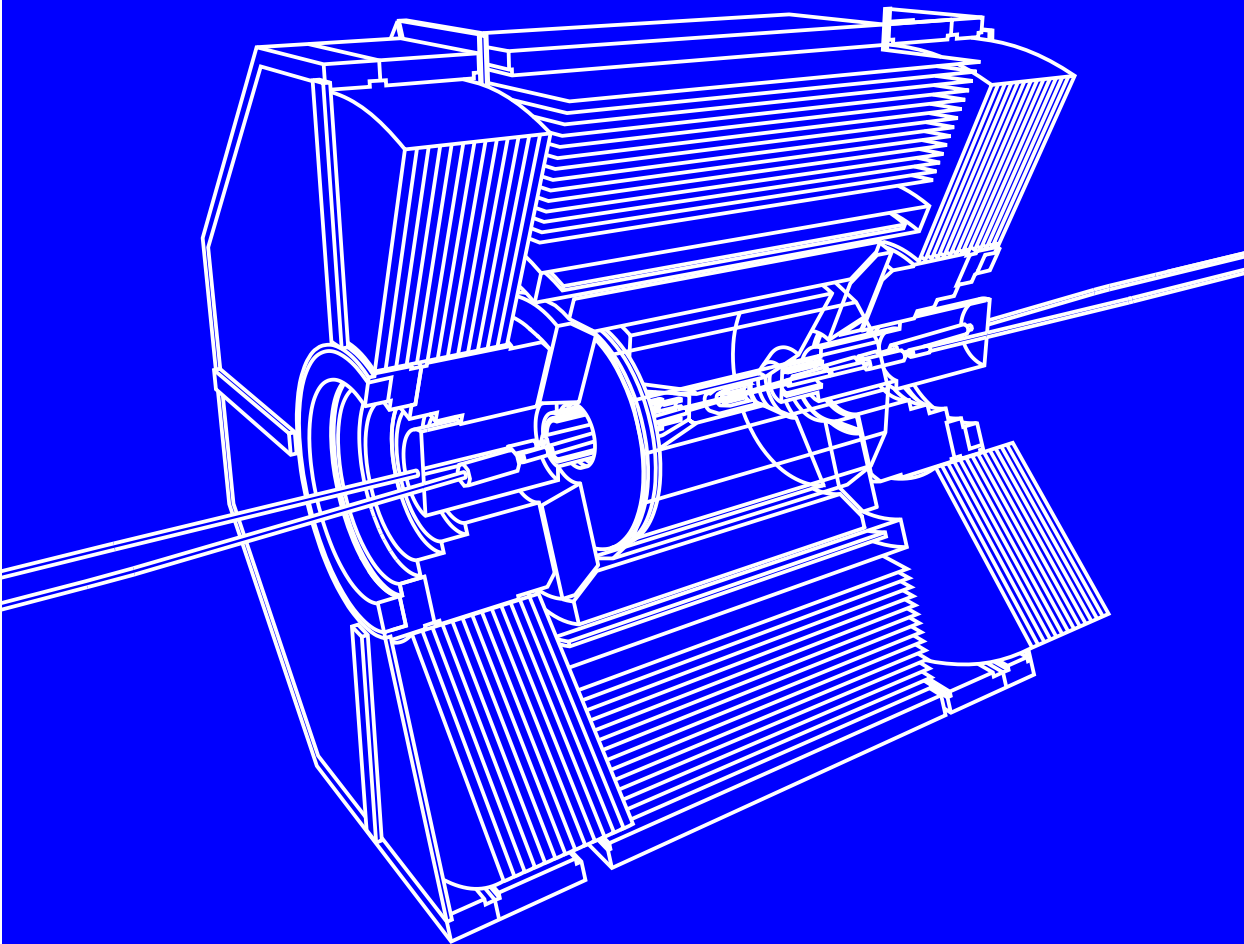Belle Ⅱ

Belle & Belle II

Takanori Hara (KEK)

takanori.hara@kek.jp

9 June, 2015
DPHEP Collaboration Workshop @ CERN

Belle Data

Belle : started in 1999, data-taking completed in 2010

still keep analysing the Belle data

in parallel with the construction of Belle II detector and computing

Integrated Luminosity(log)

$[fb^{-1}]$

Y(1S) : 6 fb^{-1}
Y(2S) : 24 fb^{-1}
Y(3S) : 3 fb^{-1}
Y(4S) : 711 fb^{-1}
Y(5S) : 121 fb^{-1}
off-reso. : ~100 fb^{-1}

> 1.0 ab^{-1}

as of July, 2010

~770M BB pairs @ Y(4S)

1998/1/1   2000/1/1   2002/1/1   2004/1/1   2006/1/1   2008/1/1   2010/1/1

Size of storage for Data

RAW	raw data	~1000 TB
DST	prescaled data (1-1/400)	320 TB
mDST	reconstructed info.	150 TB

Format : panther (Belle's own bank system)
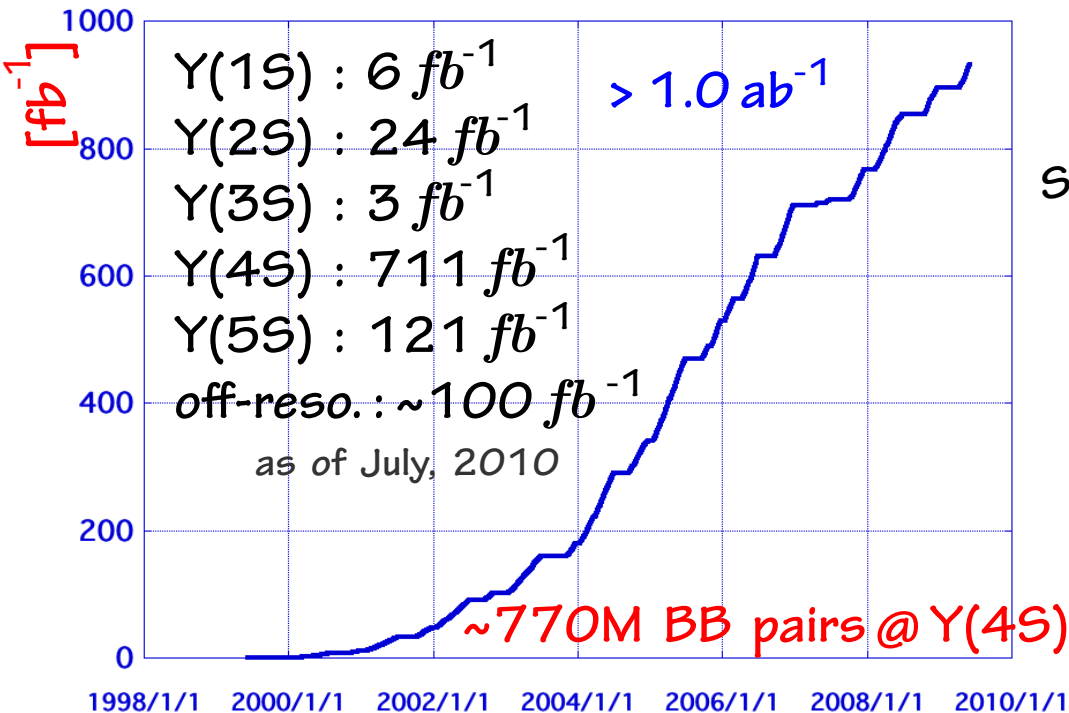
Size of storage for MC

mDST	recon. info. + MC true	800 TB

10 streams for bb + 6 streams for udsc

we have two versions of mDST
  one : w/ old-tracking (=conformal finder)
  the other : w/ new-tracking (=+Hough finder)

mDST stored on both Disk and Tape
raw data + DST stored on Tape

Belle's policy

Preserve all RAW and mDST, at least.

Keep the current computing environment AMAP
(Library, Database, data format)

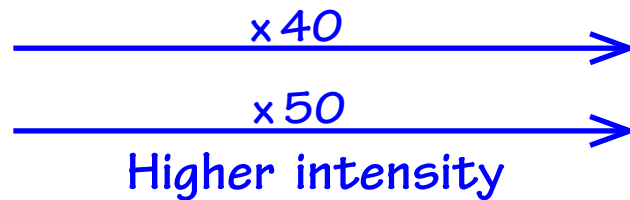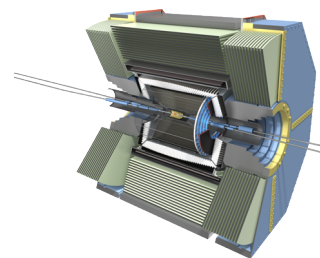Belle I data will be used by the Belle community
until the time the statistics of Belle II > Belle I data set

Belle

Belle II

	Belle	Belle II
Accelerator	KEKB	SuperKEKB
Beam Energy (GeV)	3.5×8 ($\gamma = 0.425$)	4×7 ($\gamma = 0.28$)
CM energy	......, Y(4S), ...	......, Y(4S), ...
Luminosity ($cm^{-2}s^{-1}$)	2.1×10^{34}	8×10^{35}
Total data (ab^{-1})	1	50

x 40

x 50

Higher intensity
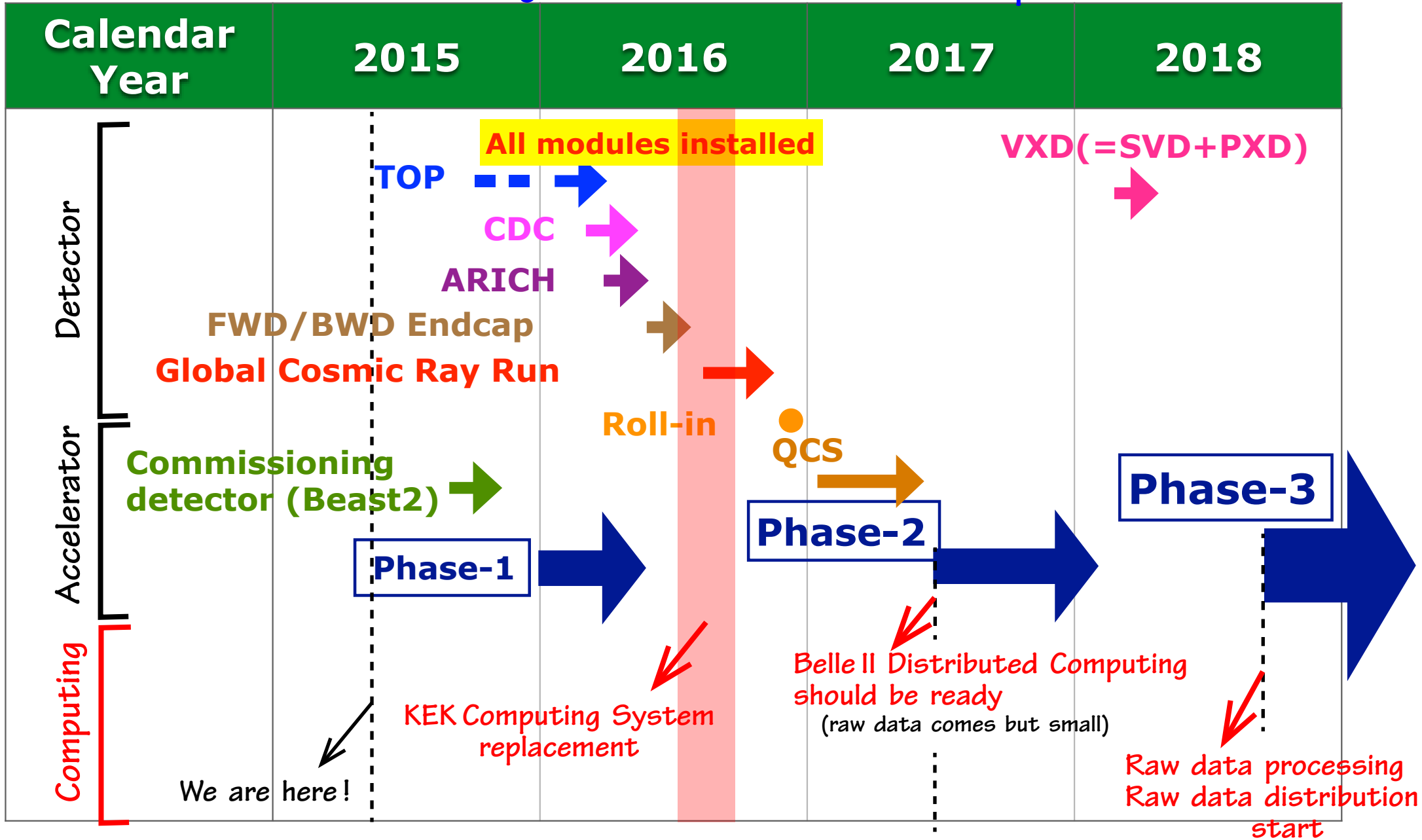
raw data : ~1PB
mDST data/MC : 0.15/0.8 PB
(for one version)

raw data : ~100PB
(another raw data copy outside KEK)

Computing   one big center @ KEK (non-grid) ⟶ world-wide distriuted computing

SuperKEKB/Belle II Time line

KEK is the hosting institute of the Belle II experiment

Calendar Year	2015	2016	2017	2018

Detector

All modules installed

VXD(=SVD+PXD)

TOP

CDC

ARICH

FWD/BWD Endcap

Global Cosmic Ray Run

Roll-in

Accelerator

QCS

Commissioning detector (Beast2)

Phase-1

Phase-2

Phase-3

Computing

KEK Computing System replacement

Belle II Distributed Computing should be ready
(raw data comes but small)

Raw data processing
Raw data distribution start

We are here!

How long ?

The data taking of the Belle experiment finished in 2010

But in 2018, physics run with the Belle II full detector starts

(physics run w/o VXD starts in 2017)

. We have to work on updating the Belle detector
. We will have a new computing system for Belle II experiment
(in 2016, including resources for Belle I + J-PARC)

We are aware of the need for Belle data preservation

Belle Y(4S) data will be superceded in 2019

It is important to validate Belle II initial results

Y(nS) data are unique data sets

Belle data should be preserved at least until 2020 ?

Until that, the Belle Data will be stored on storage
within the Belle community

Strategy for data preservation

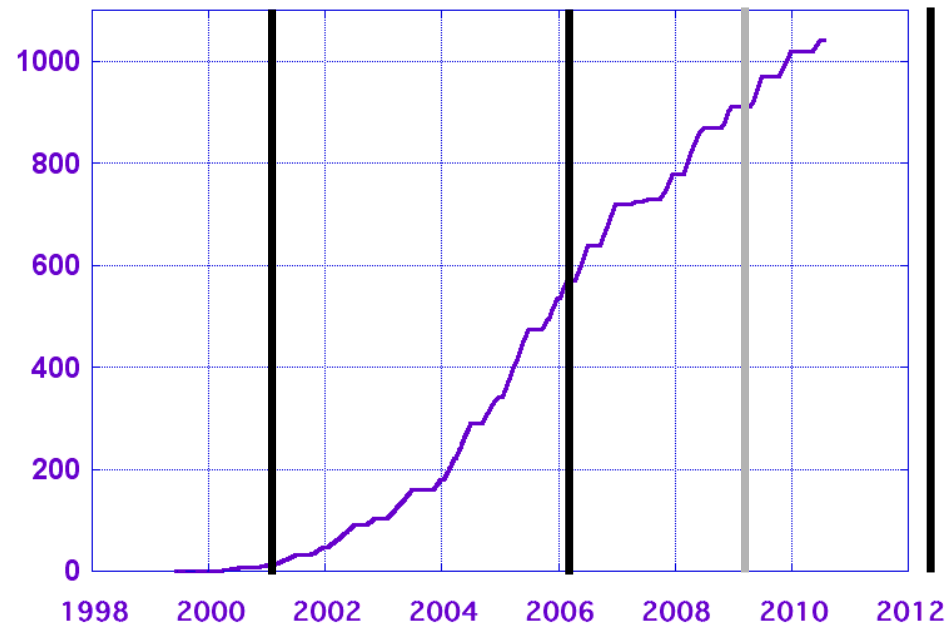| Year (contract) / specification	1999- (4years)	2001- (5years)	2006- (6years)	2009- (continued)	2012- (until 2016/8)
CPU [SI2k]	~100 (WS)	~1200 (WS+PC)	~42500 (PC)	~115200 (PC)	~3500 cores ~50kHS06
Disk [TB]	4	9	1000	1500	7000
Tape [TB]	160	620	3500	3500	16000 (at max)
	(Belle dedicated)	(Belle dedicated)	(Belle dedicated)	(Belle dedicated)	(Belle + other KEK projects)

Integrated Luminosity[fb-1]

Next replacement takes place

in Summer 2016

(until Summer 2020 ?)

Strategy for data preservation

Size of storage for Data

RAW	raw data	~1000 TB
DST	prescaled data (1-1/400)	320 TB
mDST	reconstructed info.	150 TB

Format : panther (Belle's own bank system)

Size of storage for MC

mDST	recon. info. + MC true	800 TB

10 streams for bb + 6 streams for udsc

These "official" data will be copied from the current computing system to the new one

no users' data will be copied by default

there will be a period that the old and new systems are operated in parallel.

The procedure must depend on which company will operate the new system

a couple of months or more...

data preservation outside KEK

After the Earthquake in 2011, we copied mDST to
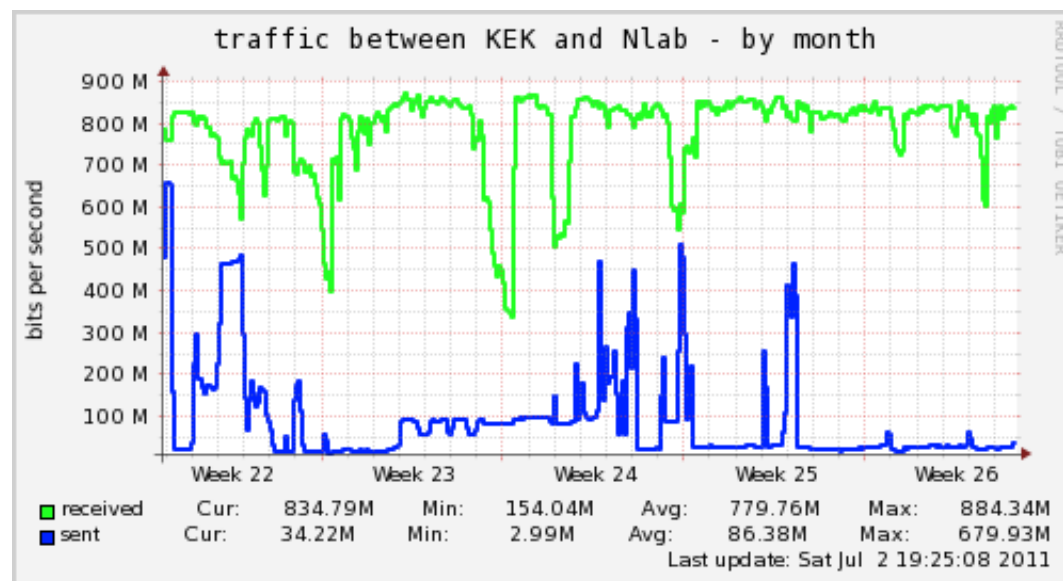PNNL(USA) and Nagoya (Japan) as of Feb. 12, 2012

PNNL(USA)

◆ 1 PB disk / 1 PB tape were prepared

- ◆ Data copy from KEK via "scp"
  - ◆ (7 TB/day at maximum, 1-2 TB/day in average)

- ◆ Only skimmed mDST (no raw data)
  - ◆ HadronBJ, tau, lepton skim, Dilep, Ypipi

- ◆ Production of MC mDST at PNNL
  - ◆ faster than the data copy
    - ◆ Y(4S) MC 2 sets
    - ◆ Y(5S) MC 2 sets
    - ◆ RareMC 150 × data

Nagoya (Japan)

- ◆ Case B Data (no raw data)
  HadronB(J), full-recon, tau_skimB
  generic MC : 6 streams
  ekpturbo files (Data and MC)
  beam background files (run-dep. and run-indep.)

- ◆ Case A Data
  HadronB(J), full-recon, tau_skimB, lowmulti
  generic MC : 4 streams

traffic between KEK and Nlab - by month

	Cur:	Min:	Avg:	Max:
received	834.79M	154.04M	779.76M	884.34M
sent	34.22M	2.99M	86.38M	679.93M

Last update: Sat Jul 2 19:25:08 2011

Strategy for analysis preservation

Make the Belle data readable within the Belle II software framework

Belle

"BASF" (= Belle AnalysiS Framework) is a software framework.

. each task (e.g. tracking) is implemented as a module

. data is exchanged through Belle Bank System

. works under the Multi-CPU System

- . Offline software : almost frozen
  software will not be updated
- . DST/MC production : not planned
  mDST is enough for analysis
- . Retention of knowledge :

    not well documented
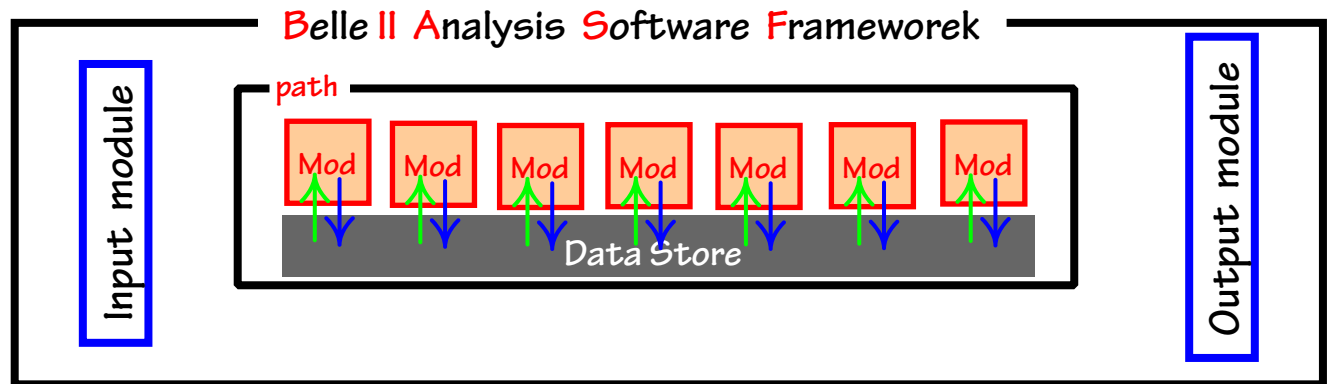 *the number of people who knows Belle software is decreasing*
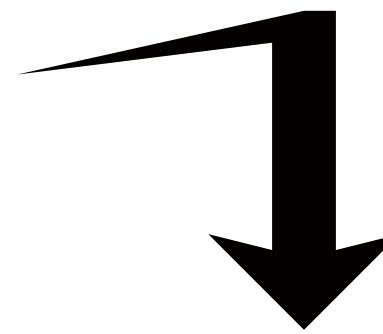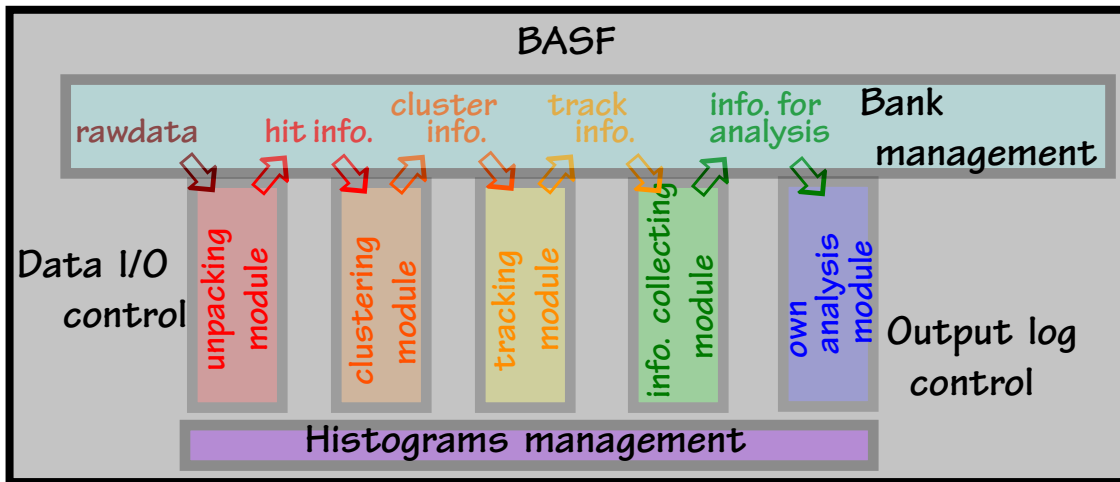
Belle II

even data Input/Output is handled by module

Panther

Belle own made bank system

Root

Strategy for analysis preservation

Two possibilities:

Reading Belle data (panther) from basf2 on the fly

- write a basf2 module, which
    - Reads in Belle data (in panther) : hard
    - Converts it to basf2 objects : need careful check
    - Writes it out in basf2 format (in root) : easy
- possible difficulties
    - panther is heavily bundled with the Belle software
    - Database access from basf2 is unclear

Converting Belle data to basf2 data format

- Write a basf module, which
    - Reads in Belle data (in panther) : easy
    - Converts it to basf2 objects : need careful check
    - Writes it out in basf2 format (in root) : not difficult
- possible difficulties
    - heavy labor for conversion
    - extra storage space

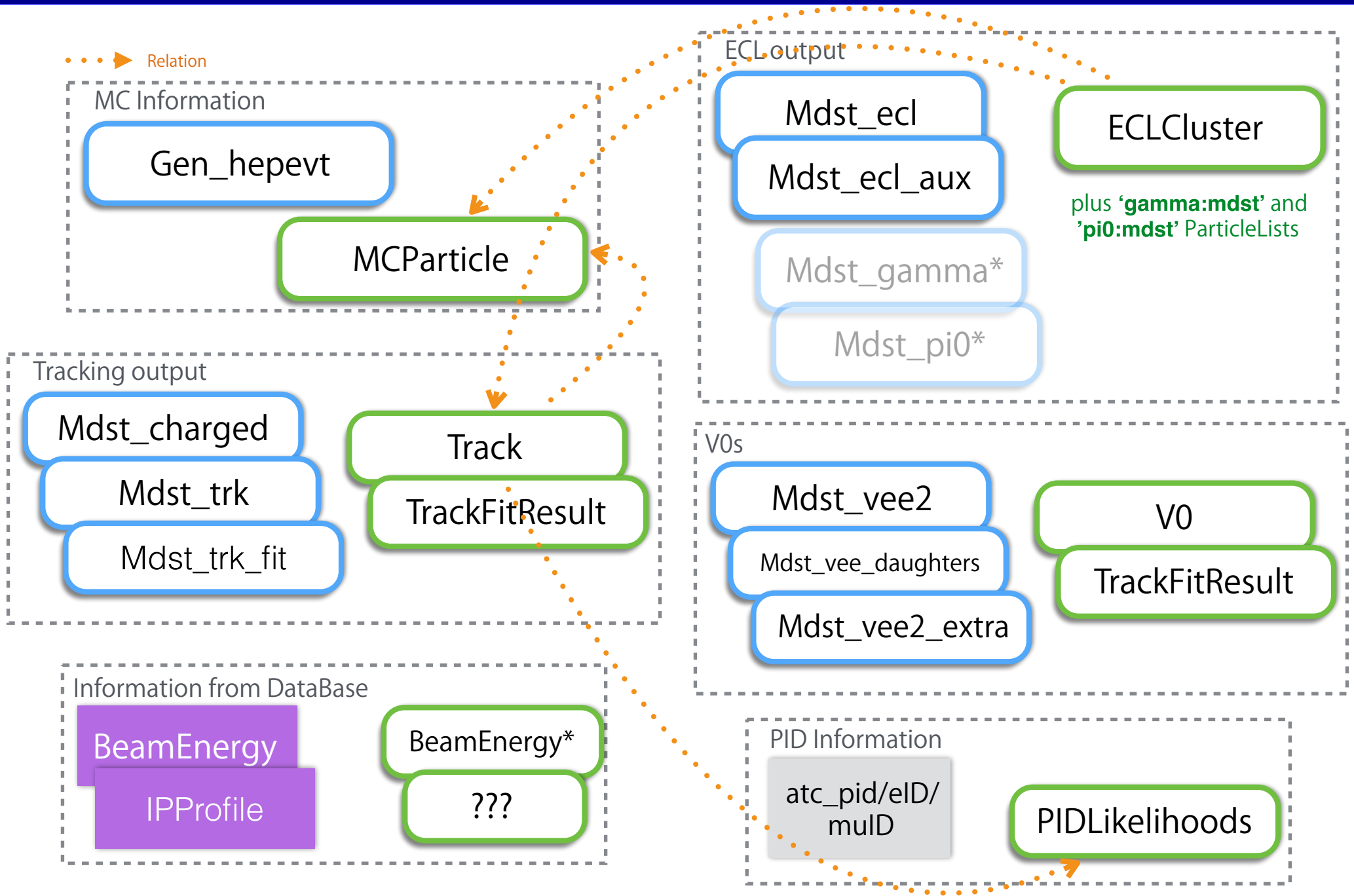# Strategy for analysis preservation

Two possibilities:

Reading Belle data (panther) from basf2 on the fly

- write a basf2 module, which
  - Reads in Belle data (in panther) : hard
  - Converts it to basf2 objects : need careful check
  - Writes it out in basf2 format (in root) : easy
- possible difficulties
  - panther is heavily bundled with the Belle software
  - Database access from basf2 is unclear

Converting Belle data to basf2 data format

- Write a basf module, which
  - Reads in Belle data (in panther) : easy
  - Converts it to basf2 objects : need careful check
  - Writes it out in basf2 format (in root) : not difficult
- possible difficulties
  - heavy labor for conversion
  - extra storage space

Convert MDST objects

Relation

MC Information

Gen_hepevt

MCParticle

Tracking output

Mdst_charged

Mdst_trk

Mdst_trk_fit

Track

TrackFitResult

Information from DataBase

BeamEnergy

IPProfile

BeamEnergy*

???

ECL output

Mdst_ecl

Mdst_ecl_aux

Mdst_gamma*

Mdst_pi0*

ECLCluster

plus **'gamma:mdst'** and **'pi0:mdst'** ParticleLists

V0s

Mdst_vee2

Mdst_vee_daughters

Mdst_vee2_extra

V0

TrackFitResult

PID Information

atc_pid/eID/muID

PIDLikelihoods

Database access

PostgreSQL interface "pntdb" in the Belle library was migrated to the Belle II lib.
Database server was activated at Karlsruhe in Germany

it is not easy to set up DB server at KEK, which is accessible from the internet

```
Belle::Beam_energy_Manager& Beam_mgr = Belle::Beam_energy_Manager::get_manager();
Beam_mgr.remove();


// Open Database connection
Belle::TPntFDDB master("rif");   // use default hostname
Belle::TPntDB constant(master, "benergy");   // implicitly call Open fddb/db


if (not constant.IsOK()) {
  B2ERROR("PantherInput: Couldn't connect to rif::benergy");
  return;
}


// Read out beam-energy of off-resonance data experiment 71!
// exp: 71: run: 1 (run indepent number is stored in 1, NOT 0), version: 2
if(constant.Get(71, 1 , 2) <= 0){
  B2ERROR("PantherInput: Cannot get data from database correctly.");
  return;
}


Belle::Beam_energy_Manager::iterator beit = Beam_mgr.begin();
if( beit == Beam_mgr.end()){
  B2ERROR("PantherInput: There is no Beam Energy data.");
  return;
} else {
  B2INFO("PantherInput: BeamEnergy is" << beit->E_beam());
}
```
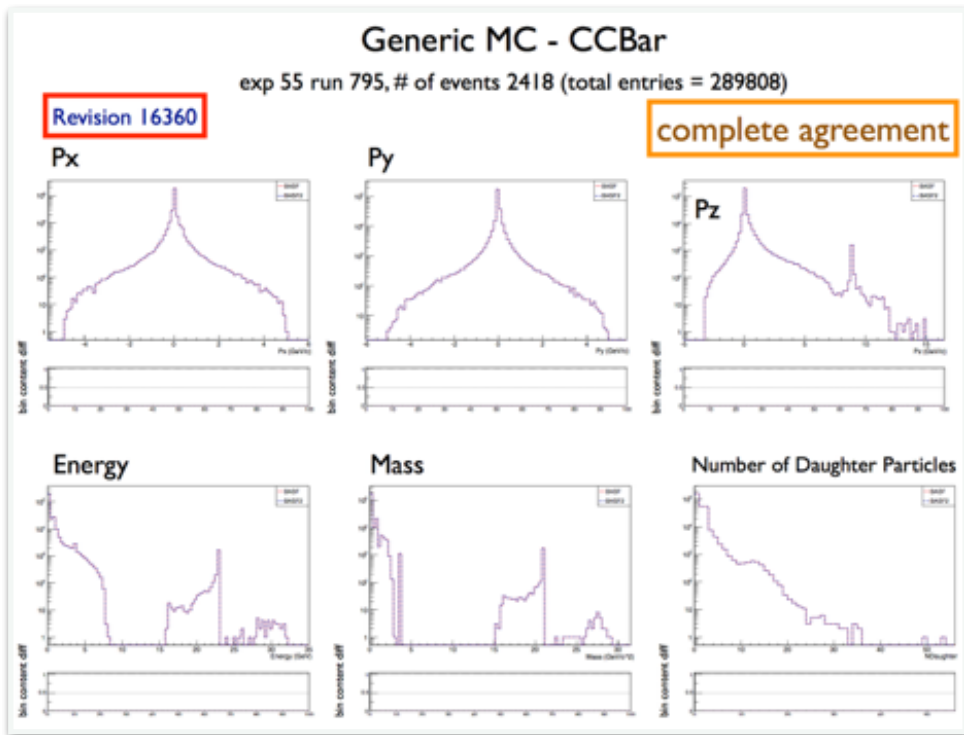
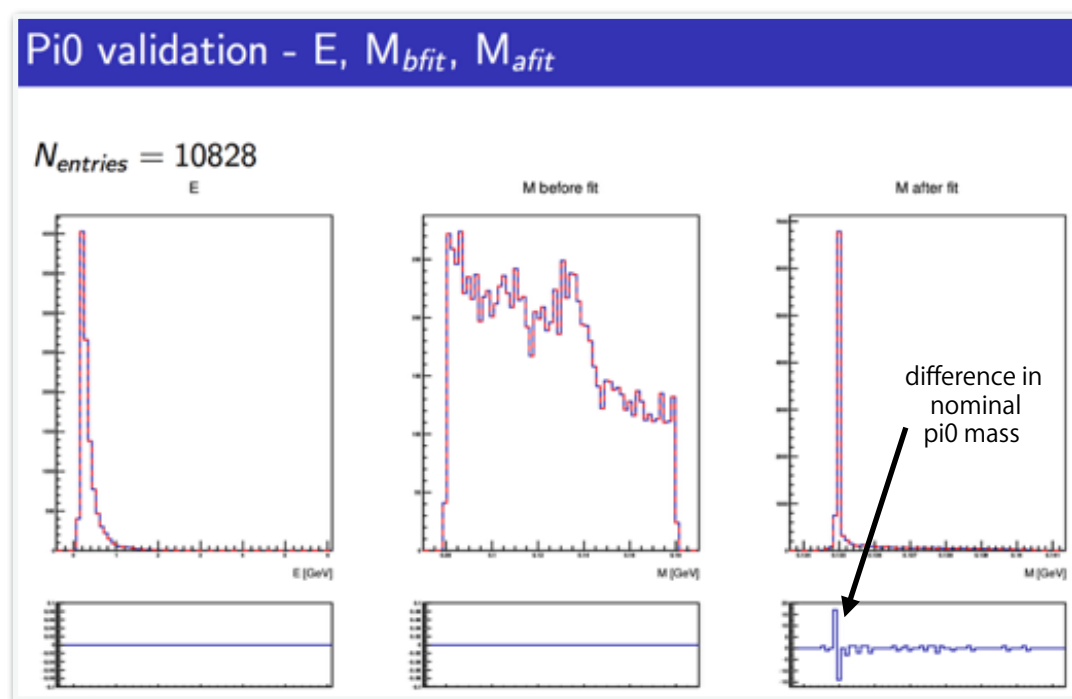proof of principle
BeamEnergy for exp71/run1
5.43429

```
[INFO] Steering file: b2bii/examples/DumpData.py
[INFO] PantherInput: BeamEnergy is 5.43429
[INFO] PantherInput: initialized.
[INFO] PantherInput: beginRun called.
```
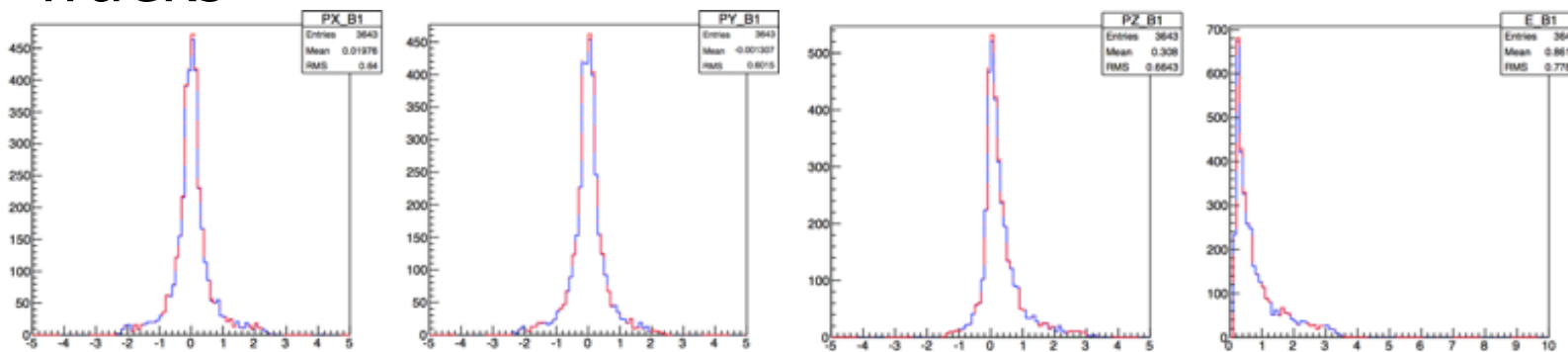
Validation

MCParticles

EM calorimeter : cluster

Tracks

Summary of Belle data/analysis preservation

> No official reprocess from the raw data.
> No official MC mass production (except user-level signal MC)
> Offline software is frozen
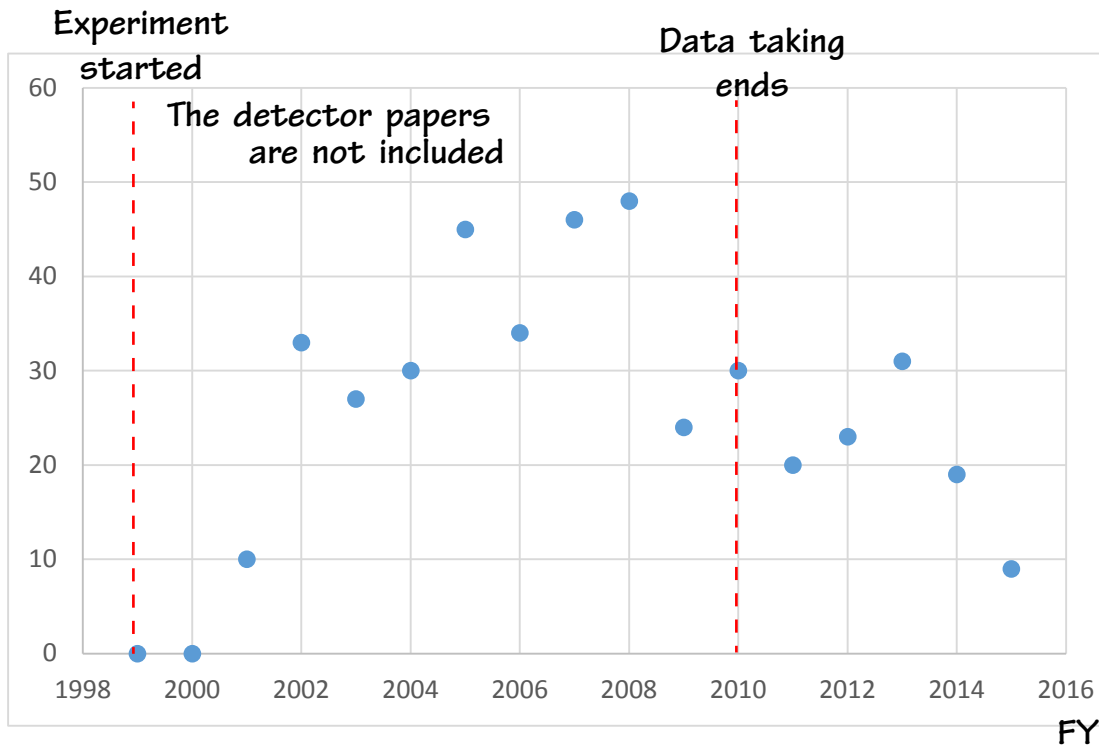
Belle data preservation

- The Belle data will be copied from the current computing system to the new one

  - New computing system starts operation in Summer 2016 (until summer 2020?)
  - two replicas of mDST (analysis-oriented data) of Data/MC in PNNL and Nagoya
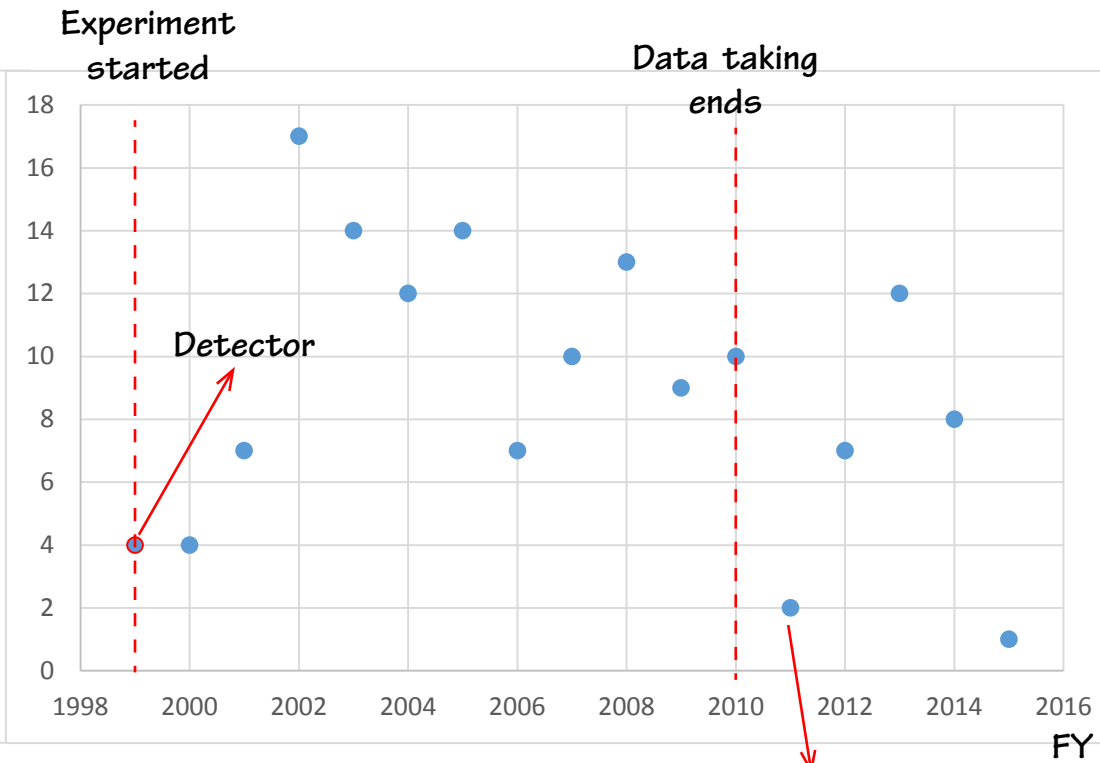
Belle analysis preservation

- The B2BII (Belle → Belle II) conversion project is well on its way

  - possible to read Belle-native mDST files and access the Belle DB within basf2
  - large fraction of the mDST objects have already been converted and the conversion has been validated
  - we aim to convert the remaining objects until June (this month)
  - next step is then to work with Belle II computing group to establish procedure to process Belle Data/MC on the distributed computing

Number of papers / PhD thesis

the number of journal publications

Experiment started

The detector papers are not included

Data taking ends

the number of PhD thesis

Experiment started

Detector

Data taking ends

Big earthquake? (March 11, 2011)

The Belle software library is maintained even after the data taking finished
code is alomost fixed...

but it works on SL5 (it also works on SL6, but everyone likes to stick on SL5.....)

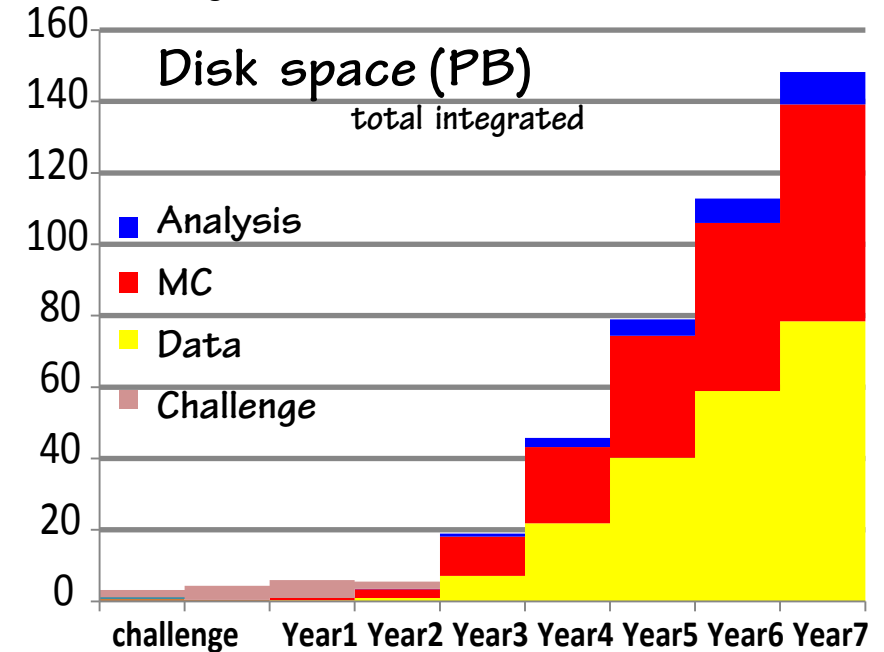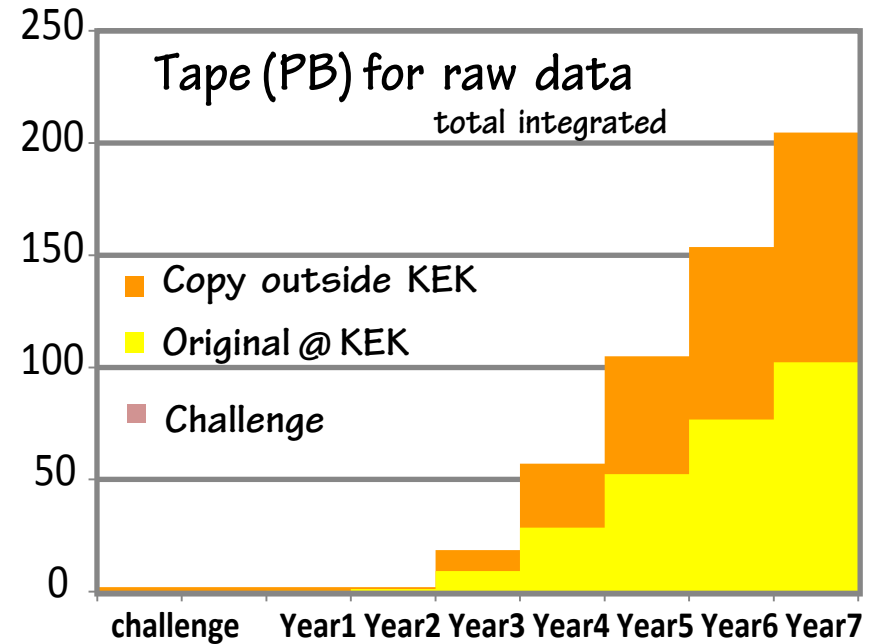The grand-reprocess was done in 2009 with better reconstruction tools (tracking, ECL clusering)

Belle II Data

version estimated in early 2014

uncertainties   Performance of accelerator

beam background condition

improvement of software

The yearly profile may change

The total at the last year should stay the same level

Tape (PB) for raw data

total integrated

- Copy outside KEK
- Original @ KEK
- Challenge

CPU (kHEPSpec)

total integrated

- Analysis
- MC (reproduce)
- MC
- Data (reprocess)
- Data
- Challenge

Disk space (PB)

total integrated

- Analysis
- MC
- Data
- Challenge
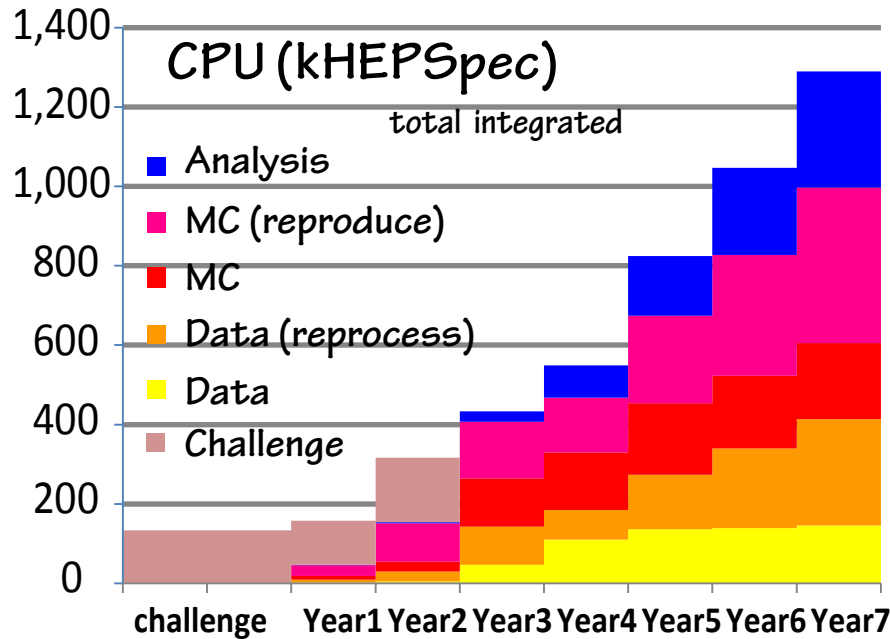
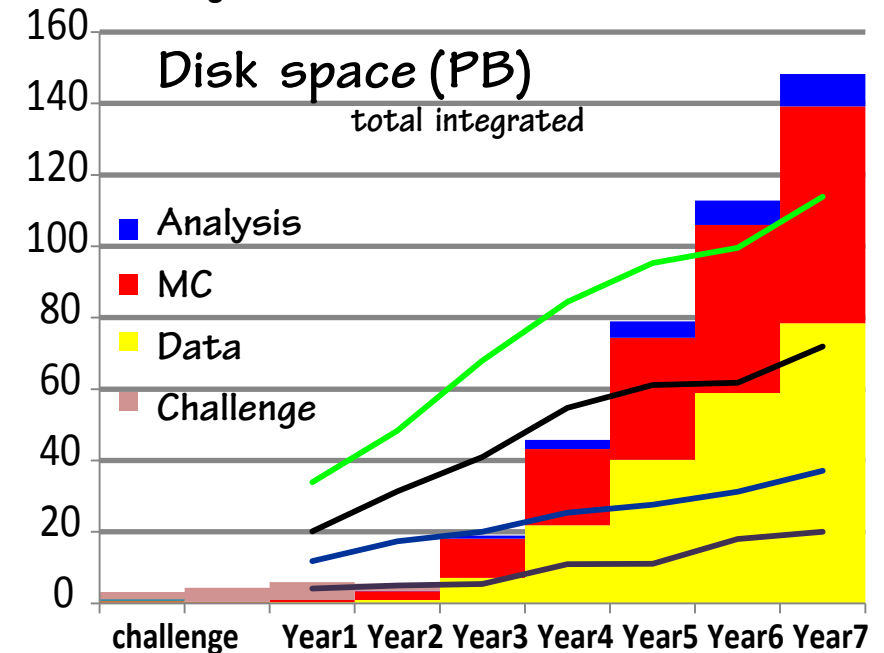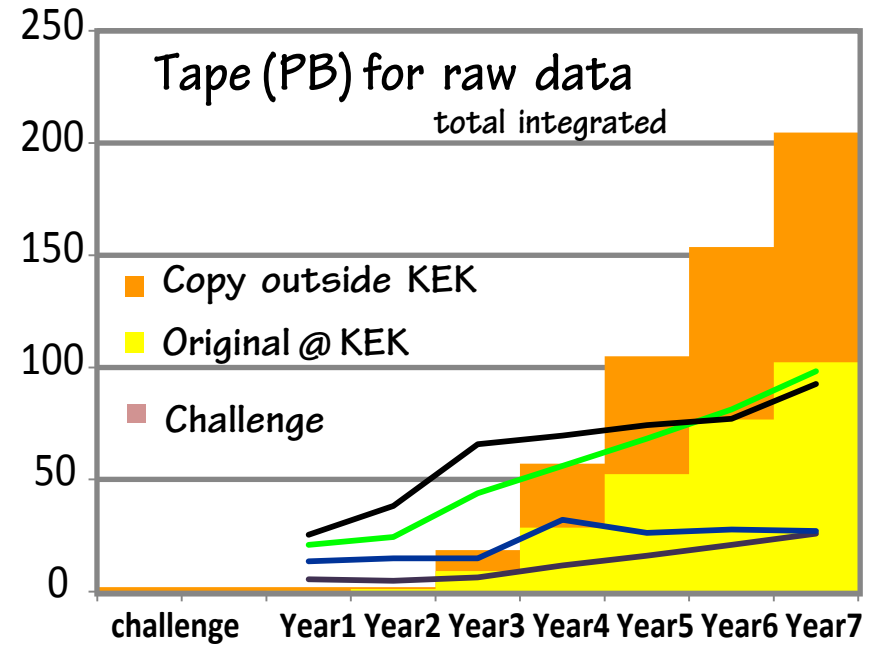# Belle II Data

LHC resources based on the published pledges

http://wlcg-rebus.cern.ch/apps/pledges/summary/

The real capacities and the usages can be different

Tape (PB) for raw data
total integrated

- Copy outside KEK
- Original @ KEK
- Challenge

challenge Year1 Year2 Year3 Year4 Year5 Year6 Year7

CPU (kHEPSpec)
total integrated

- Analysis
- MC (reproduce)
- MC
- Data (reprocess)
- Data
- Challenge

— ATLAS
— CMS
— ALICE
— LHCb

challenge Year1 Year2 Year3 Year4 Year5 Year6 Year7

Disk space (PB)
total integrated

- Analysis
- MC
- Data
- Challenge

challenge Year1 Year2 Year3 Year4 Year5 Year6 Year7

Belle II Computing Model

end of year 3

Belle II

Detector

Raw data storage and processing

Raw data duplex. processing

KEK Data Center

PNNL Data Center

Raw data

mdst Data

mdst MC

dashed inputs for Ntuple

CPU

Disk

Tape

Raw Data Center

raw data storage and (re)process

mdst storage

Asia

Europe 1

Europe 2

Regional Data Center

MC production and Physics analysis skim

GRID site

Cloud site

Computer cluster site

MC production site

user analysis (Ntuple level)

Local resource

Belle II's policy

has not been discussed seriourly yet...

To start taking data is the priority now.

only the "simple" data management plan is available
https://belle2.cc.kek.jp/~twiki/bin/view/Public/DataManagementOpenPage

http://science.energy.gov/funding-opportunities/digital-data-management/

- *"The focus of this statement is sharing and preservation of digital research data"*

- **All proposals submitted to the Office of Science (after 1 October 2014) for research funding must include a Data Management Plan (DMP) that addresses the following requirements:**

1. **DMPs should describe whether and how data generated in the course of the proposed research will be shared and preserved.**

   If the plan is not to share and/or preserve certain data, then the plan must explain the basis of the decision (for example, cost/benefit considerations, other parameters of feasibility, scientific appropriateness, or limitations discussed in #4).

   At a minimum, DMPs must describe how data sharing and preservation will enable validation of results, or how results could be validated if data are not shared or preserved.

U.S. DEPARTMENT OF **ENERGY** | Office of Science