Database ecosystem at RACF/SDCC

CARLOS FERNANDO GAMBOA

ERIC LANCON

MAGNET TEST DATABASE WORKSHOP

BROOKHAVEN | Scientific Data and
NATIONAL LABORATORY | Computing Center

Scientific Data and Computing Center

⊙ Service Operation For:

RHIC, LHC ATLAS, BER ARM, LQCD, RIKEN, BES Center for Functional Nano Materials, National Synchrotron Light Source II, National Nuclear Data Center, Simons Foundation,…

⊙ ~1500 users from 20 projects (<10 to 100+ users/project)

SDCC and large experiments

- **The RHIC Tier 0**
  - Store and process data from RHIC experiments
  - Provide analysis means for 1'200 users
  - Long term data preservation
  - Simulation resources for future programs (sPHENIX & EIC)
- **The US ATLAS Tier 1**
  - ~25% of ATLAS Tier 1 computing capacity worldwide
  - Store RAW data from LHC and from simulation
  - Distribute data to the 4 US Tier 2 sites + analysis site (SLAC)
  - **Analysis center for US physicists**
    - From 41 institutes (incl. 4 Nat. Labs)
    - 600 physicists, 190 PhDs
- **The Belle II data center outside Japan**
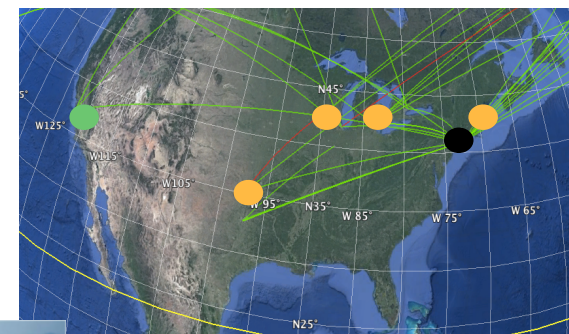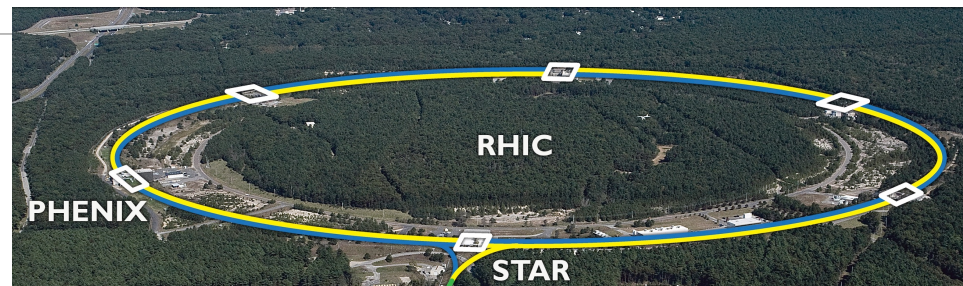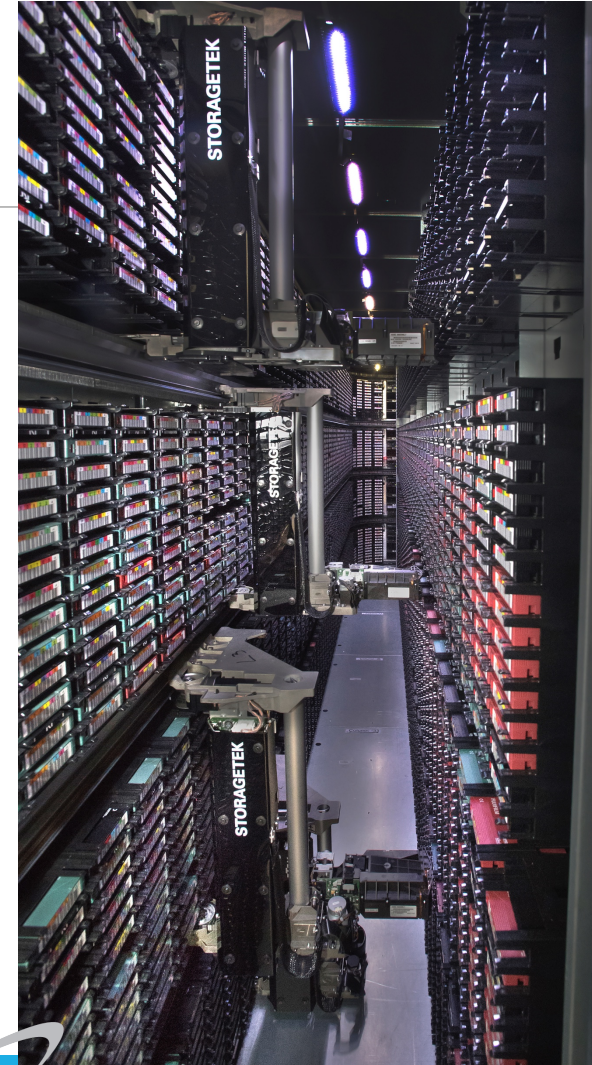
SDCC in numbers

- 90+k CPU cores — 4 PFlops
  - 3 HPC Institutional Clusters (GPU, KNL, Skylake)

- ~90 PB of disk storage
  - of various technologies

- 130+ PB of tape storage
  - Largest HPSS tape library in the US, 3rd worldwide [1]

- 2x100 Gbps connection to ESnet
  - Onsite ESNet support

BROOKHAVEN NATIONAL LABORATORY | Scientific Data and Computing Center

Databases in the SDCC

- Usability/technology driven by the application

- Relational databases most popular technology

- Commercial and open source databases. i.e DB2,MySQL and postgreSQL

General services

Storage

RELATIONAL DATABASES

Computing

BROOKHAVEN NATIONAL LABORATORY | Scientific Data and Computing Center
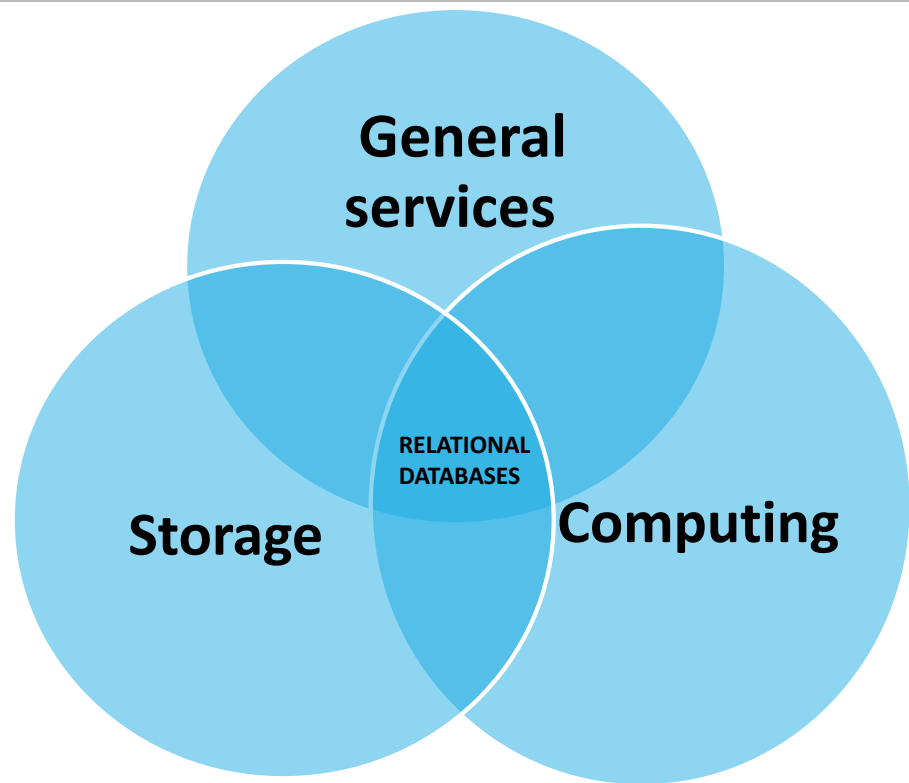
SDCC Critical services databases

⊙Storage

- ◦ dCache Storage Element's namespace uses PostgreSQL to store metadata (filenames, file ids, location,..), 432GB size .

  - ◦ Hosted in a powerful hardware (cpu, memory and storage) to ease application requirements.

⊙Tape Storage

- ◦ HPSS metadata manage all attributes and usages of files, and tapes, configuration of libraries, tape drives, 110 GB size.

- ◦ Commercial technology supported by IBM DB2

Databases as part of computer infrastructure management

PostgresSQL

Foreman and GLPI database size 10GB/each

MySQL

rt ticketing system

3 GB DB size

Databases in the SDCC, as a service for scientific experiments

BELLEII CONDITIONS DATABASE

Belle II CDB Remote service accessibility (WAN/LAN)
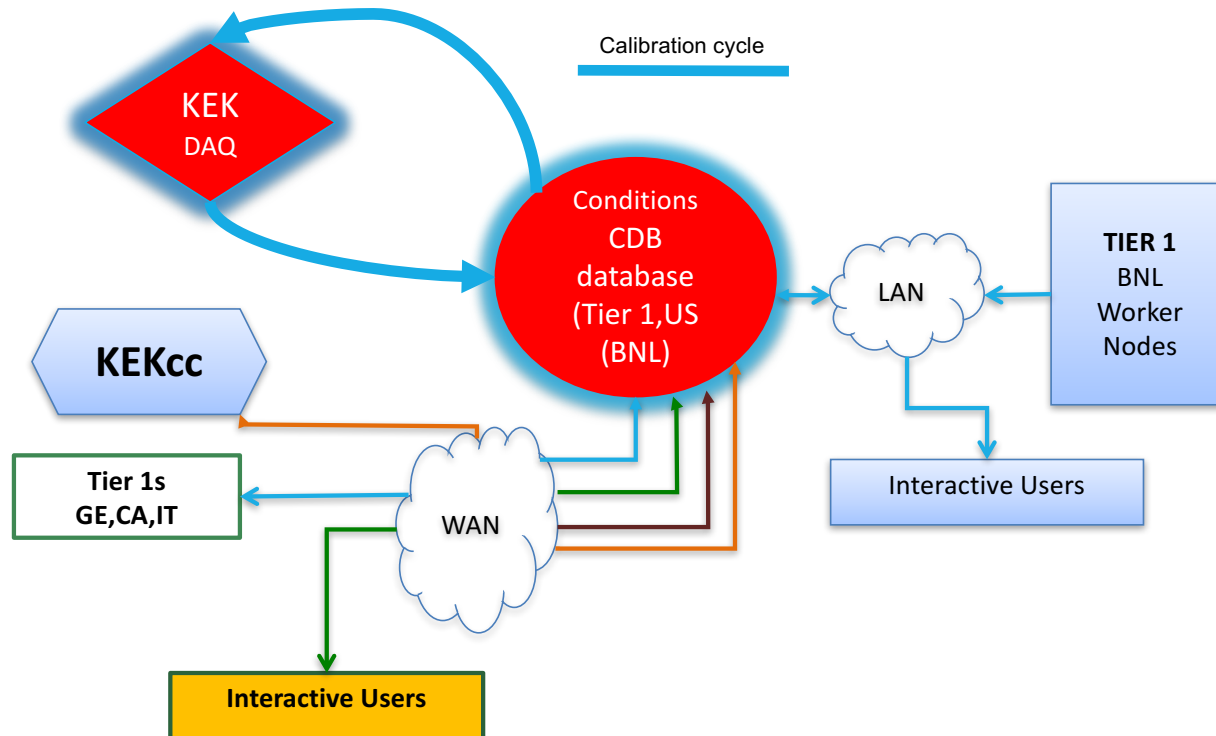
Databases in the SDCC
Belle II Conditions DataBase (CDB)

⊙ Events recorded by the Belle II detector are grouped in *runs[2]*.

⊙ A *run* sets a data taking period with stable operating conditions.

⊙ The database manages conditions data on run granularity

⊙ CDB metadata stored/accessed in the database:
  ◦ Allows identify payloads(Payload Binary objects) accessible is an external service.

[2] http://stacks.iop.org/1742-6596/898/i=4/a=042046

BROOKHAVEN
NATIONAL LABORATORY

Scientific Data and
Computing Center

Belle II CDB data model relational database

The payloads are defined for an interval of validity (IoV) and grouped into *global tags*.

BELLE II CONDITIONS DB SCHEMA

global_tag

global_tag_id	integer NOT NULL
name	text NOT NULL
is_default	boolean NOT NULL
description	text
global_tag_status_id	integer NOT NULL
global_tag_type_id	integer NOT NULL
dtm_ins	timestamp with time zone NOT NULL
dtm_mod	timestamp with time zone
modified_by	text NOT NULL

global_tag_payload

global_tag_payload_id	integer NOT NULL
global_tag_id	integer NOT NULL
payload_id	integer NOT NULL
dtm_ins	timestamp with time zone NOT NULL
dtm_mod	timestamp with time zone

payload_iov

payload_iov_id	integer NOT NULL
global_tag_payload_id	integer NOT NULL
exp_start	integer NOT NULL
run_start	integer NOT NULL
exp_end	integer NOT NULL
run_end	integer NOT NULL
dtm_ins	timestamp with time zone NOT NULL
dtm_mod	timestamp with time zone
modified_by	text NOT NULL

payload

payload_id	integer NOT NULL
basf2_module_id	integer NOT NULL
revision	integer NOT NULL
description	text
is_default	boolean NOT NULL
base_url	text NOT NULL
payload_url	text NOT NULL
checksum	text NOT NULL
payload_status_id	integer NOT NULL
deleted	boolean NOT NULL
dtm_ins	timestamp with time zone NOT NULL
dtm_mod	timestamp with time zone
modified_by	text NOT NULL

basf2_module

basf2_module_id	integer NOT NULL
name	text NOT NULL
description	text
dtm_ins	timestamp with time zone NOT NULL
dtm_mod	timestamp with time zone
modified_by	text NOT NULL

global_tag_type

global_tag_type_id	integer NOT NULL
name	text
description	text
dtm_ins	timestamp with time zone NOT NULL
dtm_mod	timestamp with time zone

global_tag_status

global_tag_status_id	integer NOT NULL
name	text NOT NULL
description	text NOT NULL
dtm_ins	timestamp with time zone NOT NULL
dtm_mod	timestamp with time zone

payload_status

payload_status_id	integer NOT NULL
name	text NOT NULL
description	text NOT NULL
dtm_ins	timestamp with time zone NOT NULL
dtm_mod	timestamp with time zone

app_message

app_message_id	integer NOT NULL
code	char NOT NULL
message	text NOT NULL
dtm_ins	timestamp with time zone NOT NULL
dtm_mod	timestamp with time zone

DATE 20170928171657

BROOKHAVEN NATIONAL LABORATORY | Scientific Data and Computing Center

Belle II CDB general service architecture

1. Metadata stored in a persistent database repository

2. HTTP REST API presented by b2s

3. Caching technology to offload the b2s/database service

4. Payloads are stored in a shared file system

File system (GPFS)
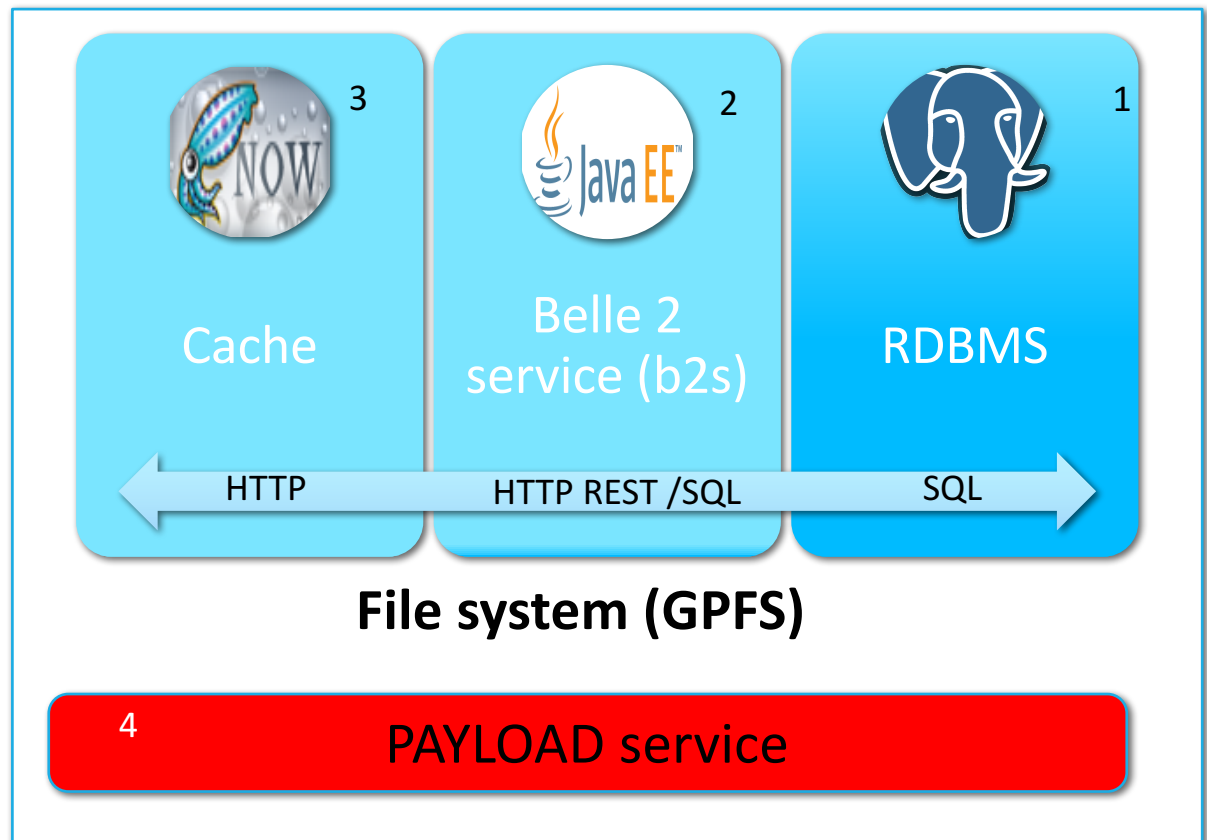
Belle II CDB deployed service
belle2db metadata service

- Architecture deployed using Kubernetes / docker framework
- Database replicated for reliability purposes
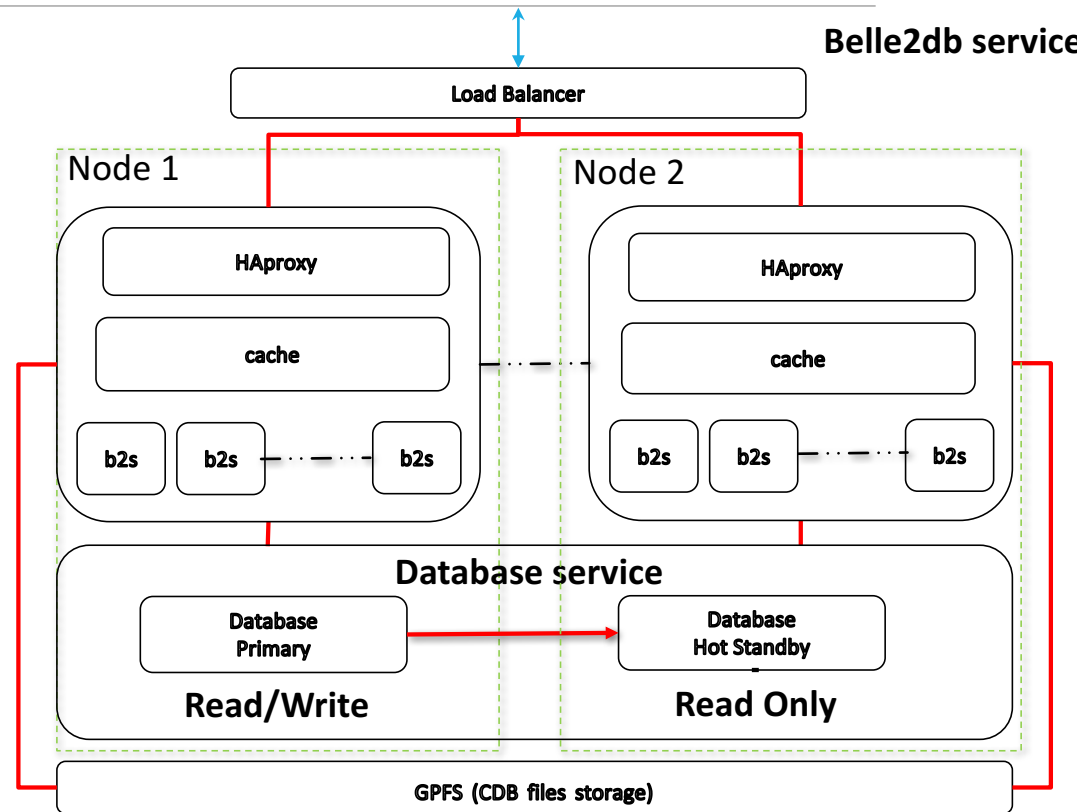- Hardware:

  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 service

Load Balancer

Node 1

HAproxy

cache

b2s | b2s | b2s

Node 2

HAproxy

cache

b2s | b2s | b2s

Database service

Database Primary

Database Hot Standby

Read/Write

Read Only

GPFS (CDB files storage)

BROOKHAVEN NATIONAL LABORATORY | Scientific Data and Computing Center

Belle II CDB deployed service payload service

◉ **General Parallel File System (GPFS) used to store CDB payload**

◉ **Hardware**

- ○ 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

Belle2db-files service
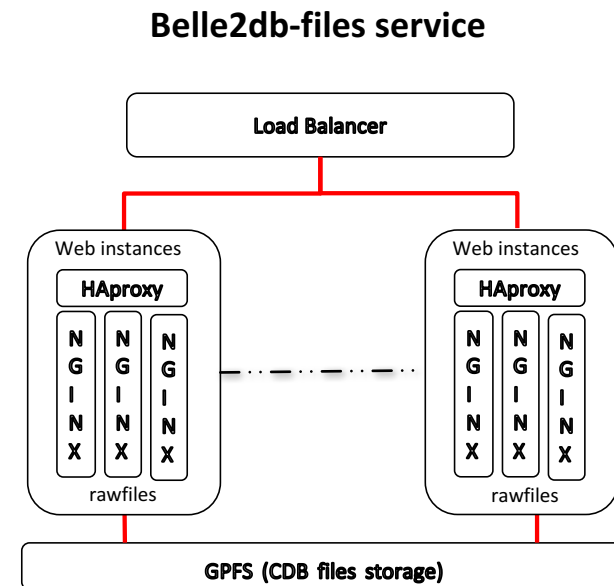
Summary

- Overview of the SDCC database services presented.

- SDCC expertise in deploying and operating different database technologies to host metadata/data critical for application performance and resiliency.

- SDCC supports databases core for HEP scientific experiment operation.