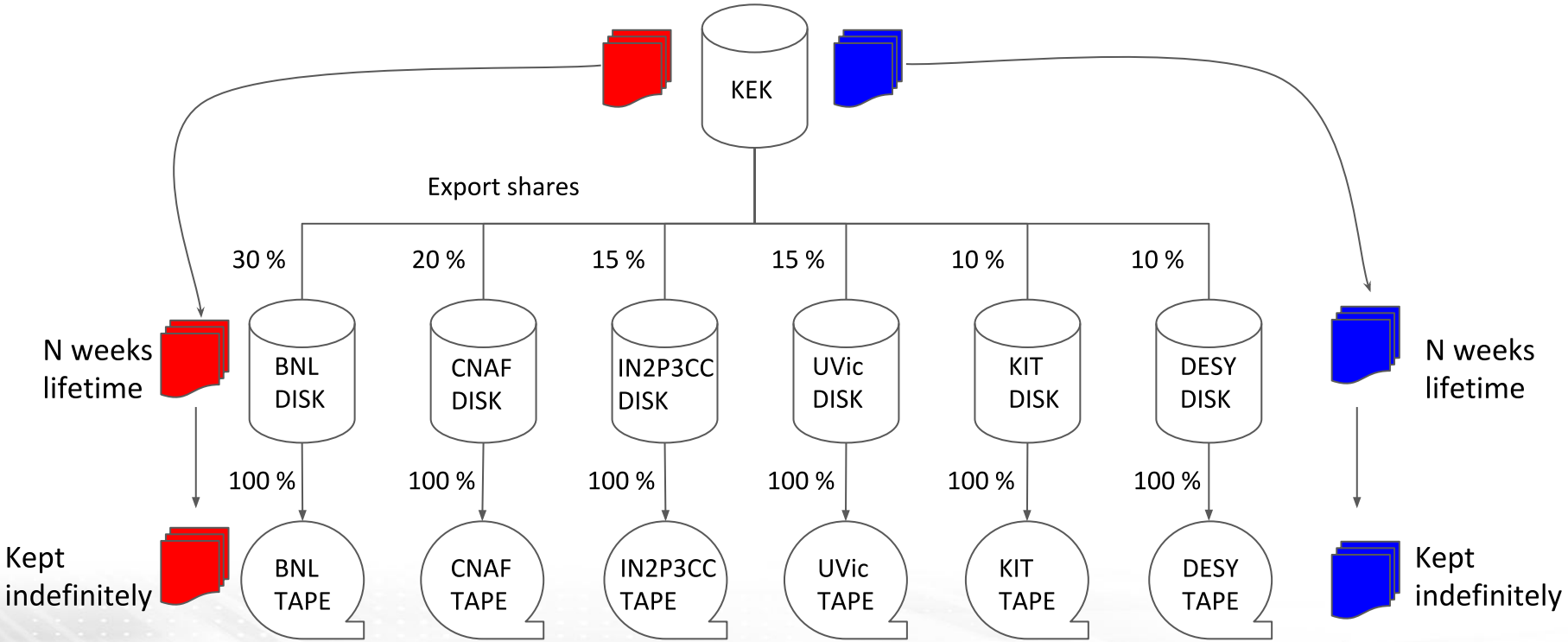
Rucio new development for BelleII

- To fit Belle II needs, a lot of developments were performed. The most important ones are :
  - Change the current DDM API to use Rucio : i.e. the API methods names do not change but Rucio is used behind. This allows the other services interacting with DDM not to change anything.
  - Rucio File Catalog plugin in BelleDirac
  - Chained subscriptions : To transfer datasets from a site A to B then from B to C
  - New daemon in Rucio to submit to external services (InfluxDB, ActiveMQ, ElasticSearch) : Needed to populate the monitoring
  - New dashboards for transfers/deletion monitoring as well as accounting

RucioFileCatalogClient in BelleDirac

- Dirac supports 2 file catalogs LFC and DFC
  - Belle II uses LFC
- In order to move to Rucio, a Rucio File Catalog plugin was developed and integrated in BelleDirac :
  - The new catalog implement all the LFC methods that are used in Belle II
  - The choice of the catalog is done via the DIRAC Configuration
  - Bulk methods on the server side were implemented to reduce number of requests and roundtrips (provide faster response)
- The plugin will eventually be merged into Vanilla DIRAC

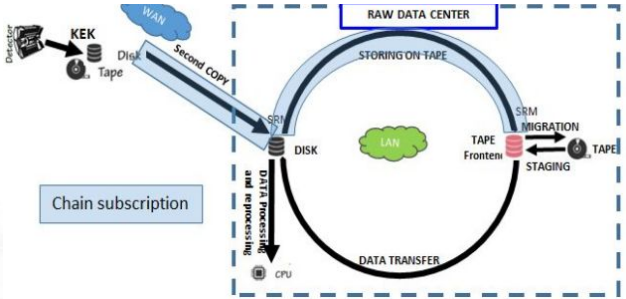
Chained subscriptions



Evaluation of Rucio performances

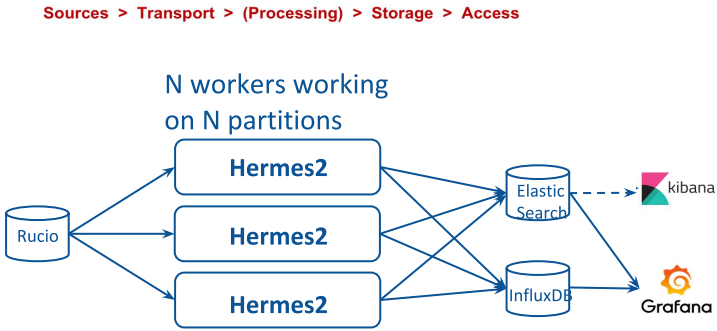
- A huge number of tests were conducted to validate Rucio :

	Goal	Description
Read/Write test at N (N>2) times the current number of files in LFC	Validate the production Rucio instance	Run jobs on different sites reading/writing data to Rucio
Functional tests between all the sites	Identify potential problems at the sites, test transfer/deletion	Insert data at BNL and export to all other sites. Delete when transfered
Export tests from KEK to RAW DC	Validate the subscriptions	Insert data at KEK and export to the RAW DC using subscriptions
All the above	Validate the monitoring	Collect the information from the different transfers (LAN/WAN) and expose it
Test RFC with FS	Validate the RFC implementation	Run some jobs in the FS with RFC



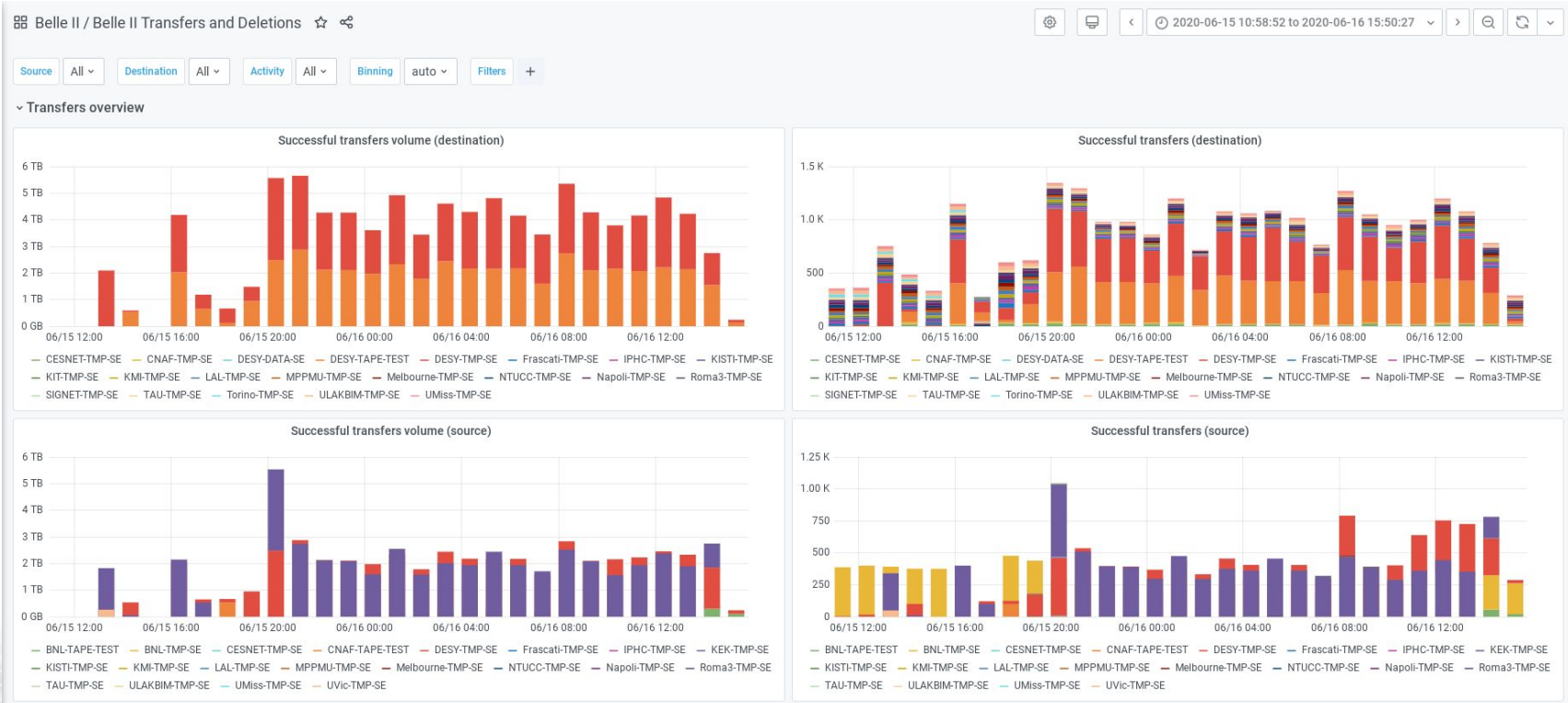
Monitoring

- A simplified monitoring stack wrt the one used by ATLAS has been implemented. It relies on a new rucio daemon called Hermes2 :
  - Rucio daemon that allows to aggregate and send Rucio messages to different services (InfluxDB, ElasticSearch, ActiveMQ, etc.)
  - Horizontally scalable
  - Lightweight : Don’t need a Kafka/Spark cluster
- Easy to deploy for any other collaboration planning to use Rucio
- Grafana dashboards to monitor transfer and deletion as well as space accounting available



Belle II monitoring stack

Monitoring example

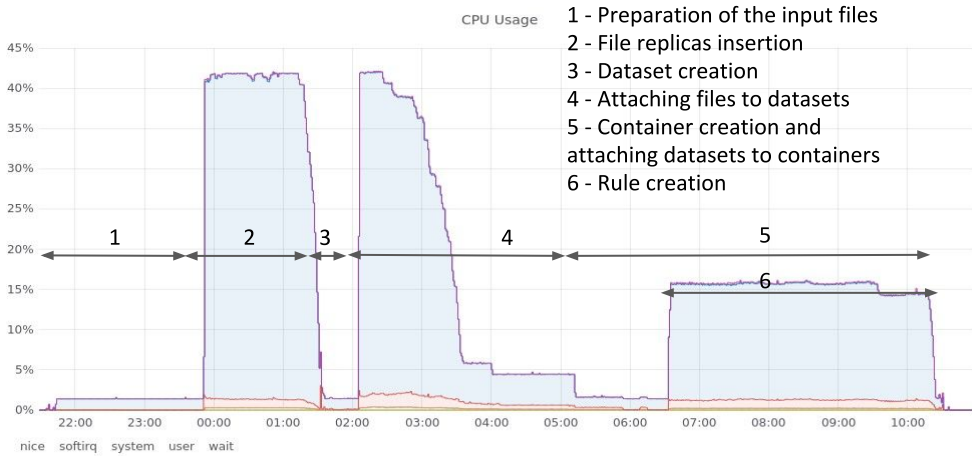


Migration plans

- The migration from the current DDM using LFC as catalog involves the migration of all the files registered in the LFC into Rucio
- To achieve this, the LFC needs to be set to read-only and an import procedure was developed (see next slides for more details)
- The new version of BelleDirac which is Rucio compatible needs to be deployed on the dirac server. Similarly, all the clients need to be updated.
- The whole migration process (including the draining of the grid and the different imports should last about 3 days). A schedule for a migration during the winter shutdown is being discussed

Migration tests

- Migration tools were developed to import the LFC content to Rucio
- Several migration tests were conducted to validate them
- The import of all replicas currently stored in the LFC (~80M files) + creation of the datasets and rules can be achieved in about 14 hours



Load on the migration machine during one full import test

Conclusion

- Rucio was integrated with Belle DDM. All the tests performed so far demonstrated that it works as expected
- Many of the developments performed for Rucio (Rucio File Catalog plugin in DIRAC, chained subscriptions, etc.) can be used by other collaboration
- We hope to switch Rucio to production in the coming months