Experience supporting Belle II CDB server Infrastructure for Phase 3

Carlos Fernando Gamboa, Ruslan Mashinistov, Maxim Potekhin
Brookhaven National Laboratory

Benedikt Hegner, Brookaven National Laboratory
Current affiliation: CERN

Martin Ritter
Ludwig-Maximilians-Universität München

Marko Bracko
University of Maribor and Institut Jozef Stefan

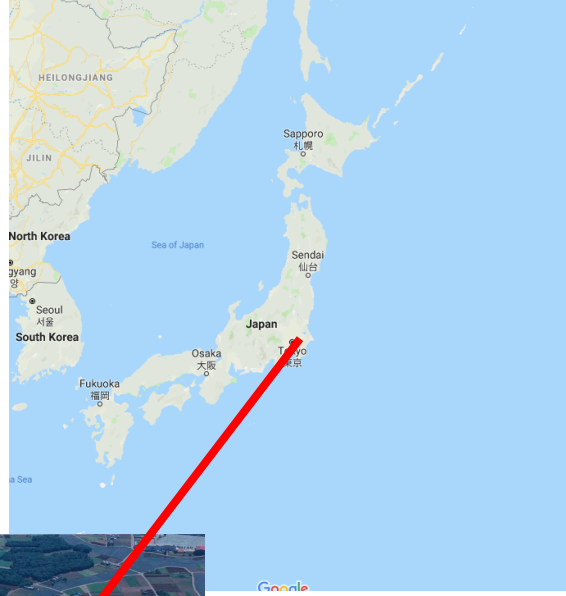
BROOKHAVEN NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

24th International Conference on Computing in High Energy and Nuclear Physics – CHEP 2019

BROOKHAVEN SCIENCE ASSOCIATES

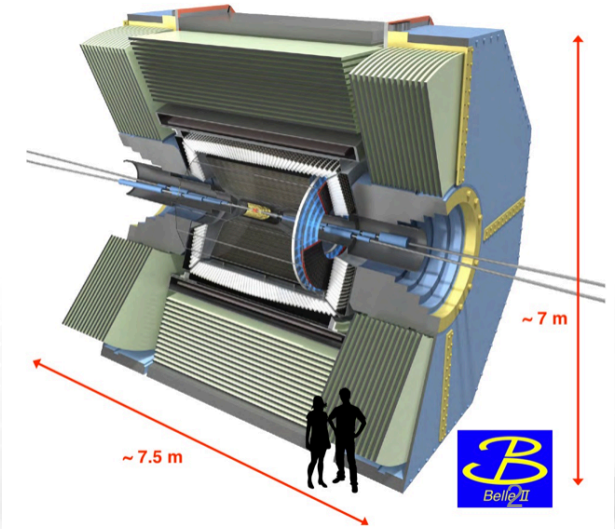
Belle II experiment hosted at KEK Tsukuba, Japan



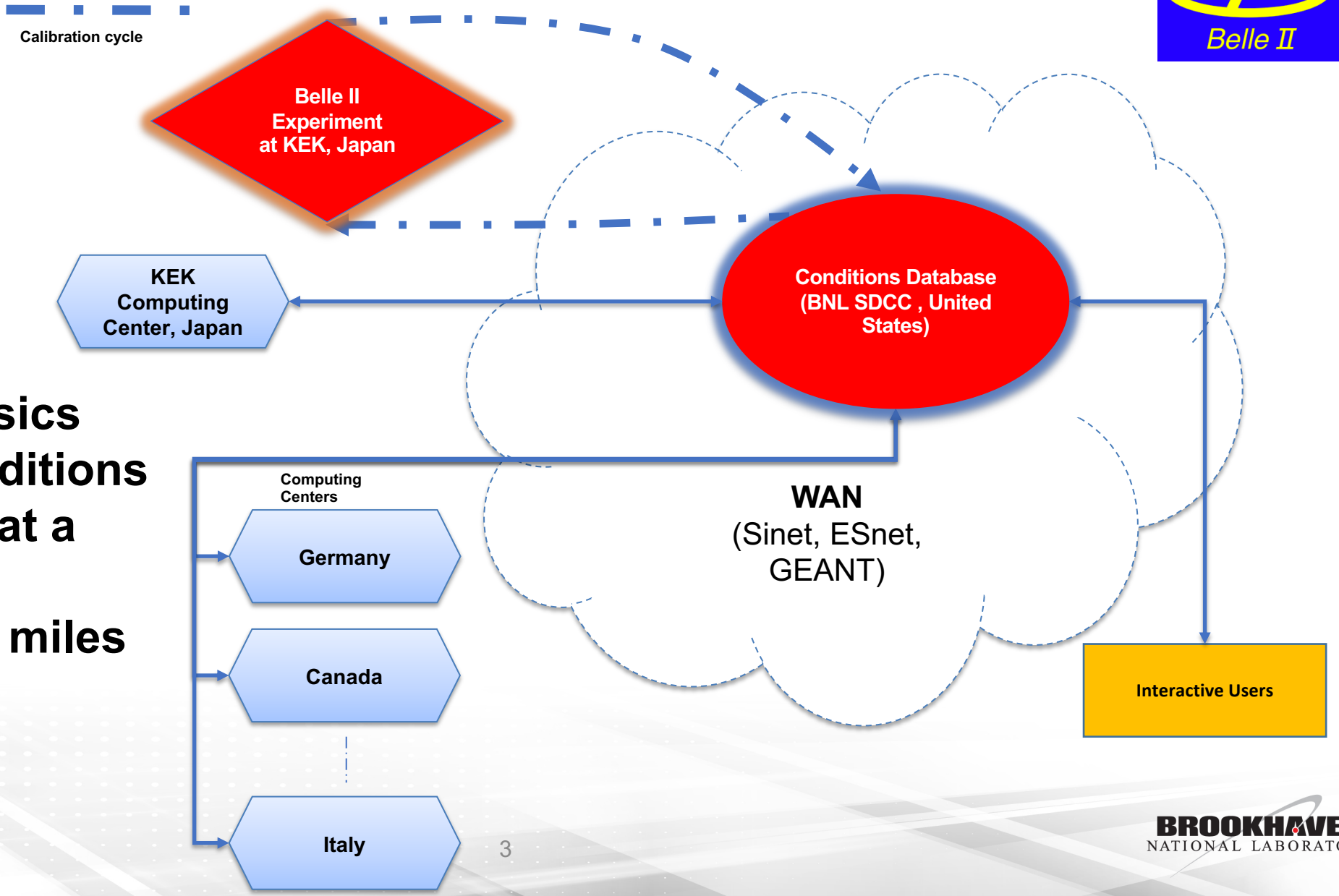
Belle II designed to find physics beyond the Standard Model of particle physics

SuperKEKB accelerator designed to increase the instantaneous luminosity in 40x with respect to KEKB accelerator, and expected to deliver 50x more data to Belle II compared to its predecessor Belle

“Phase 3” - operation with complete Belle II detector: First run from March 2019 to June 2019, operation will continue until 2029



Belle II Conditions Database accessibility overview



First particle physics experiment's conditions database hosted at a remote location, ~10,815 km/6,720 miles apart.

Belle II Conditions Database (CDB)

Events recorded by the Belle II detector are grouped in runs

- A run sets a data taking period with stable operating parameters
- The database manages conditions data on run granularity

Conditions data is composed by metadata and files, stored for persistency in a:

- Shared filesystem → **files or payloads**

  Payload binary objects containing information like (calibration, Beam Parameters, etc.)
- Relational Database Management System (RDBMS) → **Metadata**

  Allows identify payloads which are accessible in an external service

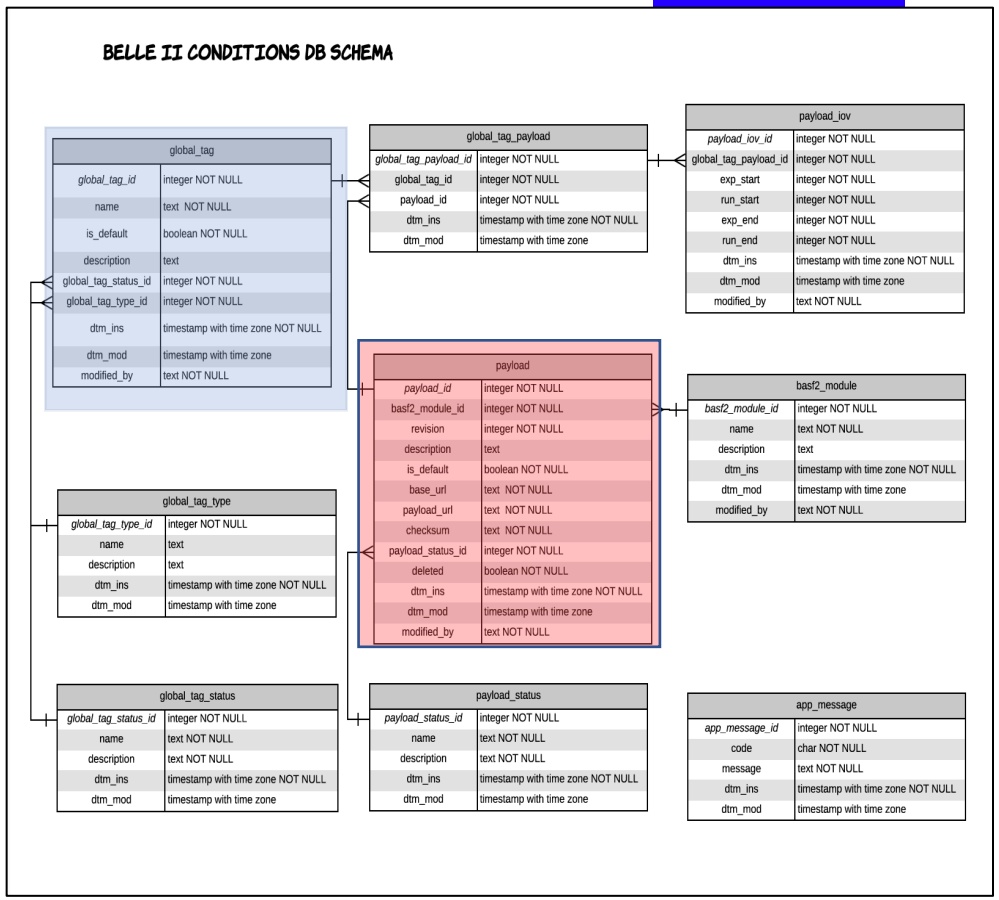
Belle II CDB data model relational database

Payloads contain a sample of conditions data (i.e. BeamParameters) stored in a file:

- File type is agnostic for the CDB server
- ROOT file format when restricted at client side
- CDB server metadata keeps track of the checksum of the file

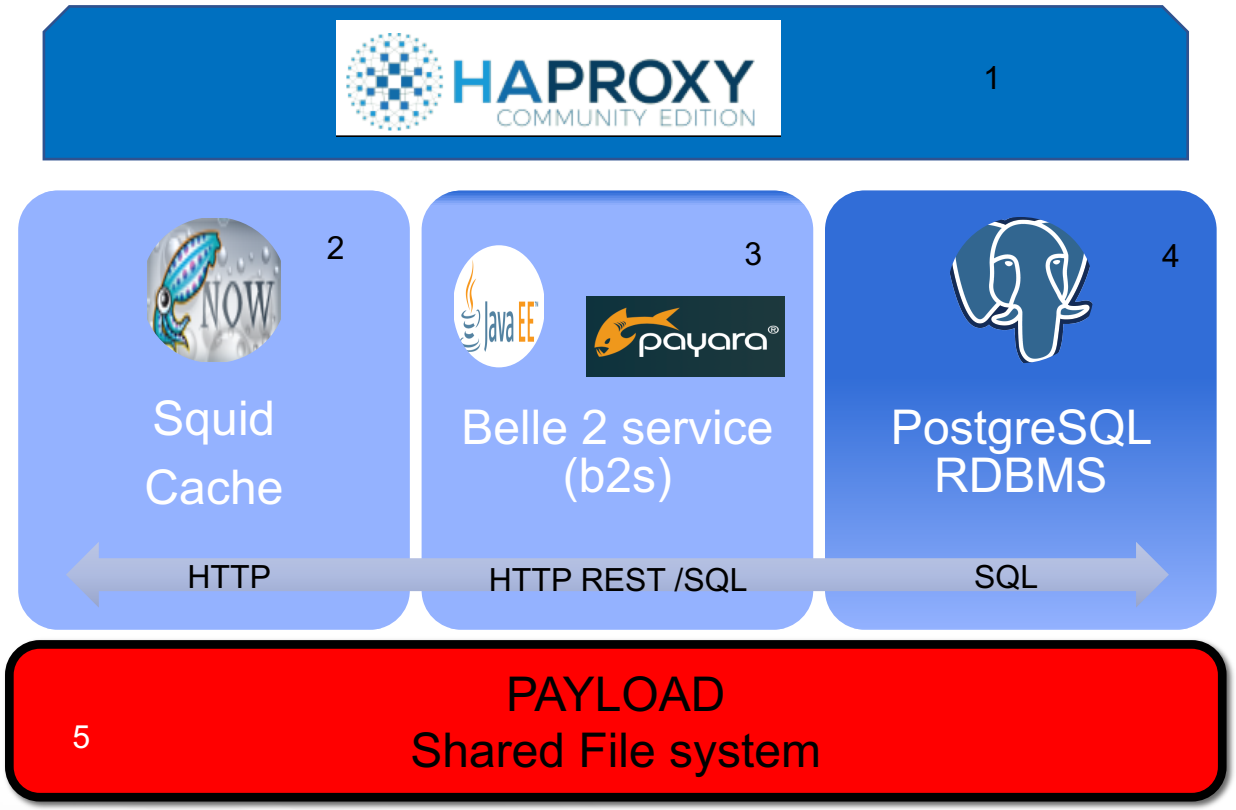
Intervals of Validity (IOV) specify starting and ending experiments and runs for a given payload for a specific global tag. Can be a fixed run range (closed) or starting at a given run (open)

Global tag (GT) contain list of IOV-payload relationships and are used to select a complete set of conditions for a given reprocessing effort

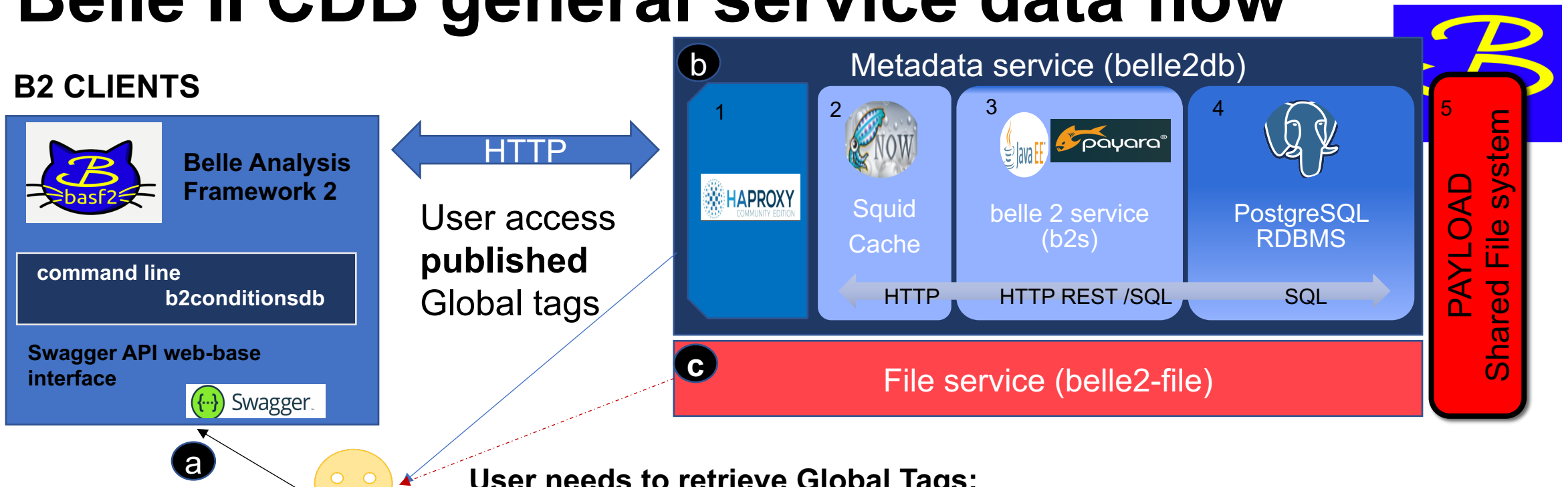


Belle II CDB general service architecture

1. HAproxy
   - Encryption/Decryption, Authorization and Authentication
   - Load balancing and resiliency
2. Caching technology
   - Allow Global Tag caching and offload dependent services
3. HTTP REST API presented by Belle 2 service (b2s) code
   - HTTP REST API Swagger enabled
4. Metadata stored in a persistent database repository
5. Payloads are stored in a shared file system
   - General Parallel File System enabled for metadata and payload service (next slide)



Belle II CDB general service data flow



User needs to retrieve Global Tags:

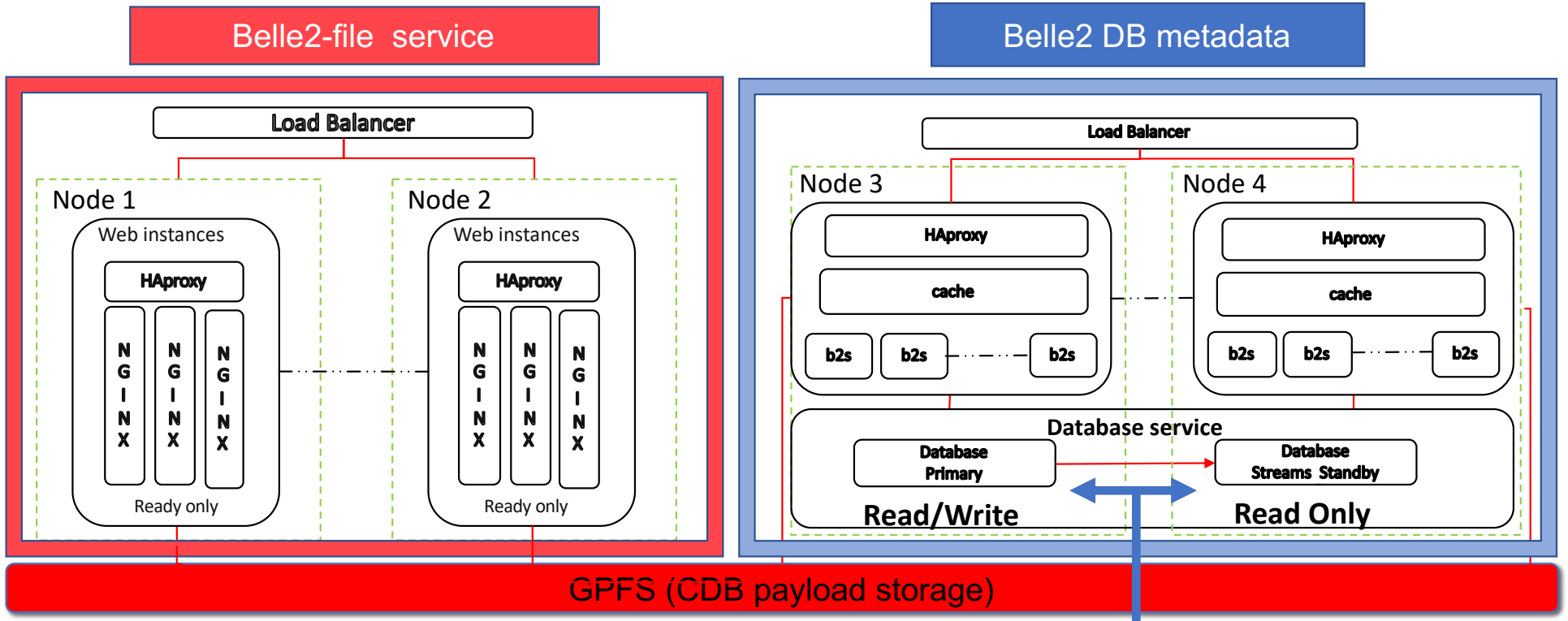
a. Uses a client to interact with the CDB metadata and file services

b. Metadata service provides information about the payloads and related IOVs of Global Tags:

- b.1 Request is Authenticated/Authorized and/or encrypted/decrypted
- b.2 The cache provide the request if possible
- b.3 The request is translated from HTTP to SQL (or vice versa) by the **b2s** service
- b.4 Database provide the metadata associated to the request

c. User retrieve payloads identified by the metadata on **b** via Belle2-files service

Belle II CDB server architecture (under the hood)
Orchestration of individual containerized components



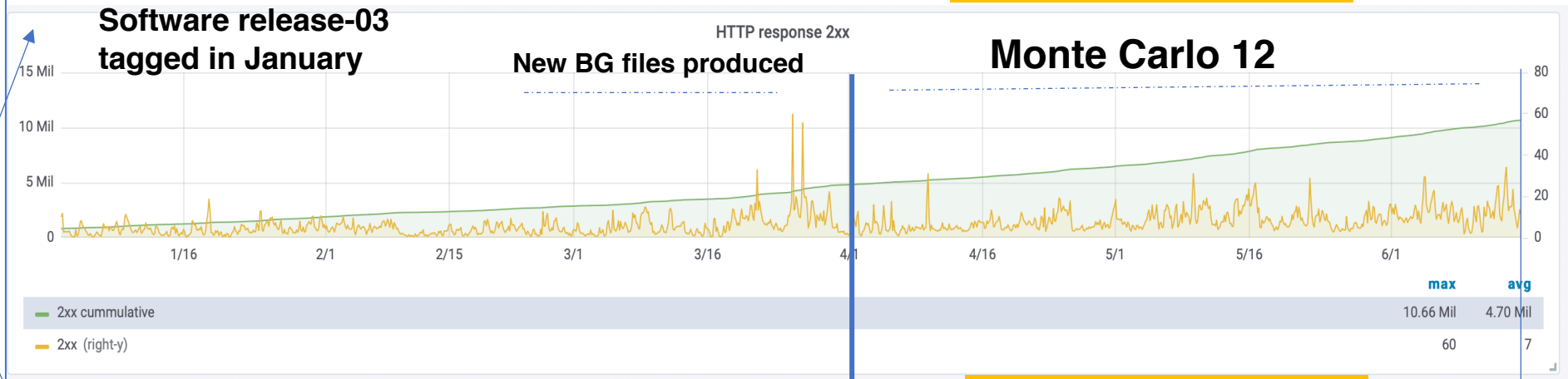
Streams standby replication technology enabled

- Database service open in READ only mode, mainly serving the ***CDB web monitor*** application
- Technology improves lag synchronization compared to the initial deployment
  - Initially hot standby was deployed for database disaster and recovery

Performance Belle II Conditions service CDB January-June 2019 all service response in terms of HTTP response code

Belle II Conditions metadata service

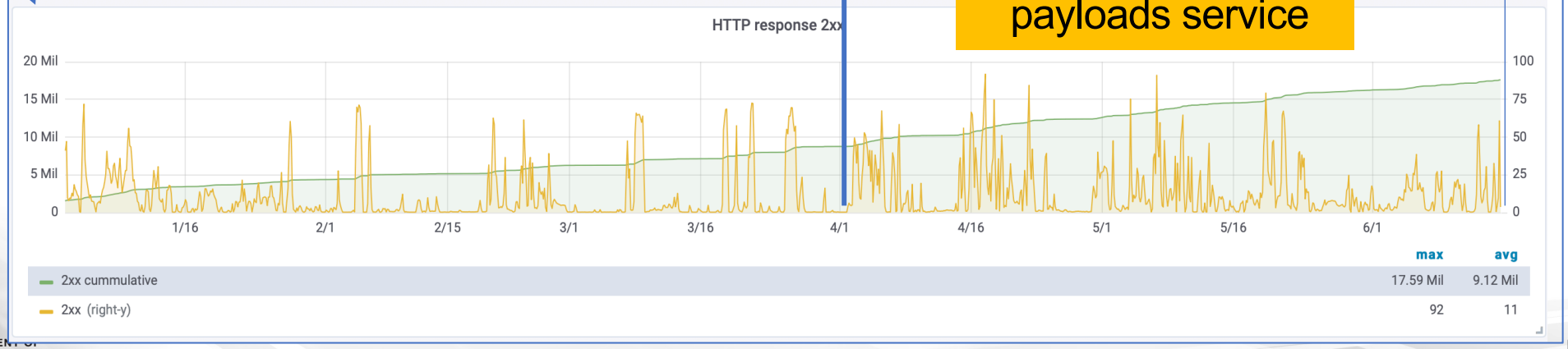
4.7M of HTTP 2XX succeeded



Average of HTTP requests

Belle II Conditions payloads service

9.5M of HTTP 2XX succeeded



CDB service resource usability within the capacity of resources provisioned.
See extra slides for capacity details.

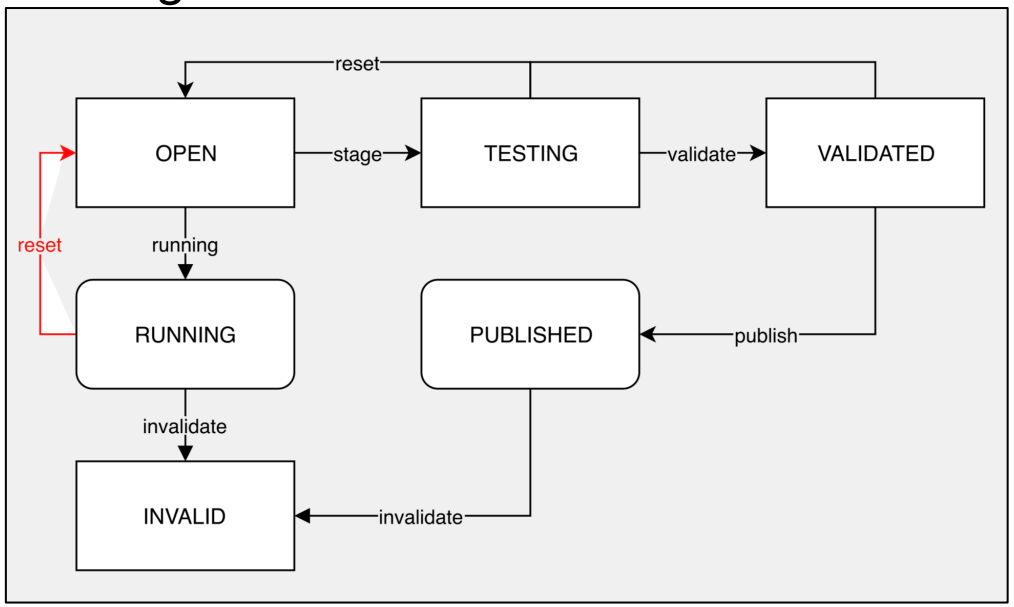
Operational Experience

Global tag state machine implementation into the CDB server

The initial design of the state machine had to be adjusted to changes in Belle II workflows

Changes included:

- Database schema modification
- New API HTTP REST endpoint
- Transition restrictions are implemented
- New functional tests server side
- Enhancement of Authorization schemes
  - Only privilege user can execute a change in the Global Tag state i.e. Open to Running



This required effort from different experts from software and database group.

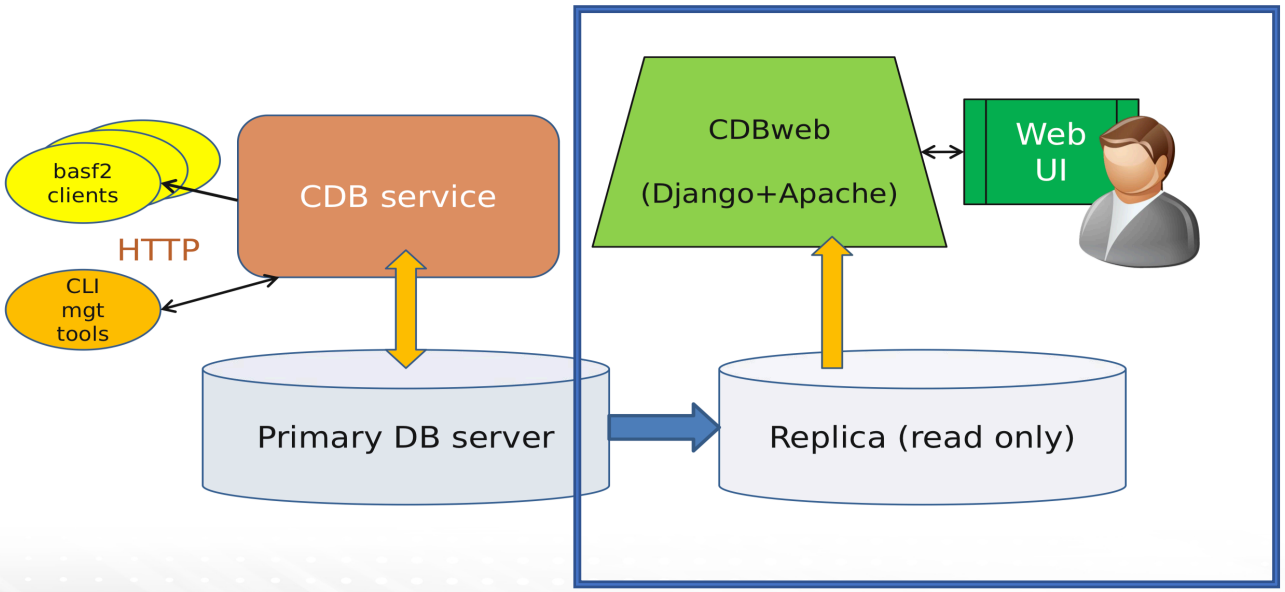
Deployed in July 1st 2019
All of this implementation deployed while minimize service disruption

Previously Global Tag states only supported three states, NEW, PUBLISHED, INVALID

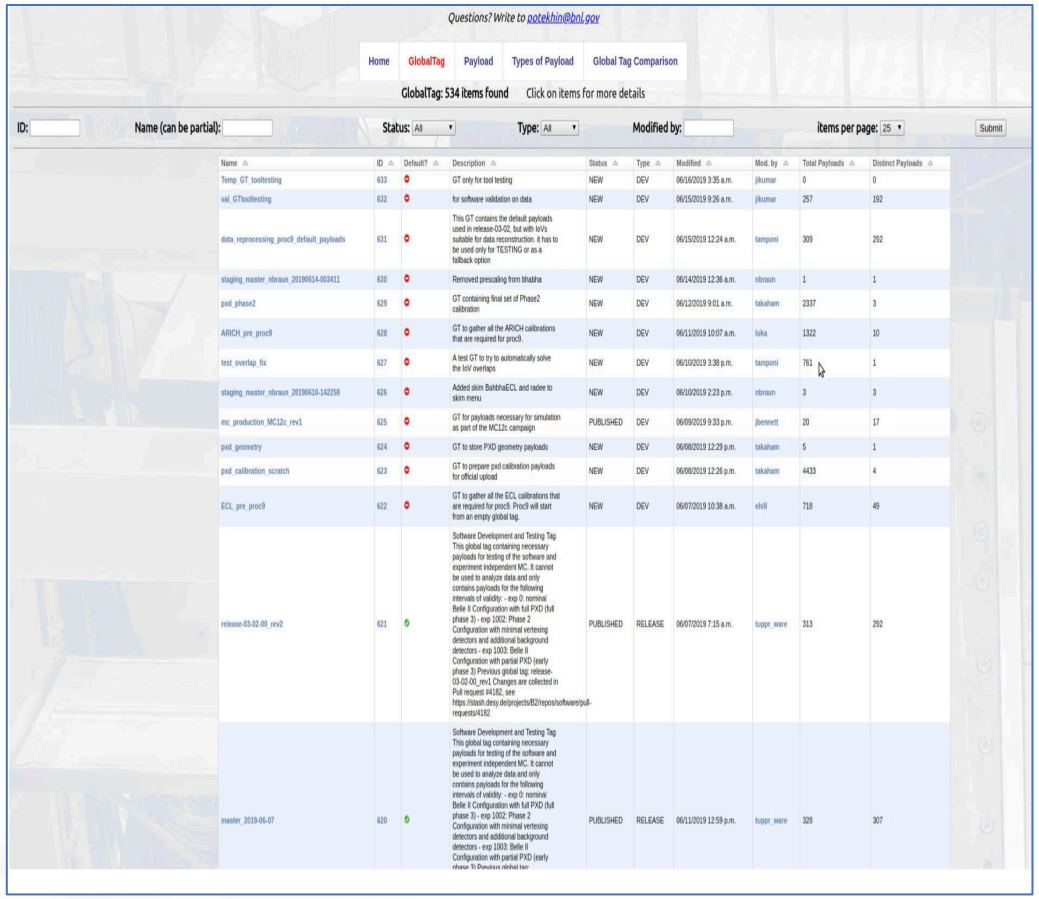
New developments: CDBweb browser

- Allows user to visualize database
- Conditions database content accessed via web browser.

Read Only Mode



The Global Tag page



Jason Web Token (JWT) authorization

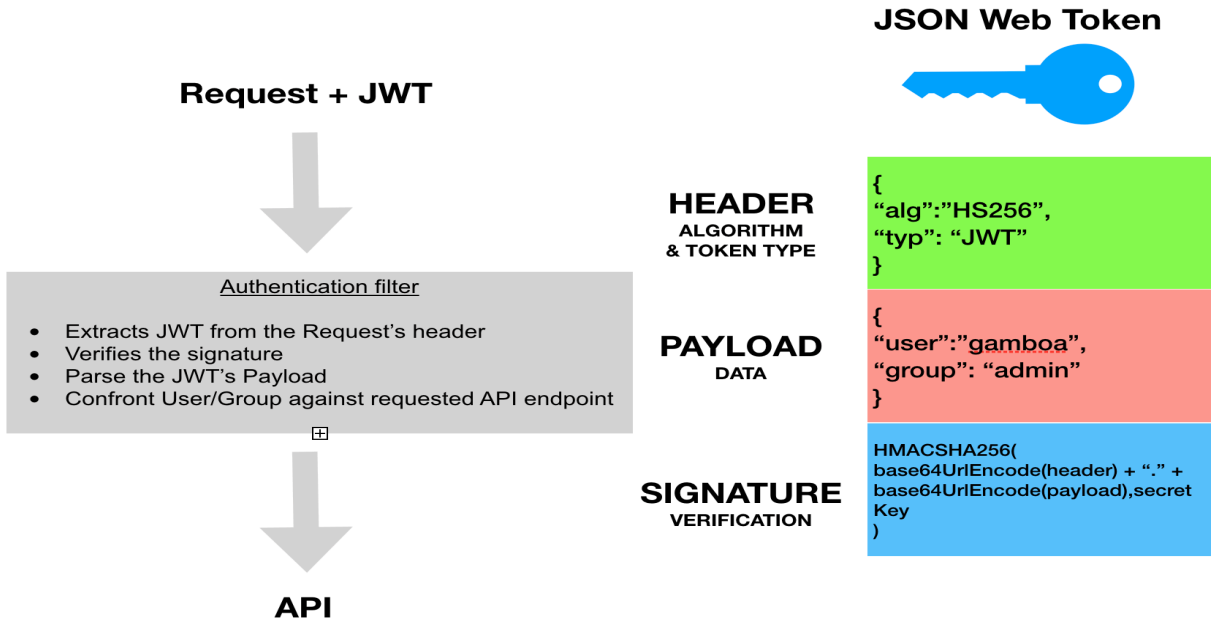
Belle II CDB is enhancing the authorization layer to enable fine-grained accessibility to serve conditions data.

Two possible options:

- Public Key Infrastructure (X.509 certificates)
- Token based infrastructure, JWT
  - Industry and HEP community are leaning towards token based technology as an alternative to provide Authorization as opposed to Public Key Infrastructure

Proof of concept of JWT for authorization implemented

Basic functionality tested via RESTful CDB API



Not a trivial task as the implementation involved troubleshooting different system used in the CDB framework

Future developments

HTTP REST API metadata

Augment current HTTP API to enhance support for Global Tag state machine

Authentication and Authorization

Extend JWT proof of concept by integrating an external token granting system and Belle 2 database clients

Summary

- Overview of the Belle II CDB database server presented
- **CDB service successfully delivered conditions data for the first run of the experiment in phase 3**
- Global tag state granularity extended
- PostgreSQL Streams standby replication of CDB metadata successfully enabled
- Alternative mechanism to visualize CDB content developed

Extra slides

Hardware capacity installed for Belle II CDB services

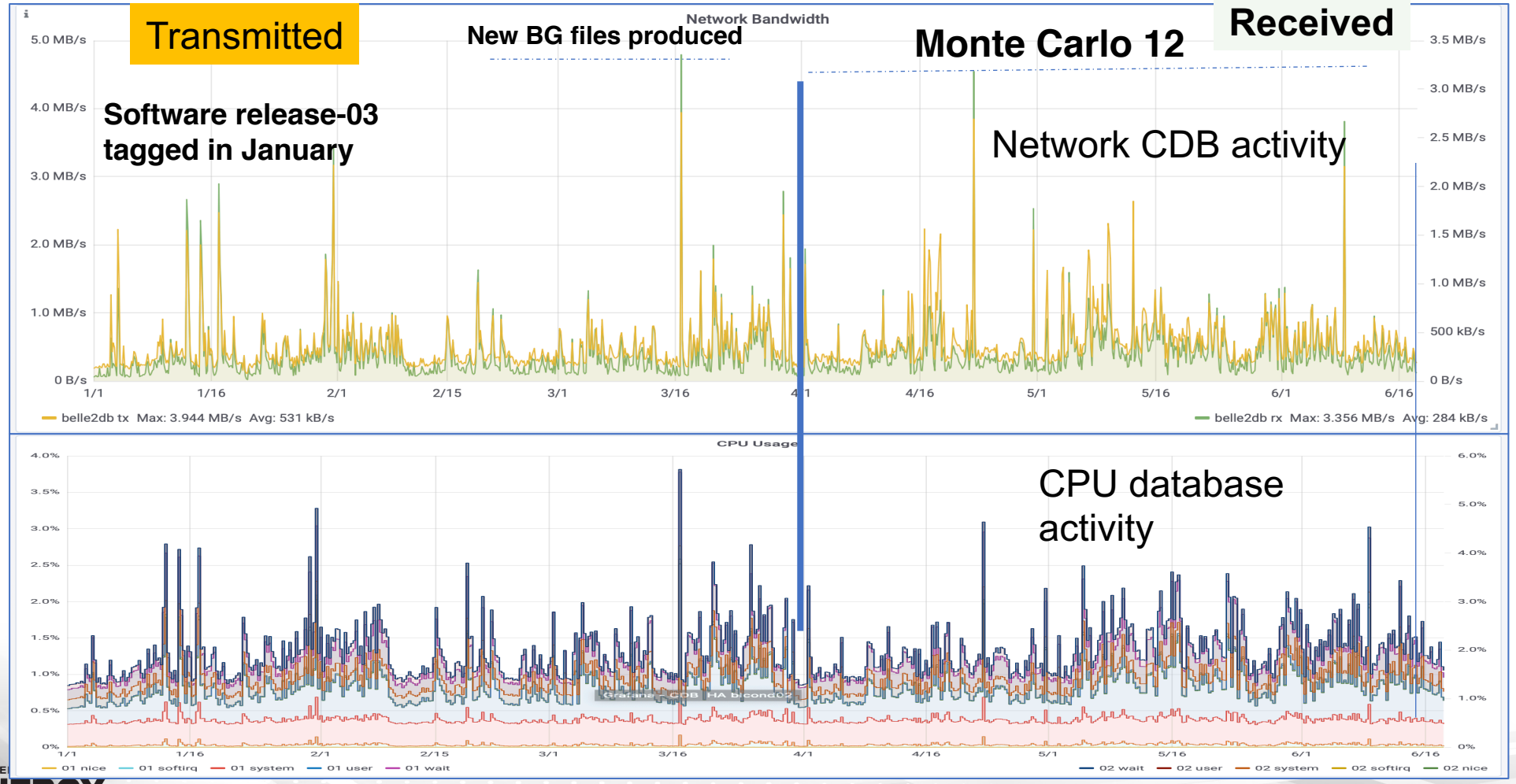
Belle2db service (metadata)

2 Nodes Dell R730xd
Two Intel(R) Xeon(R) CPU E5-2667 v4 @ 3.20GHz -> total CPU thread 32
Memory 256 GB
Disk for Database Intel DC P3700
NVMe SSD on PCI Express 800 GB
20Gb/s Channel Bounded connectivity

Belle2db-files service (payloads)

2 Nodes Dell R430
Two Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz total CPU threads 48
Memory 256GB
20Gb/s Channel Bounded connectivity

Performance CDB service (belle2db metadata service) CDB January-June 2019



Performance CDB service (belle2db-file services) from January-June 2019

CDB service resource usability within the capacity of resources provisioned despite of the increase in the usage of the belle2db-file service

Belle2db-file services files performance

Despite of the change in the access to the Belle2 DB file service, system remained stable and not major issues observed.