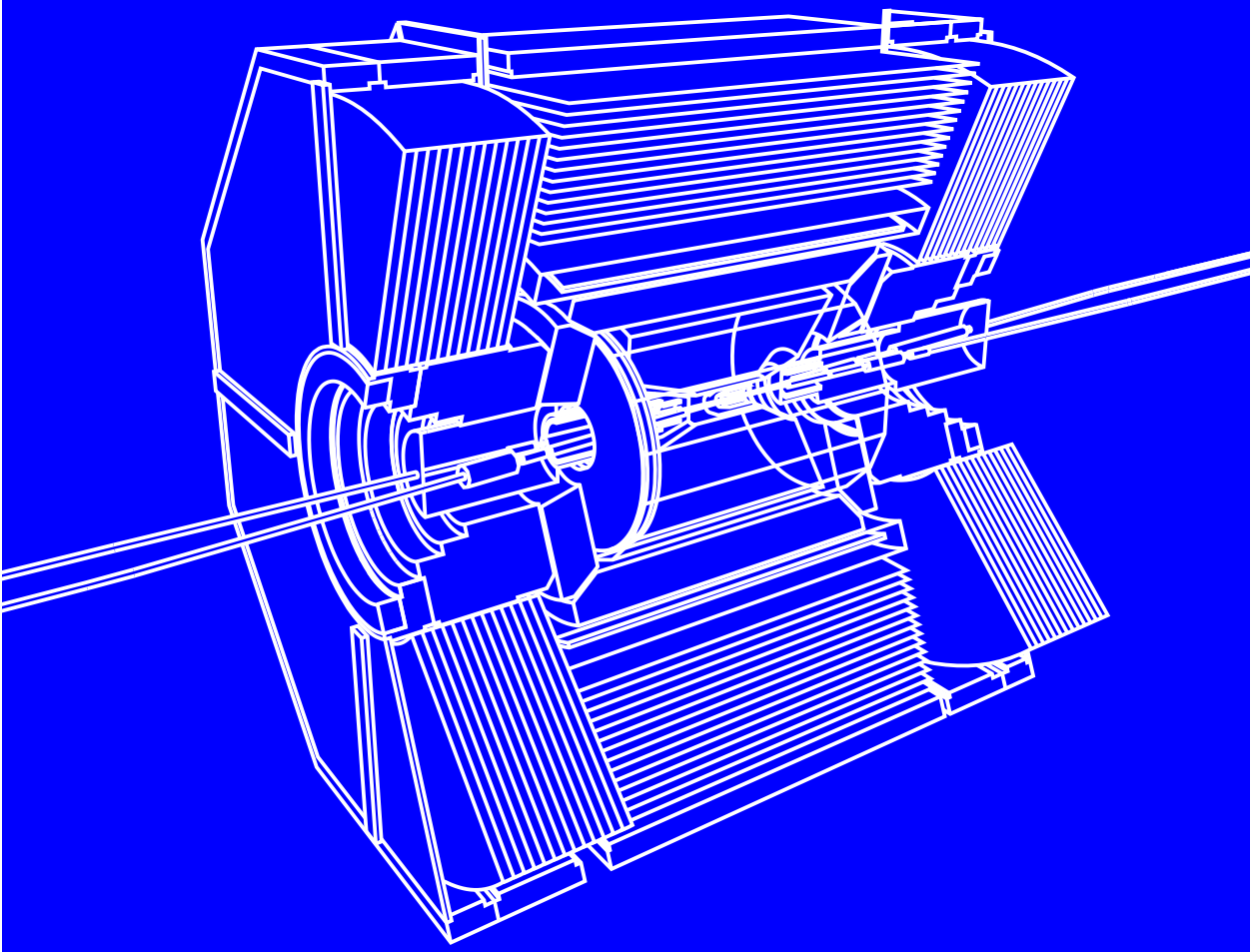
LCWS13 @ Tokyo

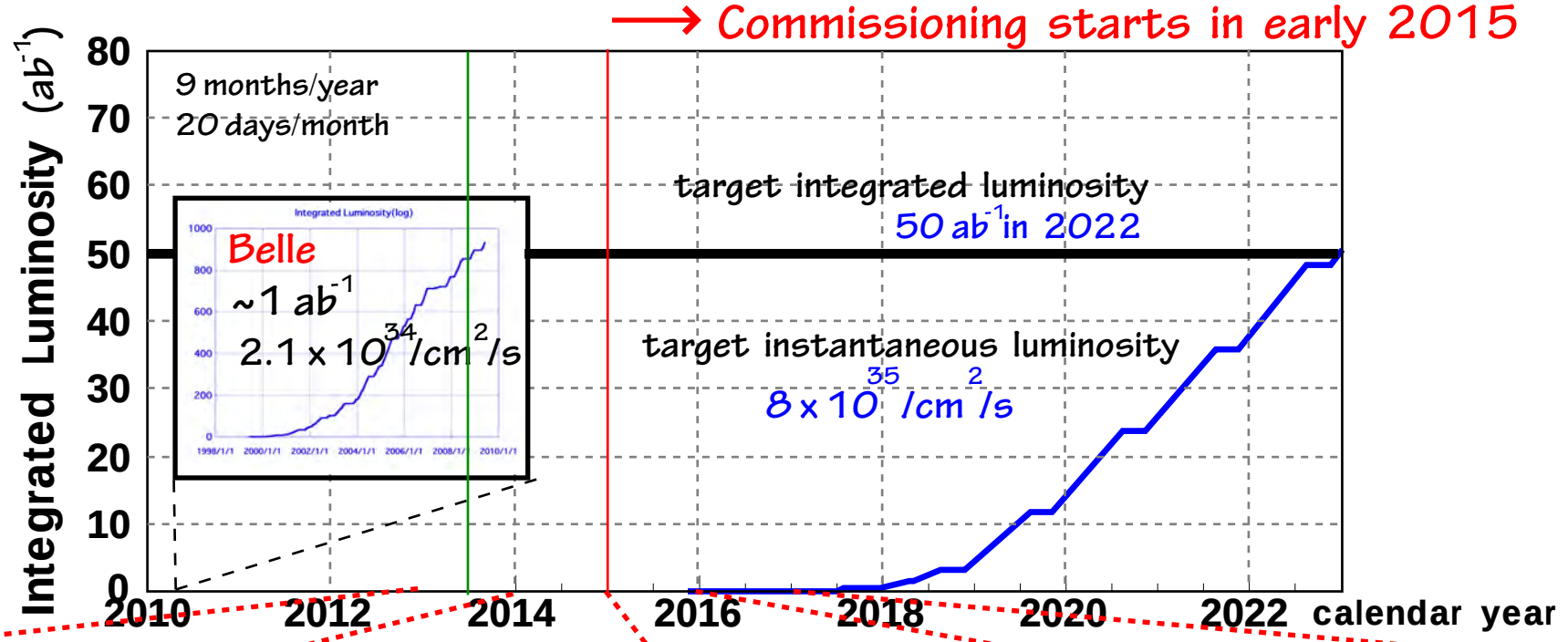
BelleII software and computing

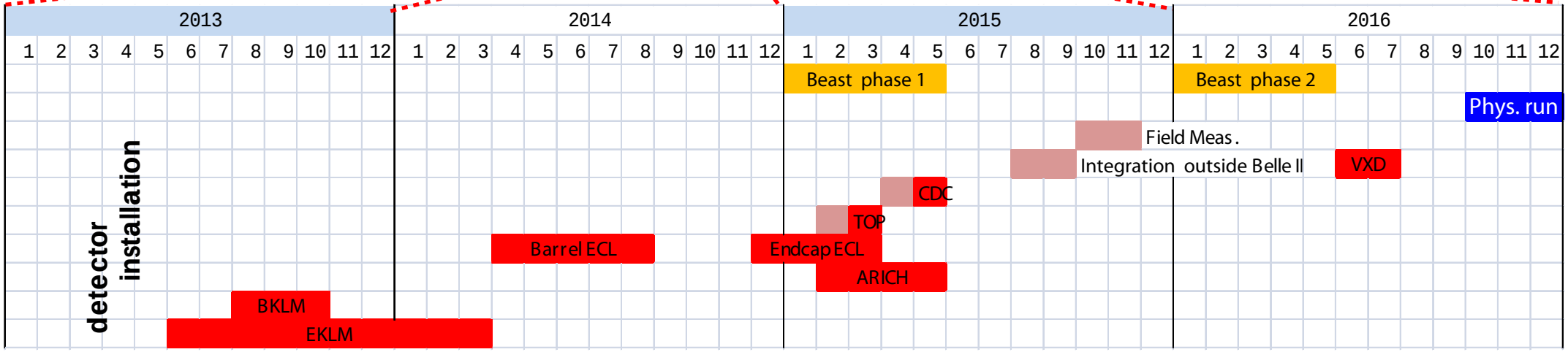


Takanori Hara (KEK)

14 Nov., 2013

Lumi. Prospect + const. schedule

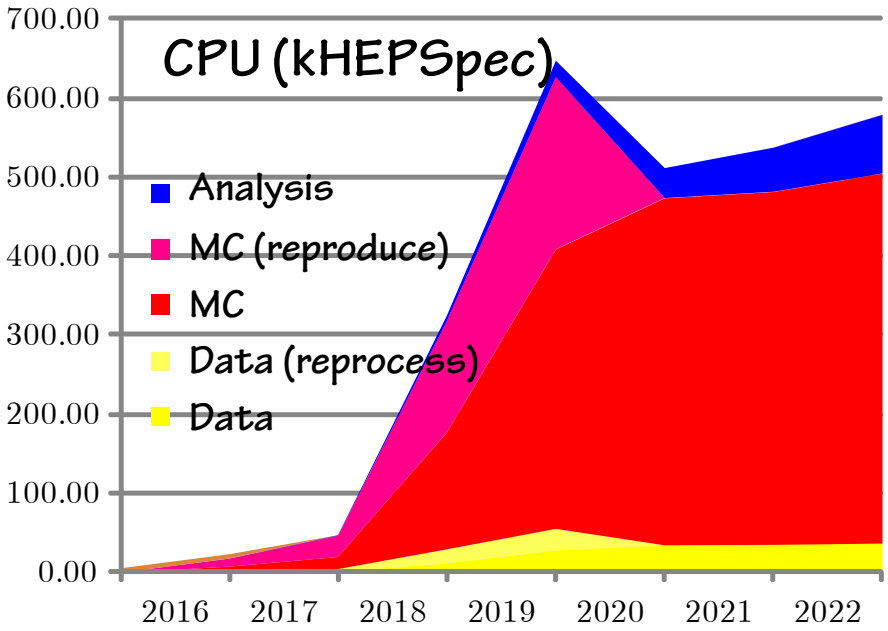


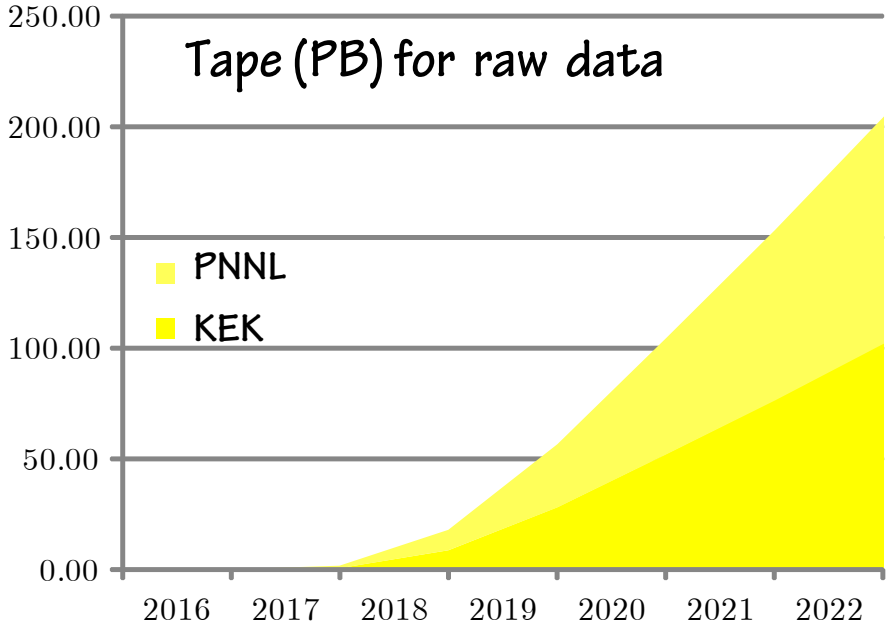


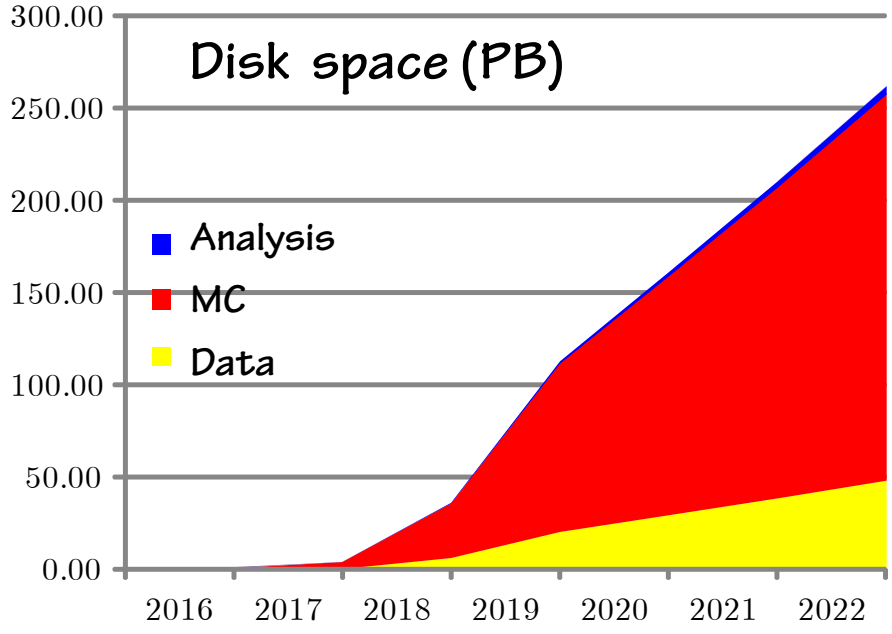
Hardware Resources for Belle II

Preliminary estimates depend on many unknown parameters

- . accelerator performance
- . data reduction
- . performance of simulation/reconstruction
- . analysis requirements, ...



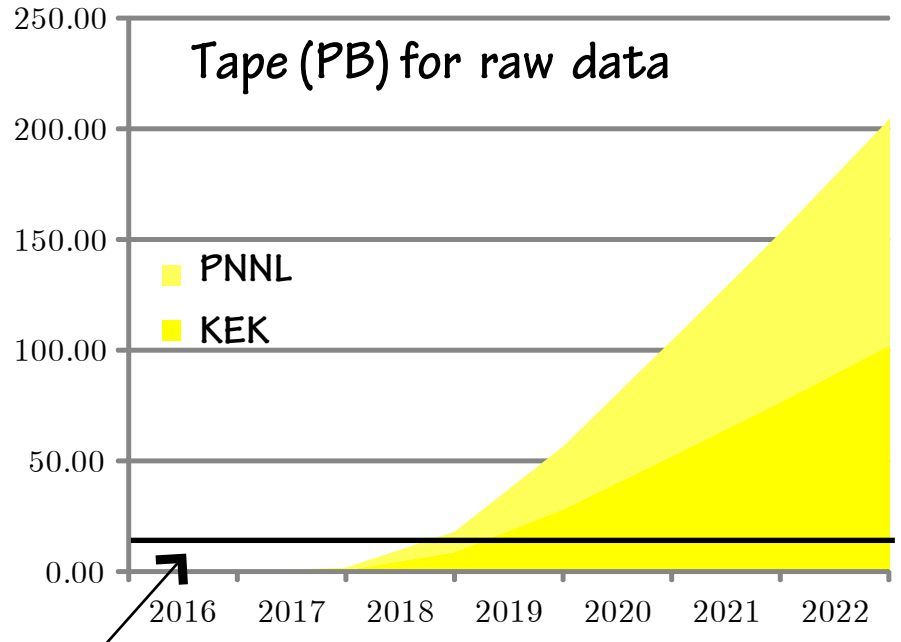


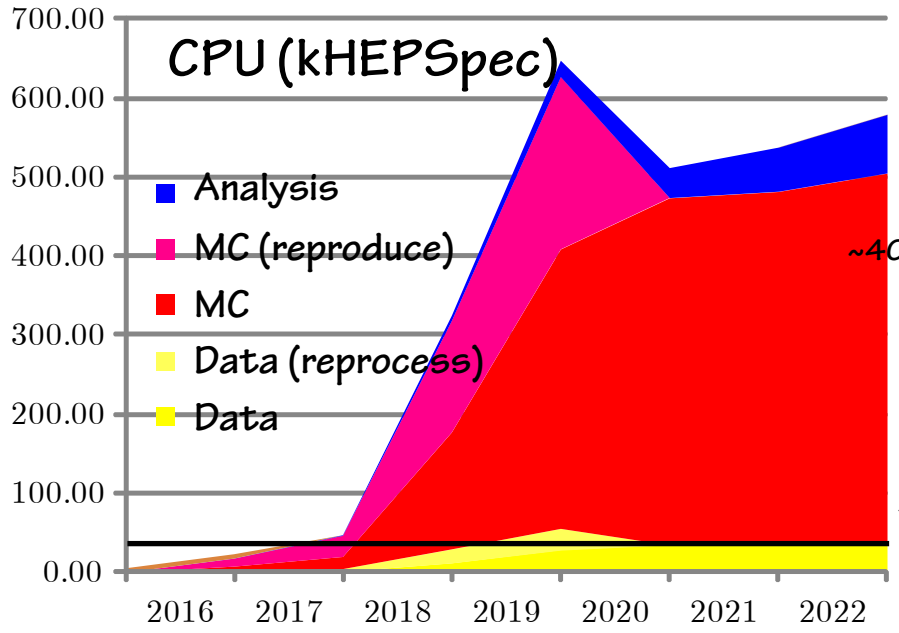


Hardware Resources for Belle II

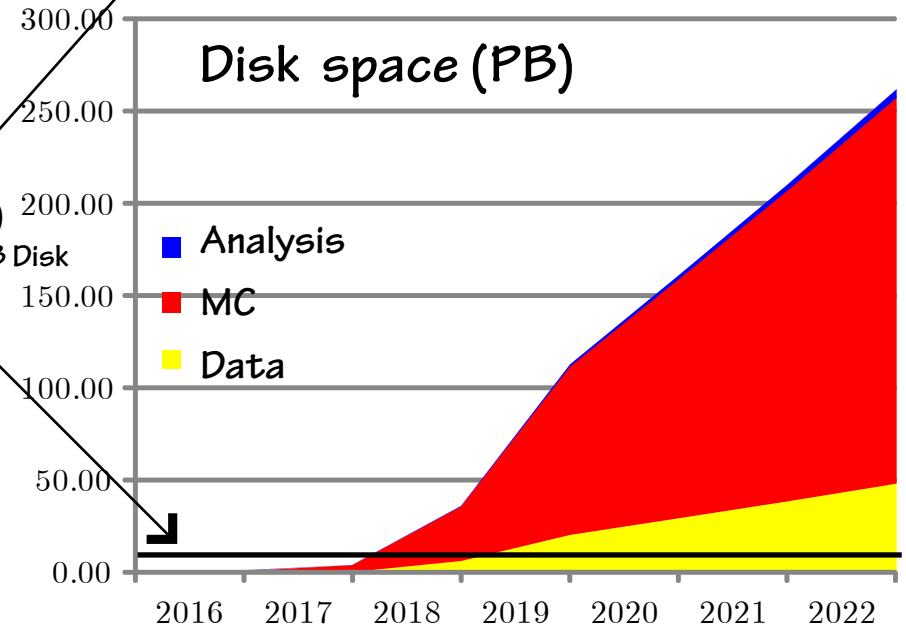
Preliminary estimates depend on many unknown parameters

- . accelerator performance
- . data reduction
- . performance of simulation/reconstruction
- . analysis requirements, ...





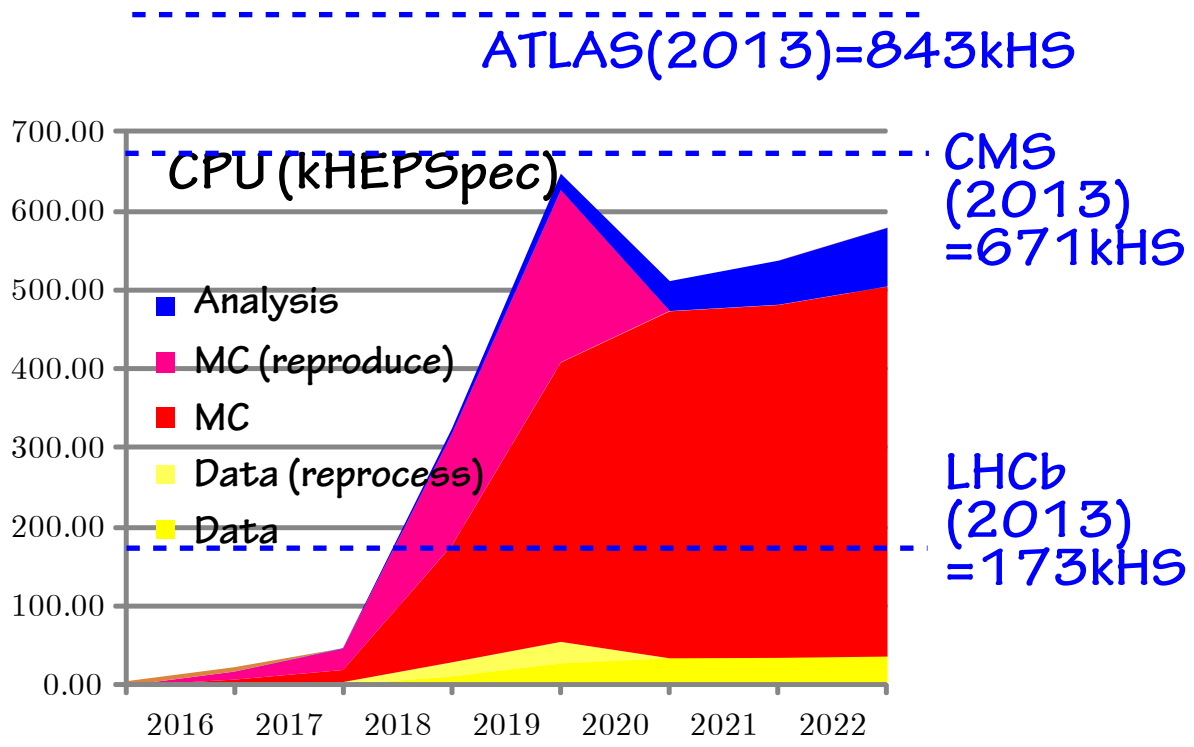
KEKCC resource
(for all projects)
~40kHS, 16PB Tape, 7PB Disk

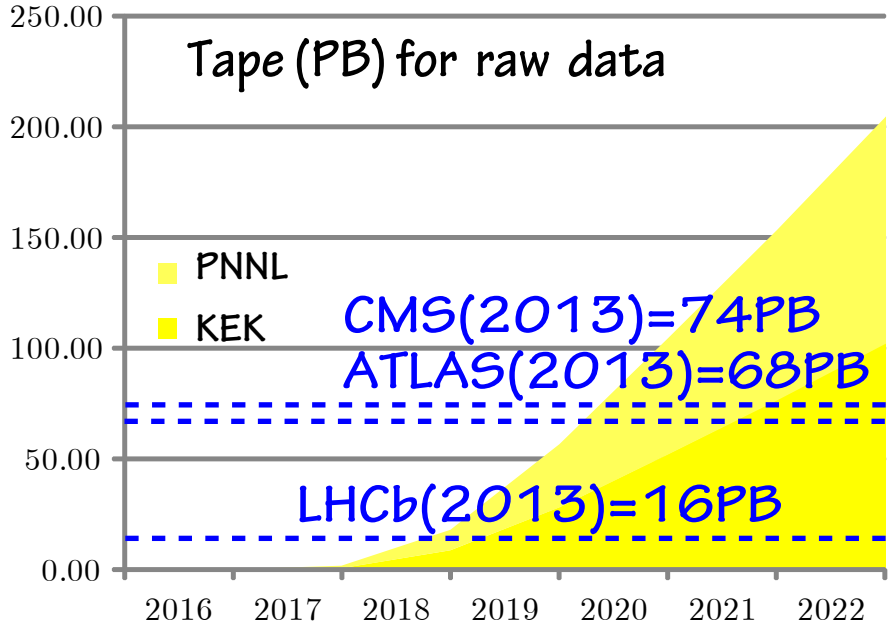


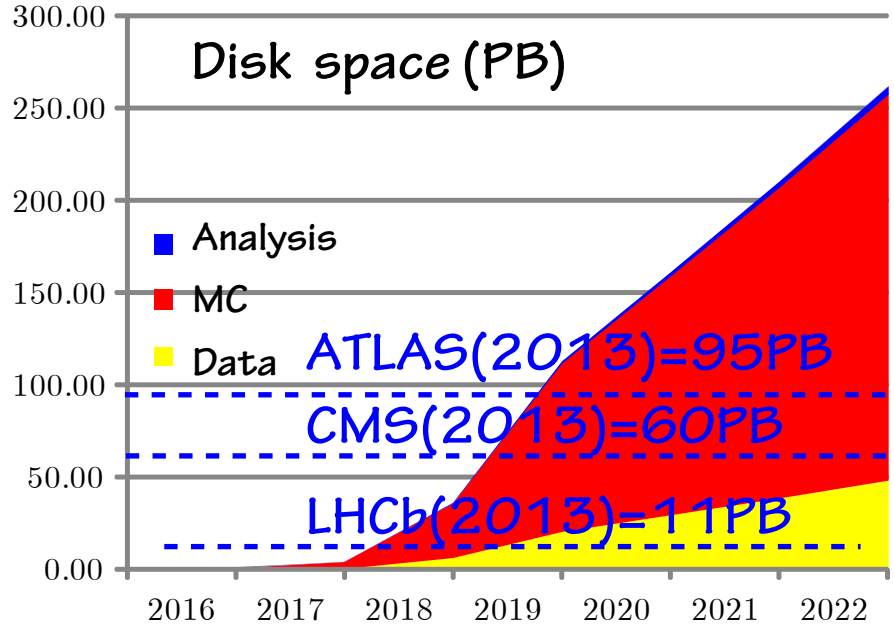
Hardware Resources for Belle II

Preliminary estimates depend on many unknown parameters

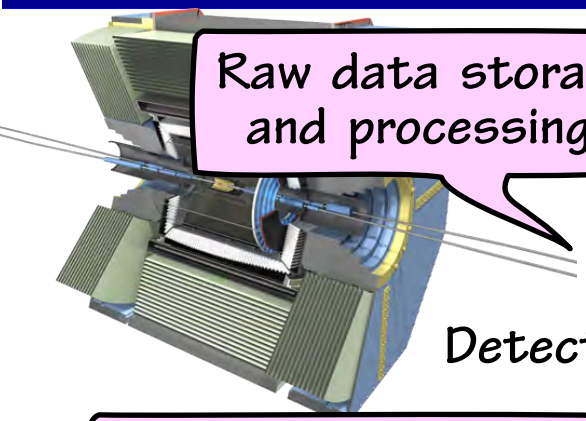
- . accelerator performance
- . data reduction
- . performance of simulation/reconstruction
- . analysis requirements, ...







Belle II Computing Model



Detector

Raw data storage and processing

KEK

PNNL

Raw data duplex. reprocessing

Raw Data

mDST Data

Disk

mDST MC

Ntuples

MC production and Ntuple production

Grid Site

Grid Site

Cloud

MC production (optional)

Ntuple Analysis

Local Resources

Local Resources

Local Resources

Local Resources

Do ntuple analysis on local resources

Software framework (basf2)

Software framework of BELLE: **BASF** (**B**elle **A**nalysi**S** **F**ramework)

Successfully used since over **10 years** for

Online

- **HLT** (High level trigger)
- **DAQ** (Data acquisition)
- **DQM** (Data quality monitor)

Offline

- BELLE detector **optimization**
- Physics **analysis**

Pros & Cons

Pros	Cons
Proven concepts	Highly optimized for BELLE
Combines a lot of knowledge	Lack of object persistency (not object oriented)

Software framework (basf2)

The **design** and **architecture** of the Belle II framework is driven by:

Resources:

Take the best ideas/concepts from other frameworks
(BASF, ILC, Gaudi, CDF, ALICE)

Use proven third-party libraries
(ROOT, boost, libxml)

Modularization
(distribute work, keep maintenance work of the core low)

Technology choices:

Framework core and libraries: **C++**

Steering/Control: **Python**

Parallel processing

Requirements:

Same framework used for both **online** (HLT, DQM)

offline event processing

Software framework (basf2)

Event processing: linear arrangement of **modules**.

A module is the smallest building block in basf2

Typical modules: ***data input***, *geometry input*, *simulation*, *tracking*, ***data output*** ...

Modules live in a **path** (module container)

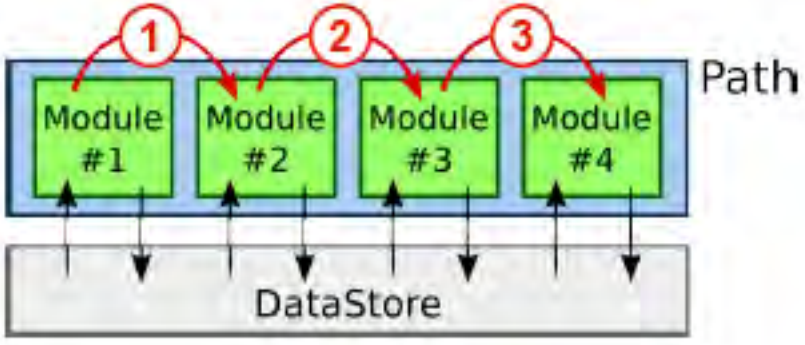
Framework executes one **module** at a time

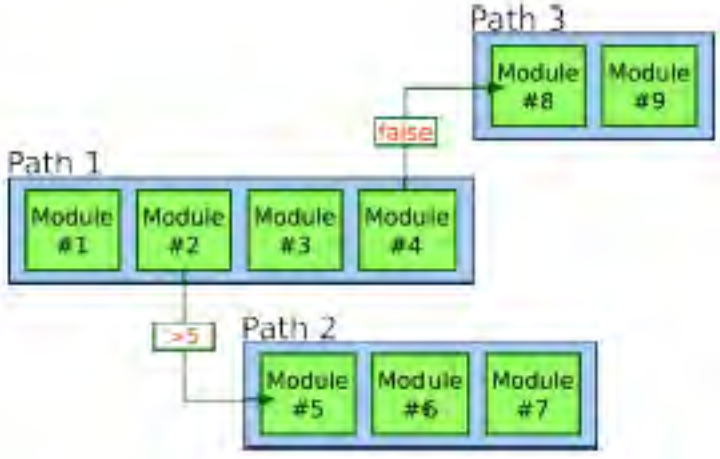
Data **input/output** managed by standard modules

Multiple paths can be connected by **conditions**

Processed event data is stored in a **common storage** (the DataStore)

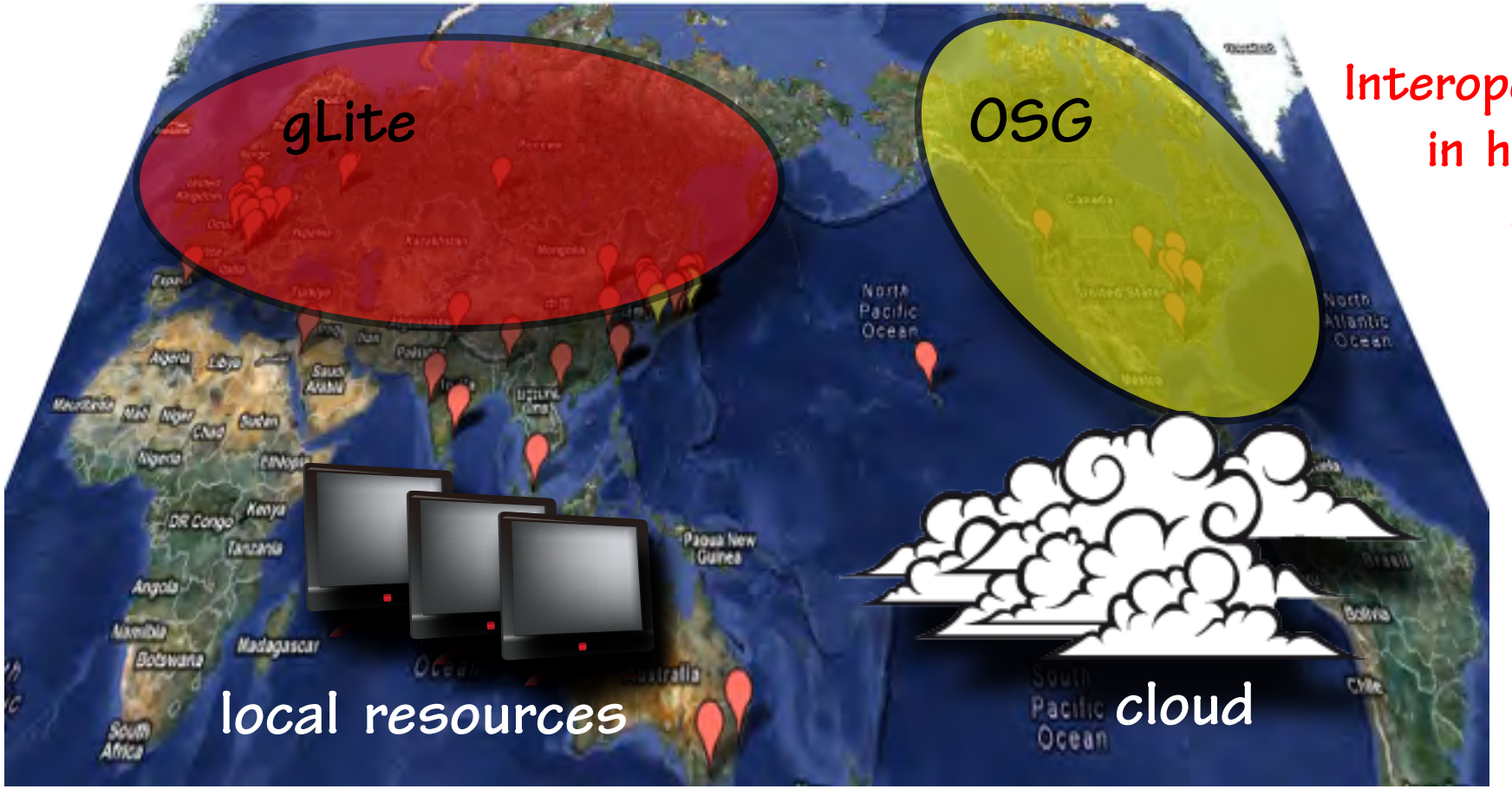
Parallel processing





Distributed Computing Model

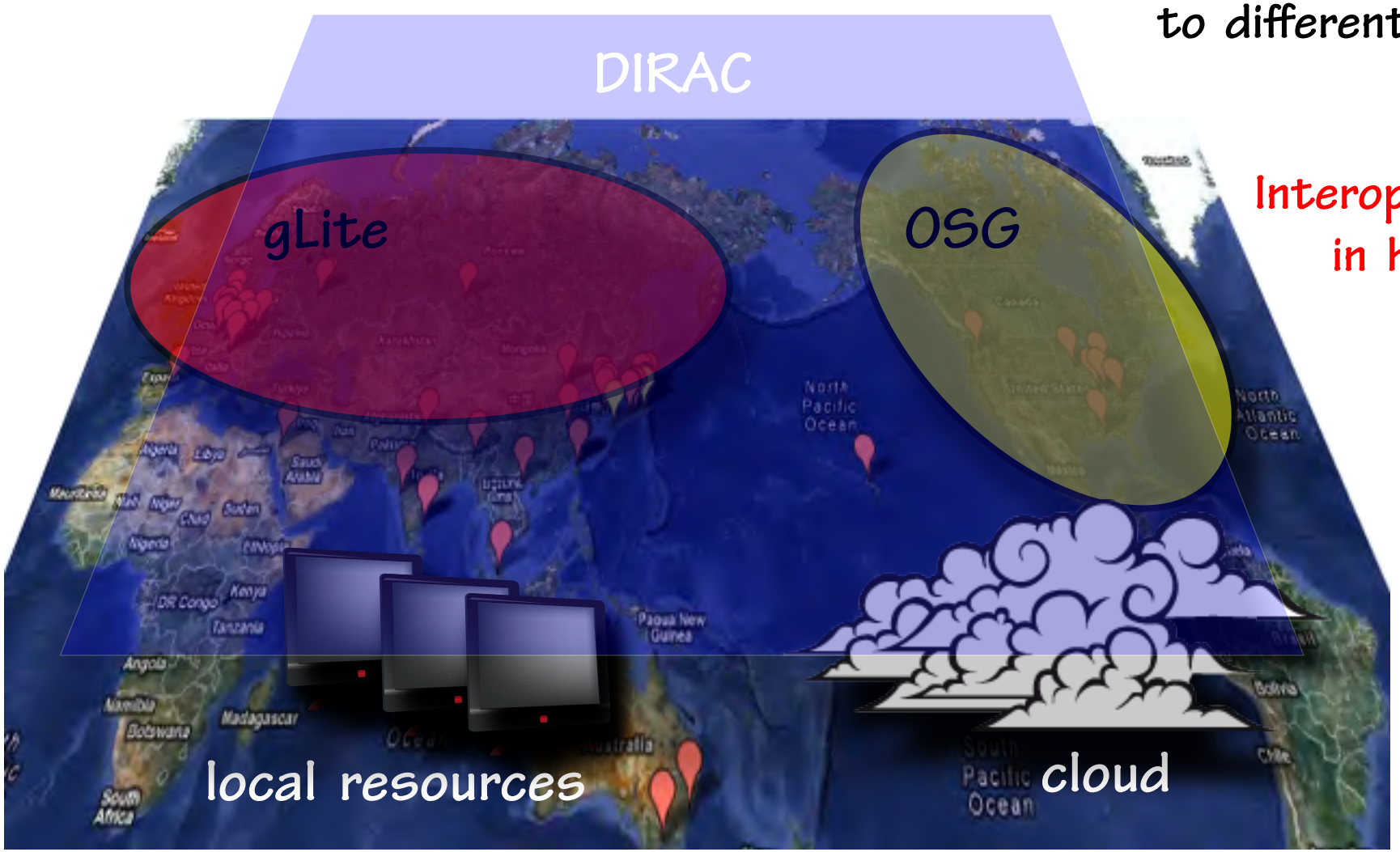
- DIRAC (developed by LHCb) Distributed Infrastructure with Remote Agent Control
  - → Pilot jobs
  - → Modular structure that enabled it possible to submit jobs to different backends.



Interoperability in heterogeneous computings

Distributed Computing Model

- DIRAC (developed by LHCb) Distributed Infrastructure with Remote Agent Control
  - → Pilot jobs
  - → Modular structure that enabled it possible to submit jobs to different backends.



Interoperability in heterogeneous computings

AMGA catalogue

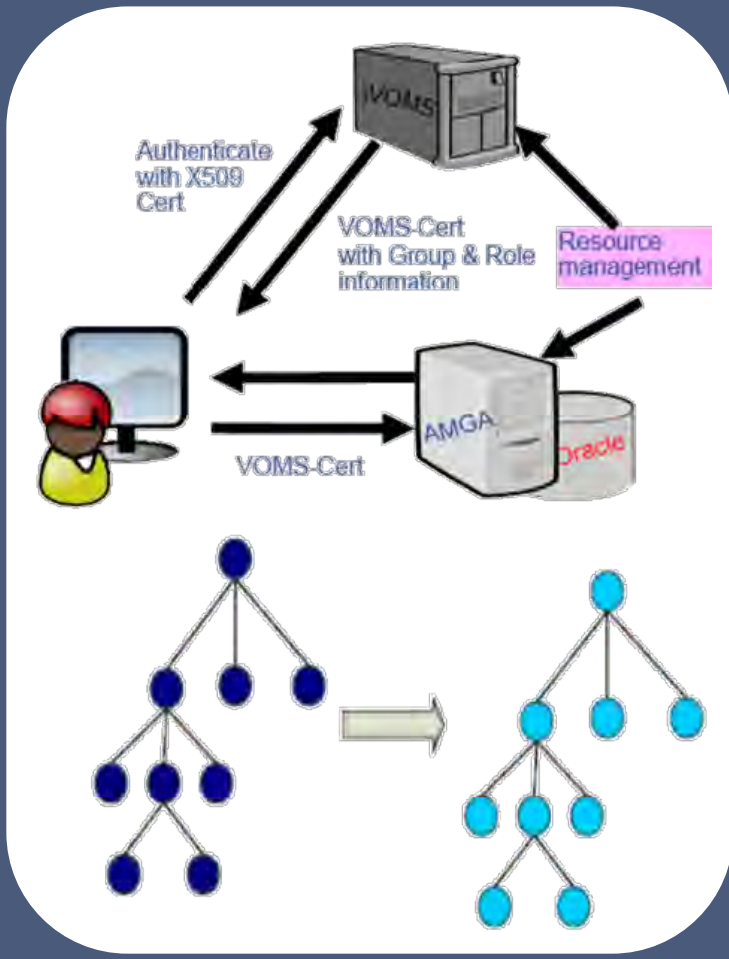
- ARDA Metadata Grid Application
  - Metadata server for GRID environment

Metadata: data of data

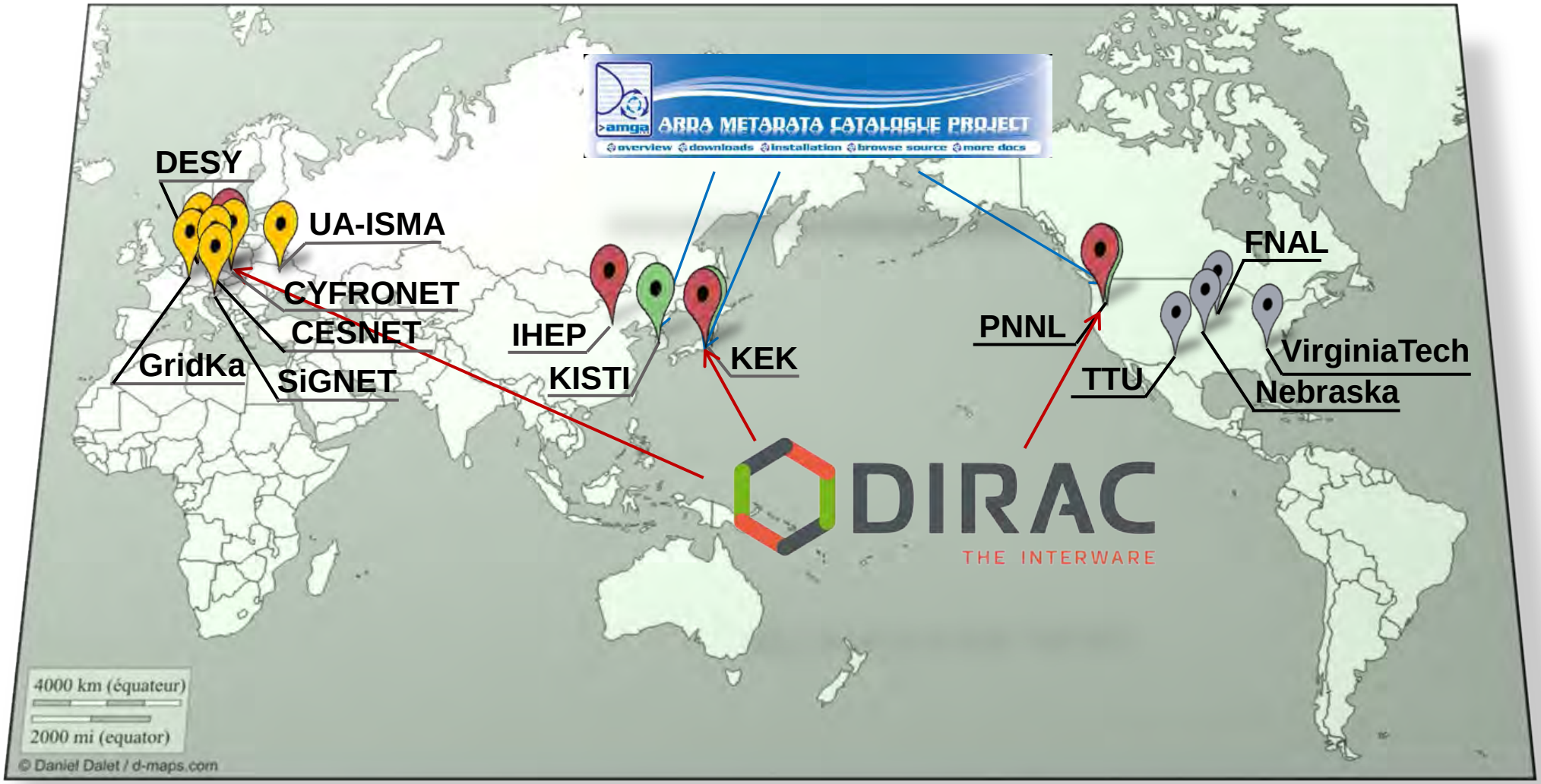
LFN, run range, software version…

- Main feature
  - Integration with GRID security
  - Secure connection using SSL
  - Replication of data
    - Asynchronous and hierarchical

e.g. replication of specific data set or run period for a GRID site



Belle II GRID sites + core servers



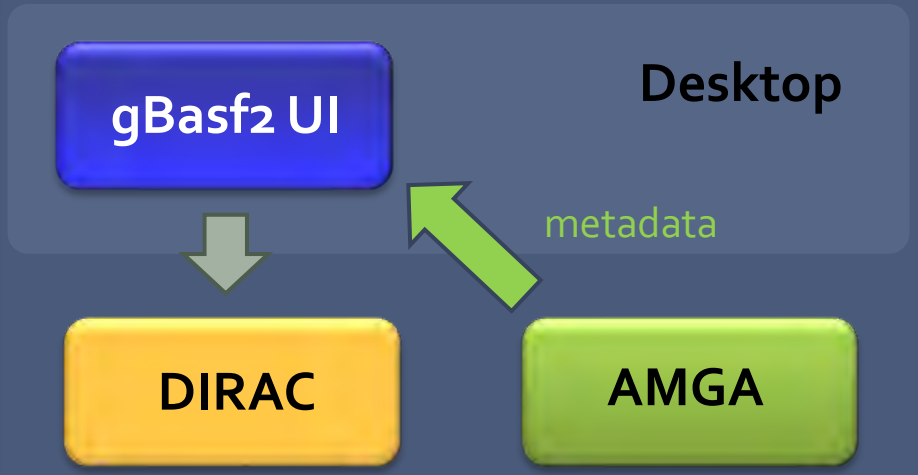
4 DIRAC servers
3 AMGA servers

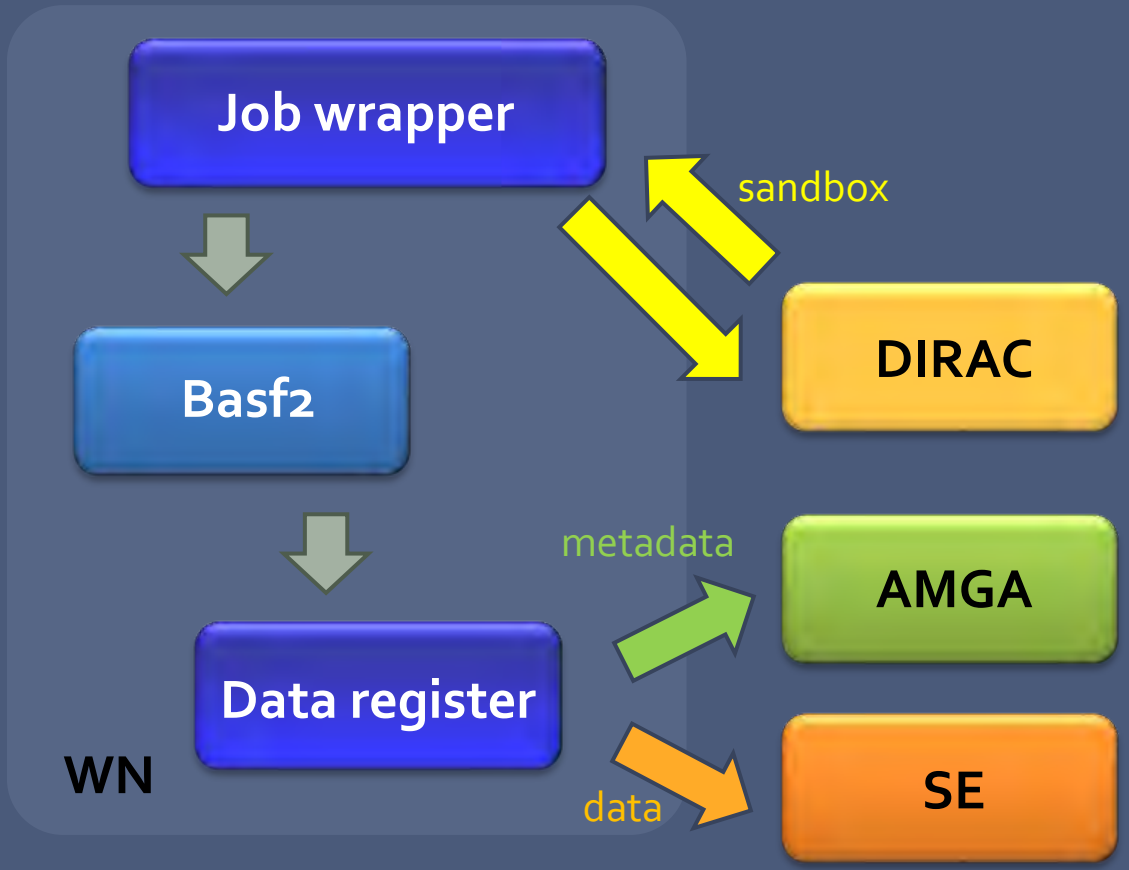
LCG sites
OSG sites

Getting increased

gbasf2

- GRID UI dedicated for Belle II
  - Based on DIRAC API (UI + job wrapper on WN)
  - Not only job submitter but a collection of job and data management tools
  - Provide **transparent user experience** of Belle II standard analysis framework (Basf2)





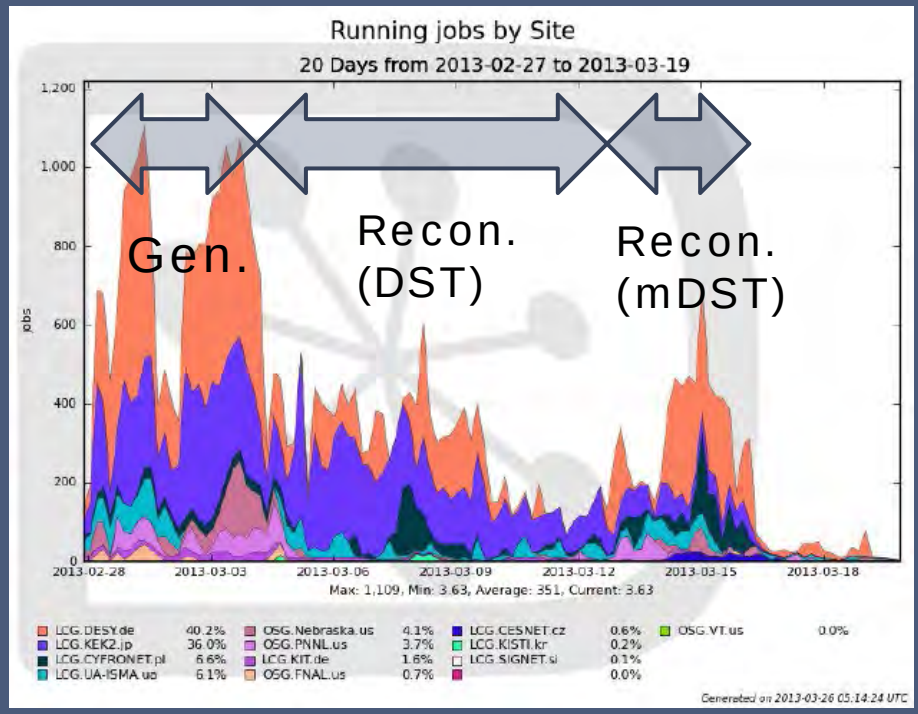

```
% basf2       analysis.py
% gbasf2 –s   analysis.py
```

Software deployment

- Belle II software (basf2 + externals)
  - Statical installation (10GB/release)
  - Installed bellow common software area (e.g. VO_BELLE_SW_DR)
  - Or CVMFS
  - CVMFS sites are getting increased
- DIRAC software (DIRAC + BelleDIRAC)
  - Dynamical installation
  - Installed below working directory
  - gBasf2 has been integrated with BelleDIRAC

First MC prod. camp.

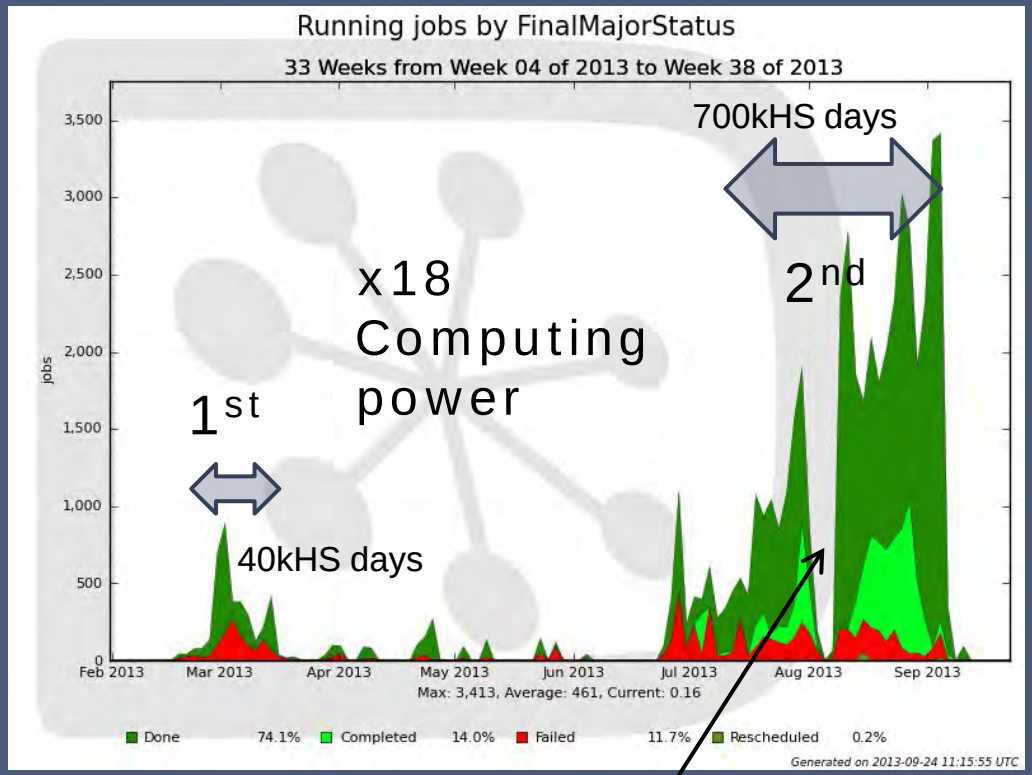
- The 1st iteration of MC mass production using Belle II software on GRID
  - Feb. 28th ~ Mar. 20th, 2013
- The main goal
  - Find possible bottle-necks at everywhere
- Two stages
  - Event generation and detector simulation
  - Reconstruction
- 60M events resulted in 190 TB data (raw level format: DST → high level format: mDST)
- 20% failure rate
  - Metadata registration
  - Input data download
  - Output data upload
  - Application errors



Second MC prod. camp.

- The 2nd iteration of MC has finished
  - July 23rd ~ Sep. 8th , 2013
- More realistic situation
  - Event generation + reconstruction
  - Background mixing
- 560M events resulted in 85 TB data
  - mDST format

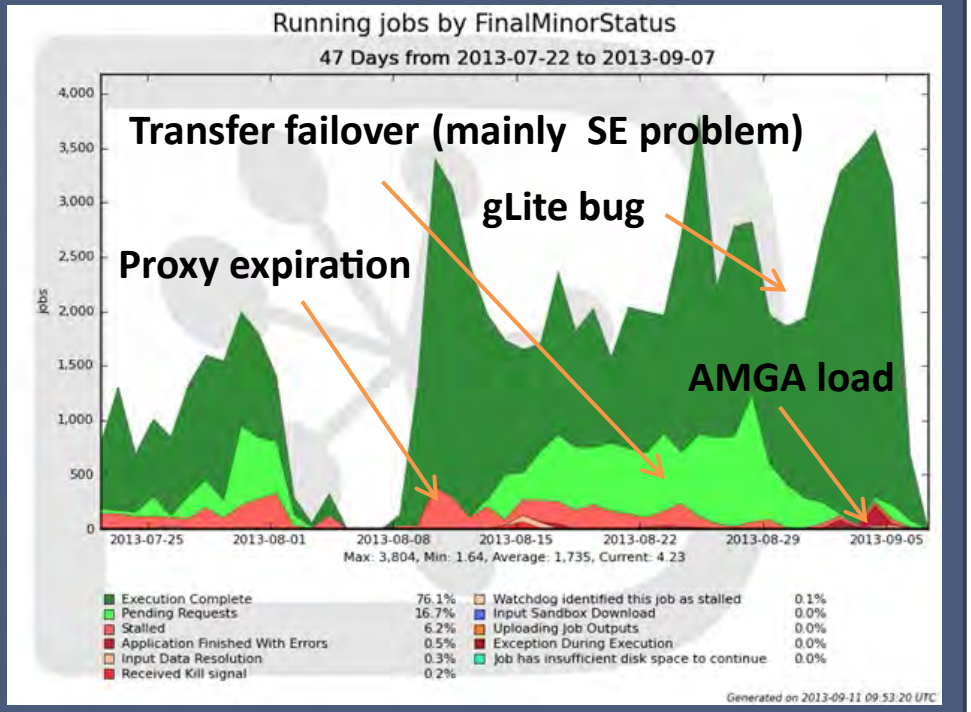
- 10% failure rate
  - Getting decreased through production
  - Final failure rate ~ a few percent
  - No application crash



KEK shutdown

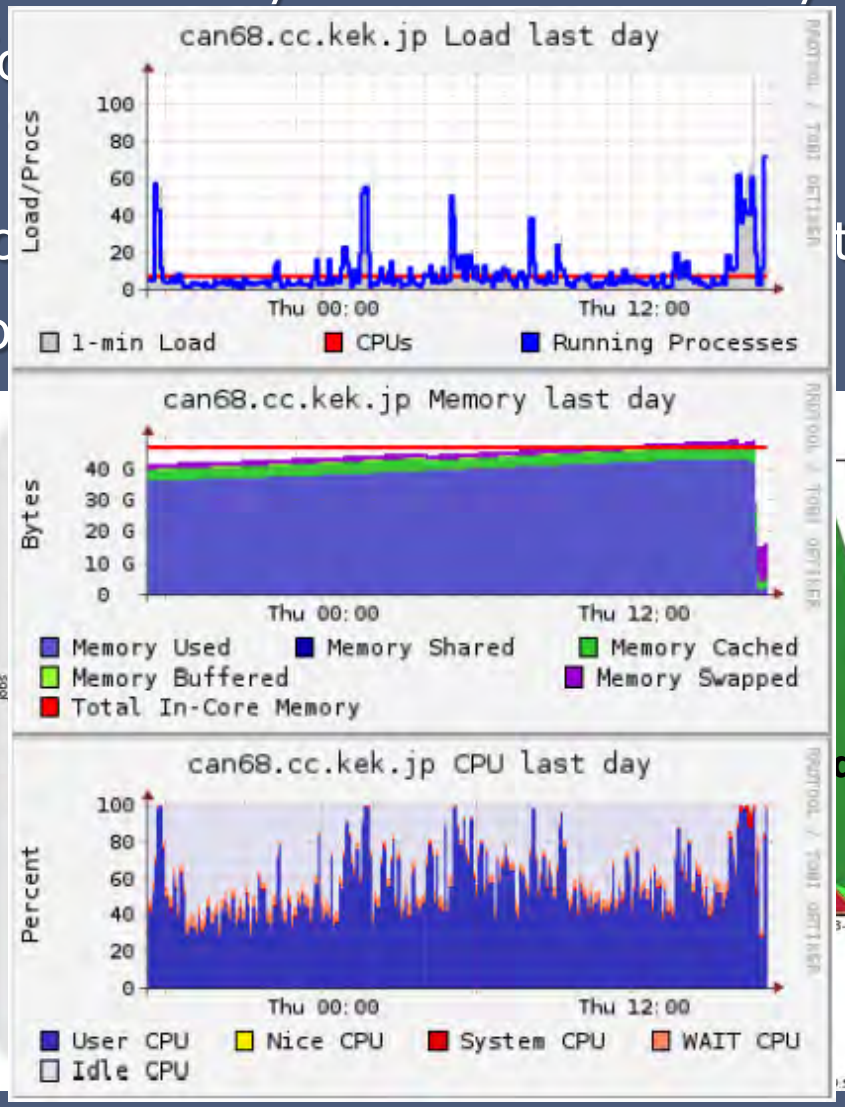
Dominant job failures

- We faced various issues but could solve them one by one thanks to kindly support from each GRID site, especially for KEK
- In early stage, dominant failures came from missing pilot jobs due to short proxy lifetime→ fixed by VOMS server configuration
- Finally dominant problems became our server deployment (exceeded the limit of single server performance)
- Modification of server configuration is ongoing so that we can handle factor 10 more jobs

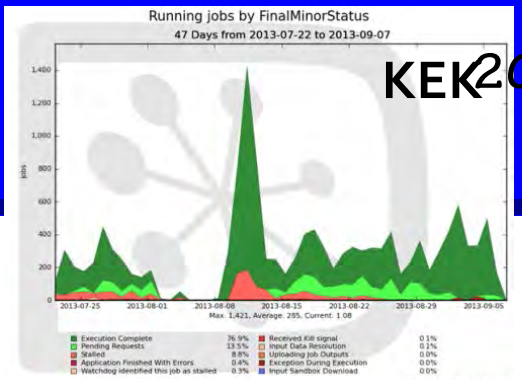


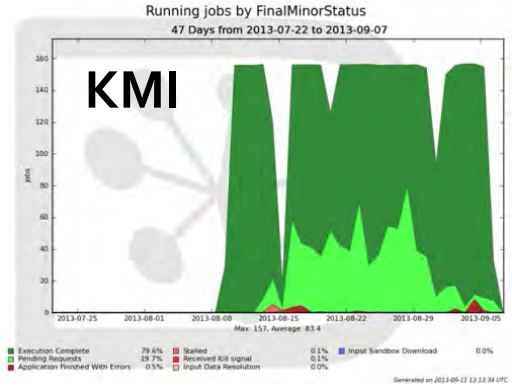
Dominant job failures

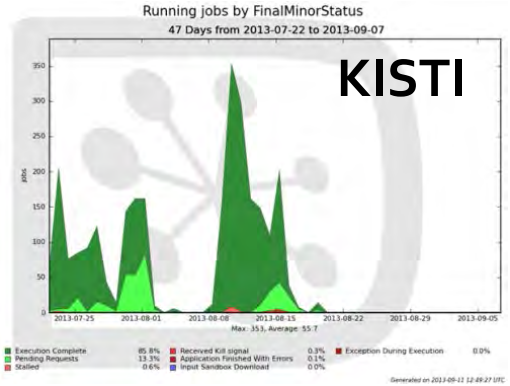
- We faced various issues but could solve them one by one thanks to kindly support from each GRID site, especially fo
- In early stage, dominant failures came fro t proxy lifetime→ fixed by VOMS server co
- Finally dominant problems became our server deployment (exceeded the limit of single server performance)
- Modification of server configuration is ongoing so that we can handle factor 10 more jobs

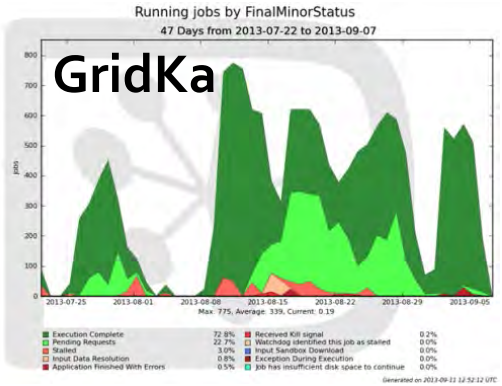


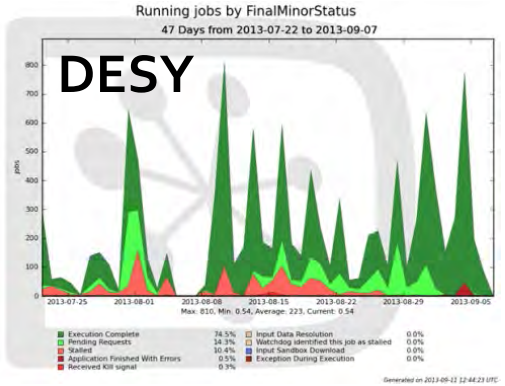
Contributing sites

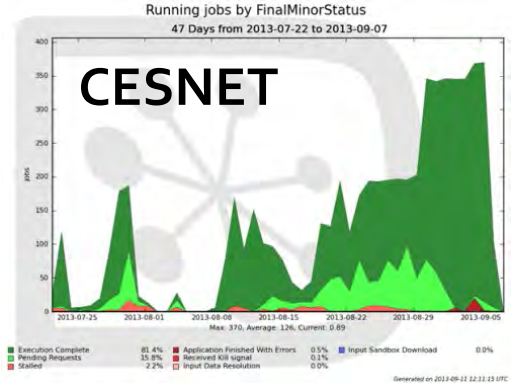


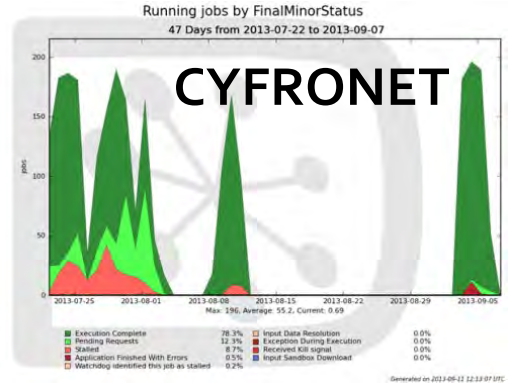


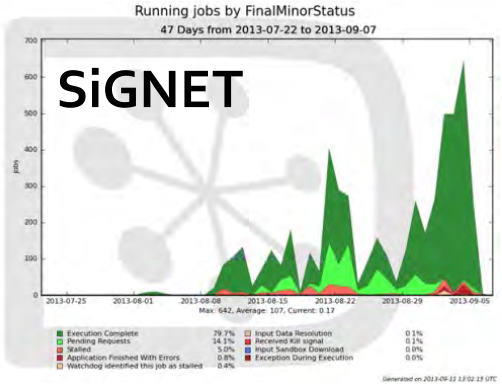


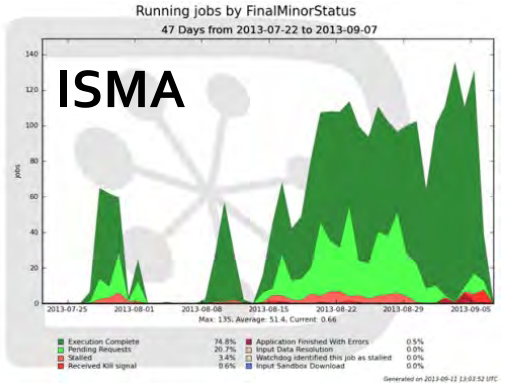


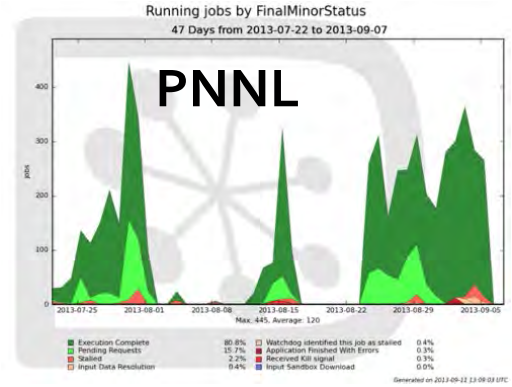


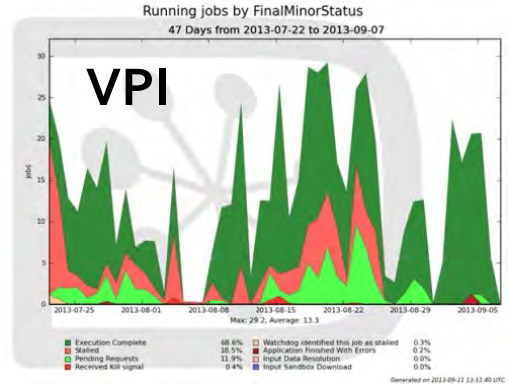


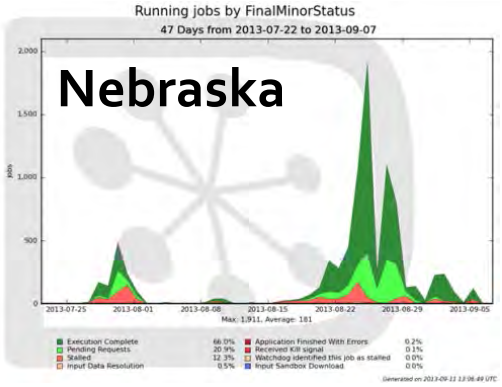


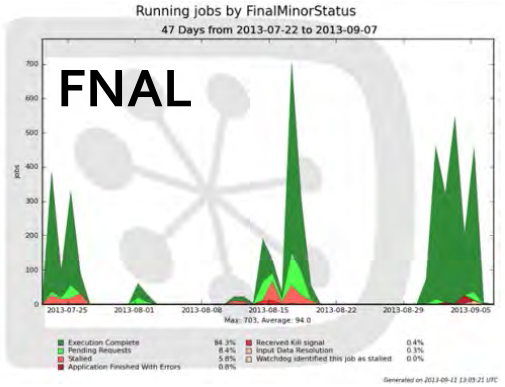






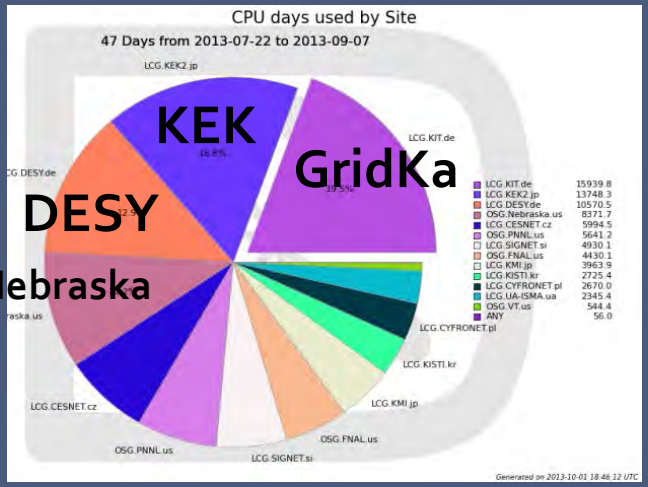


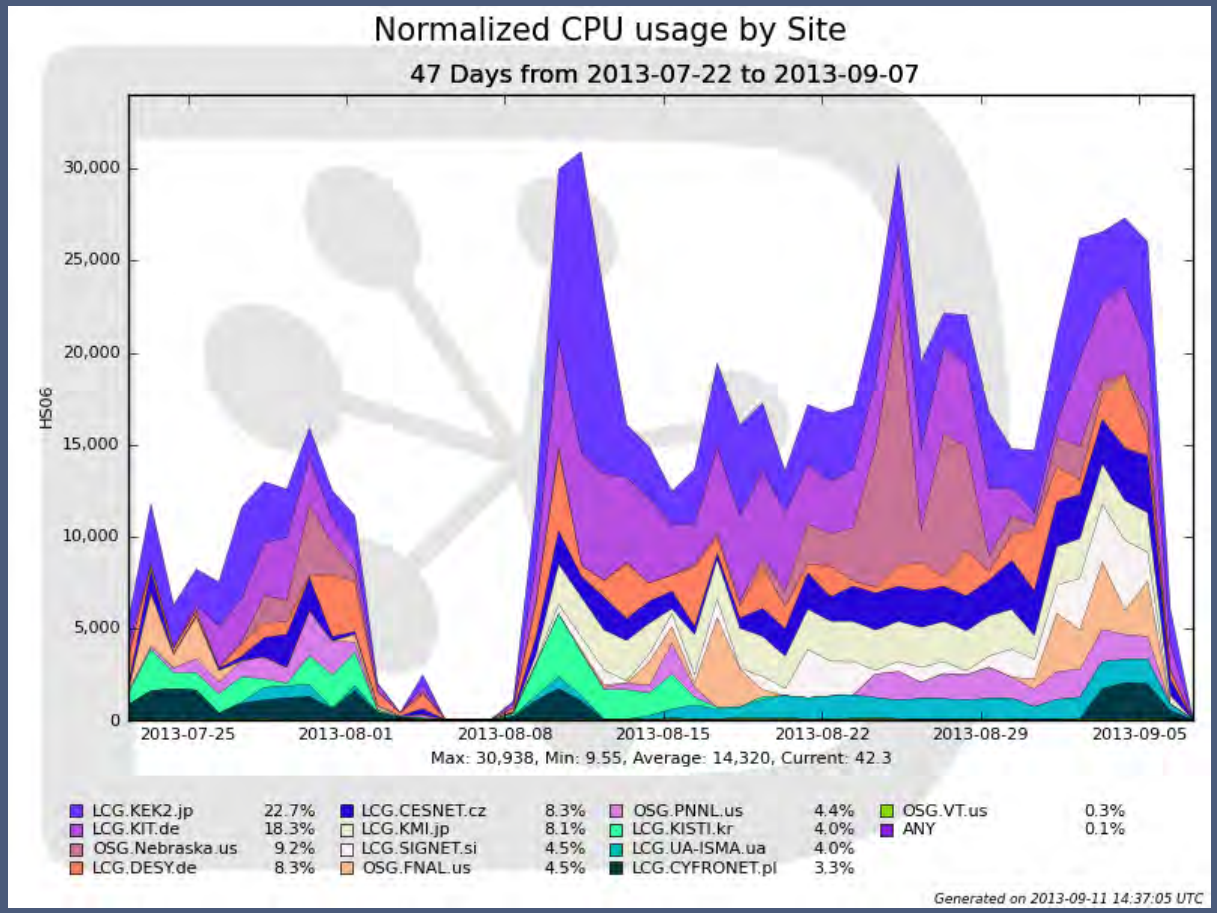




Contributing sites

- Extraordinary contributions were rare
- Variety of GRID sites gave generally constant production rate

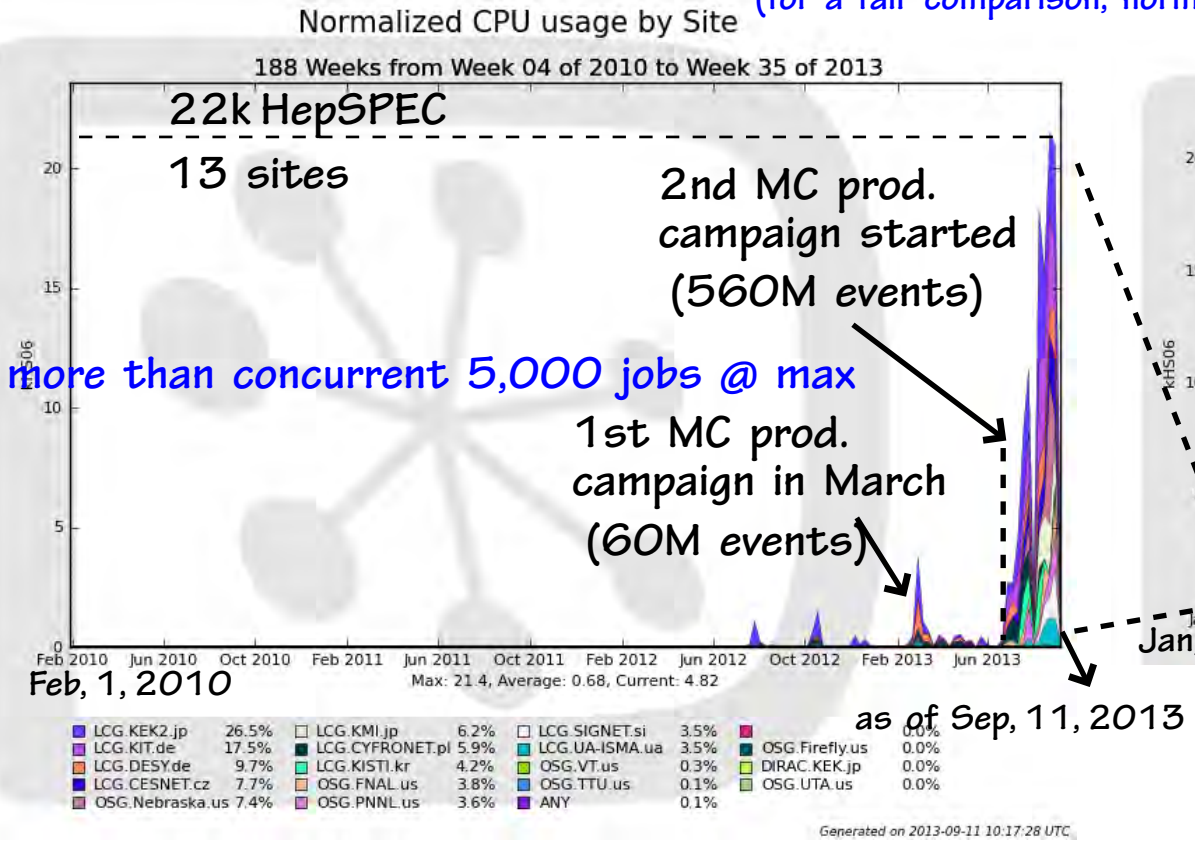




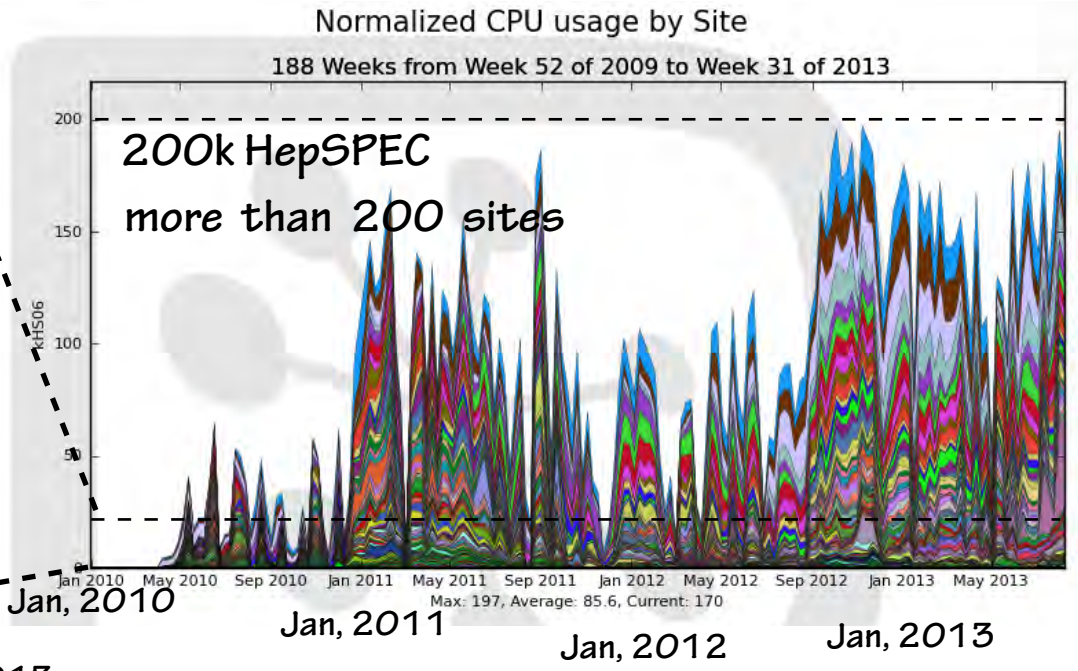
Comparison w/ LHCb

Belle II normalized CPU

(for a fair comparison, normalized for 188 weeks)



LHCb normalized CPU (Jan. 2010-)



we have reached the level of the early stage of LHCb

Thanks to the detailed tuning of DIRAC/AMGA, many MC production's shift workers, stable Belle II software and each site joining the Belle II computing !!

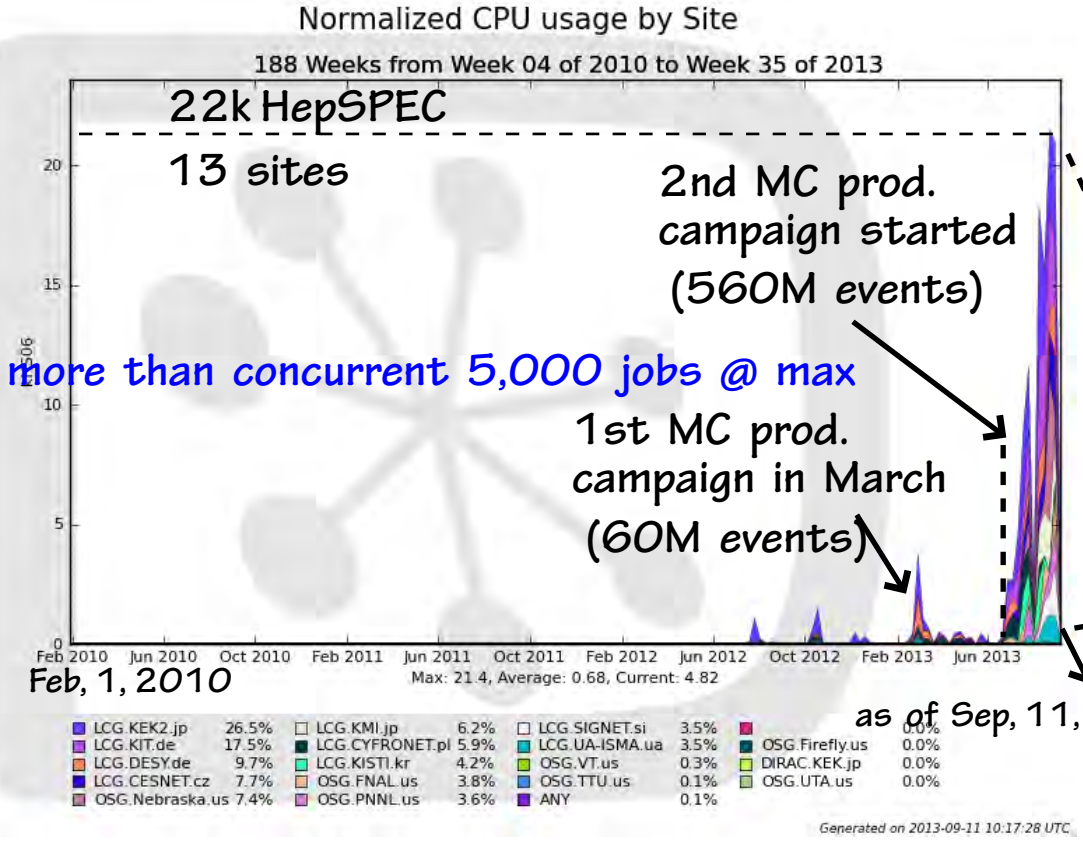
- Storage (Disk, Tape) : ~0.3 PB
  - Only for MC data

Storage (Disk, Tape) : ~26 PB

including Data, MC, User data

Comparison w/ ILC

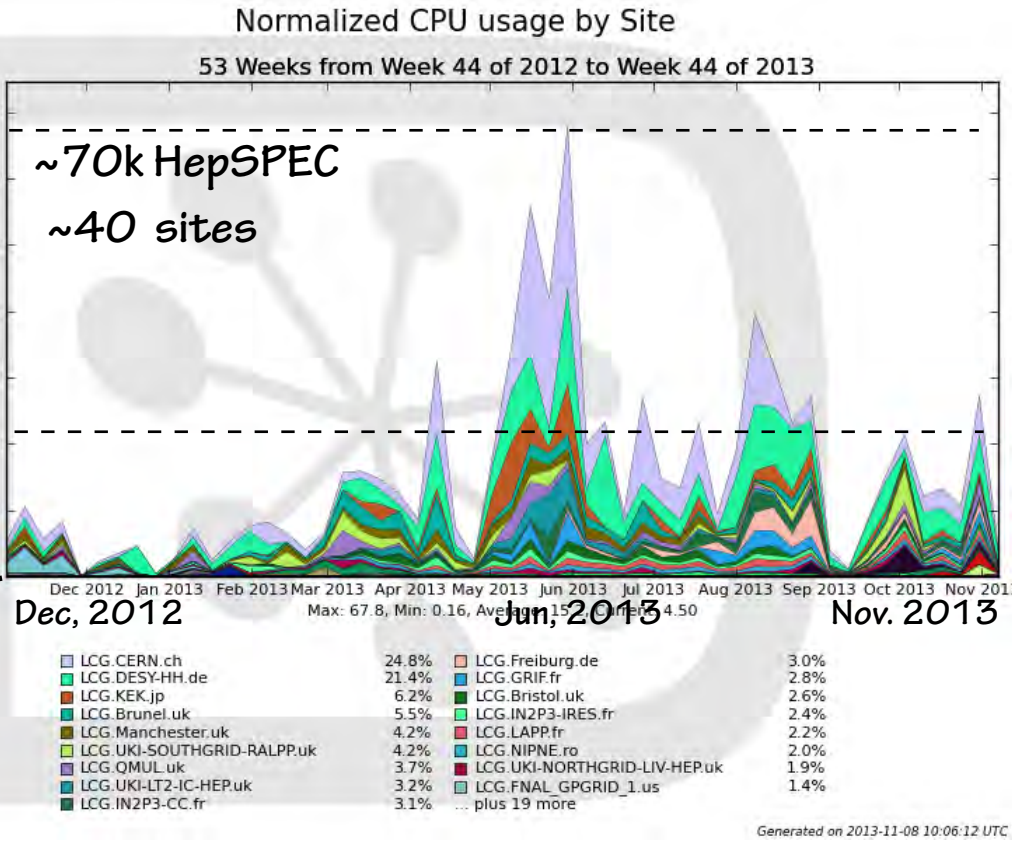
Belle II normalized CPU



Feb, 1, 2010

as of Sep, 11, 2013

ILC normalized CPU (Dec. 2012-)

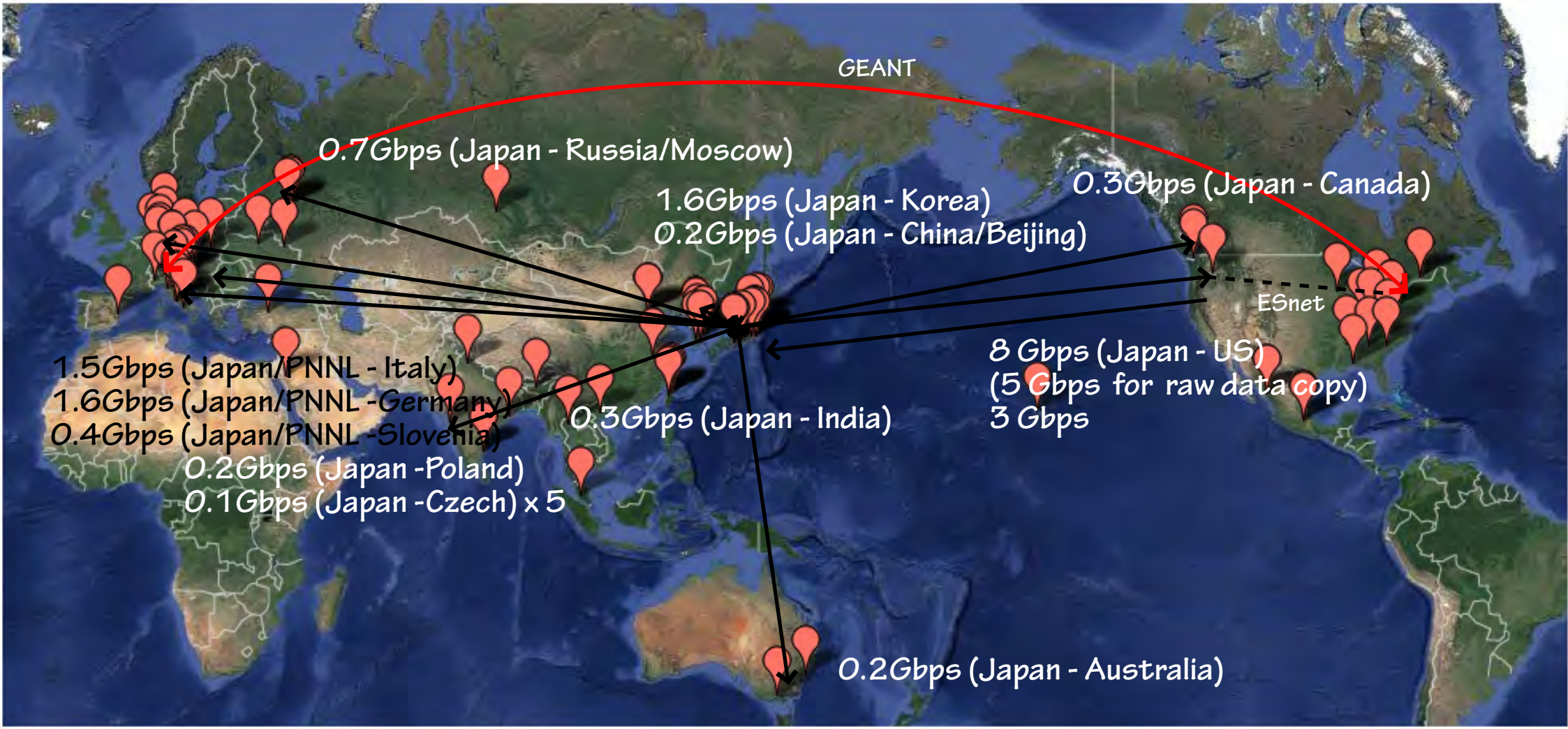


Dec, 2012

Jun, 2013

Nov. 2013

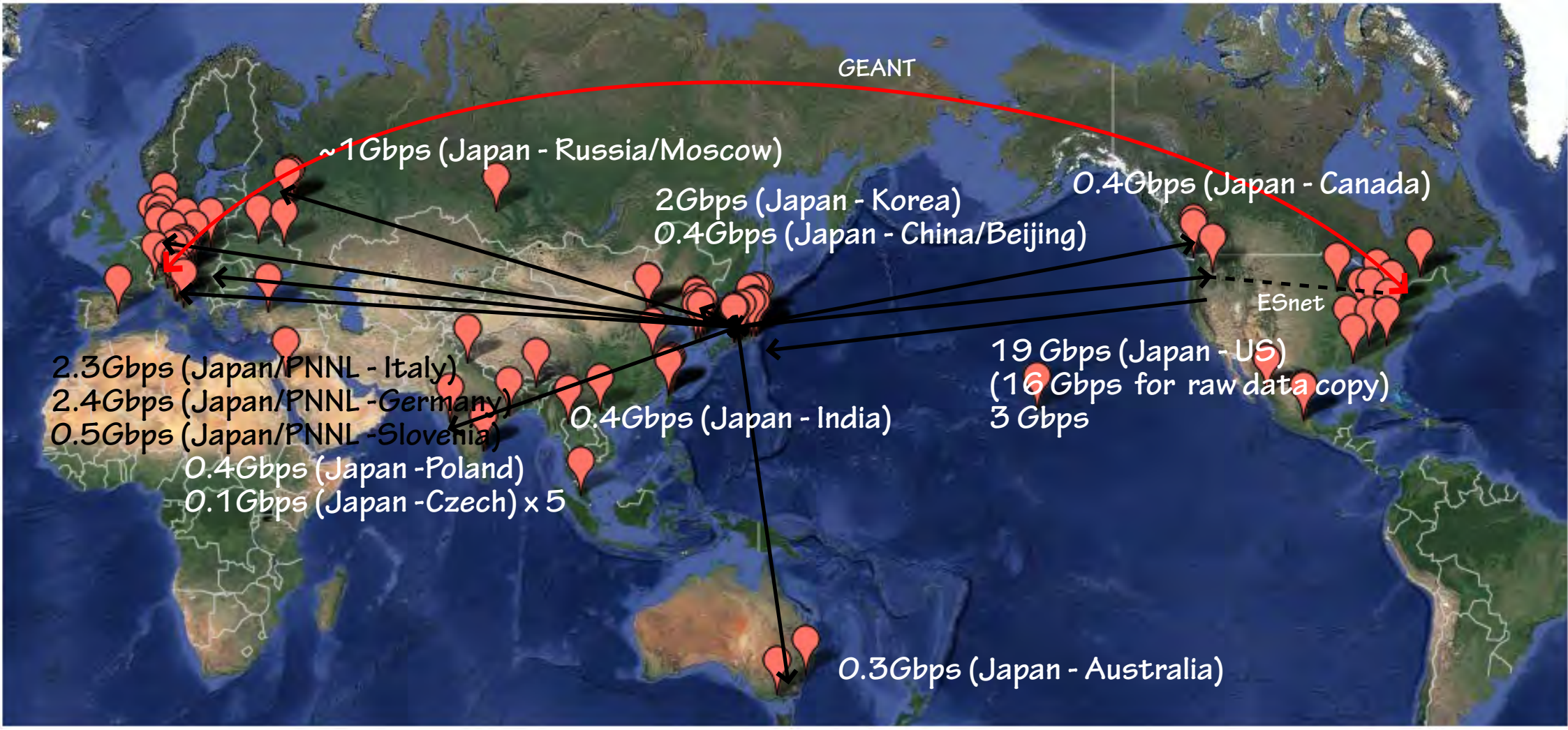
Network bandwidth @ ~2018



Japan/PNNL - XX : mdst transfer from Japan and/or PNNL + data transfer between XX and other sites
Japan - XX : data transfer between XX and Japan + other sites

as of Sep 26, 2013

Network bandwidth @ ~2022



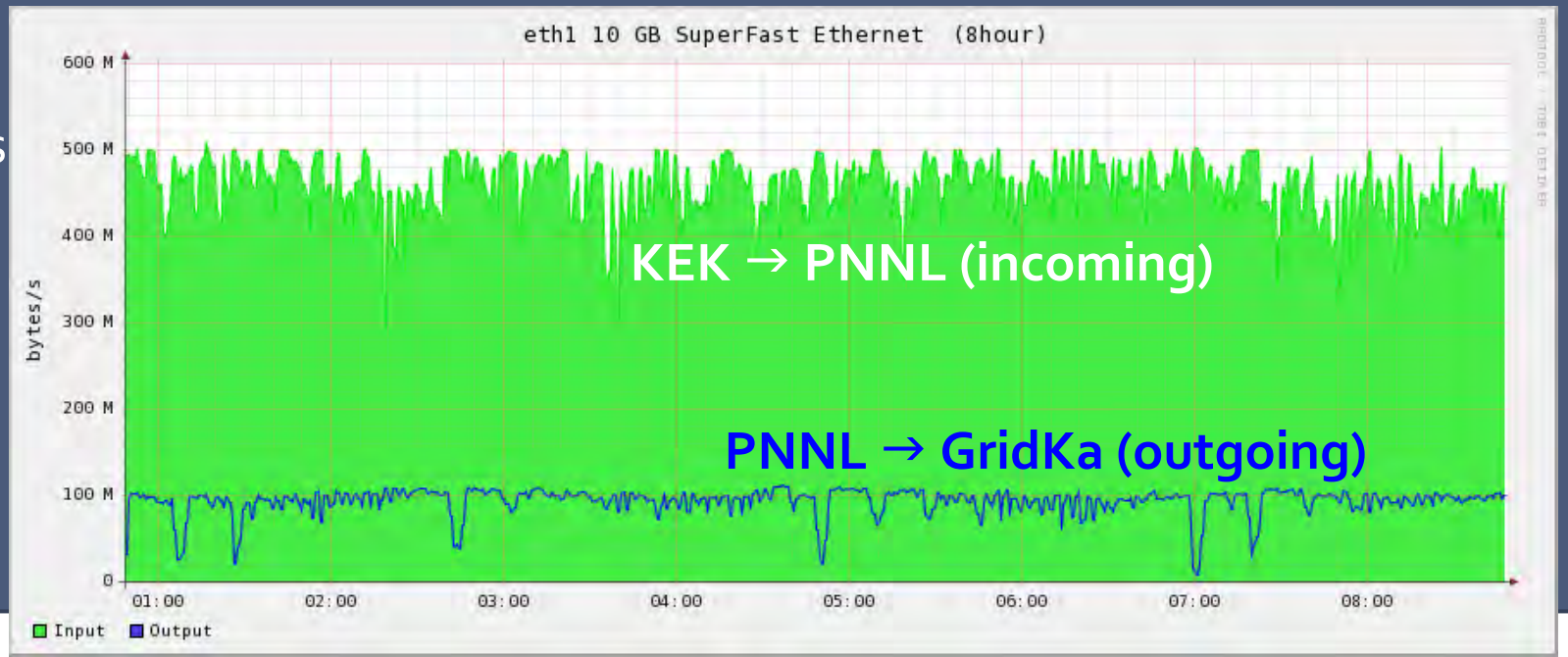
Japan/PNNL - XX : mdst transfer from Japan and/or PNNL + data transfer between XX and other sites
Japan - XX : data transfer between XX and Japan + other sites

as of Sep 26, 2013

Data challenge

- It is essential to establish and keep good network connection among GRID sites
  - Raw data transfer from KEK to PNNL
  - mDST data (data and MC) deployment to regional analysis center
  - Use of mDST data by each computing GRIDs, clouds
- Data transfer test using the MC campaign products
  - In May 2013, we performed massive data transfer using FTS2 at GridKa

500MB/s



Summary

- Belle II starts taking physics data from Oct. 2016
  (SuperKEKB accelerator commissioning starts in early 2015)
- Software framework has been established and working well
- Belle II adopts the distributed computing with DIRAC
- MC production campaign started with full detector simulation
- Data challenge also started

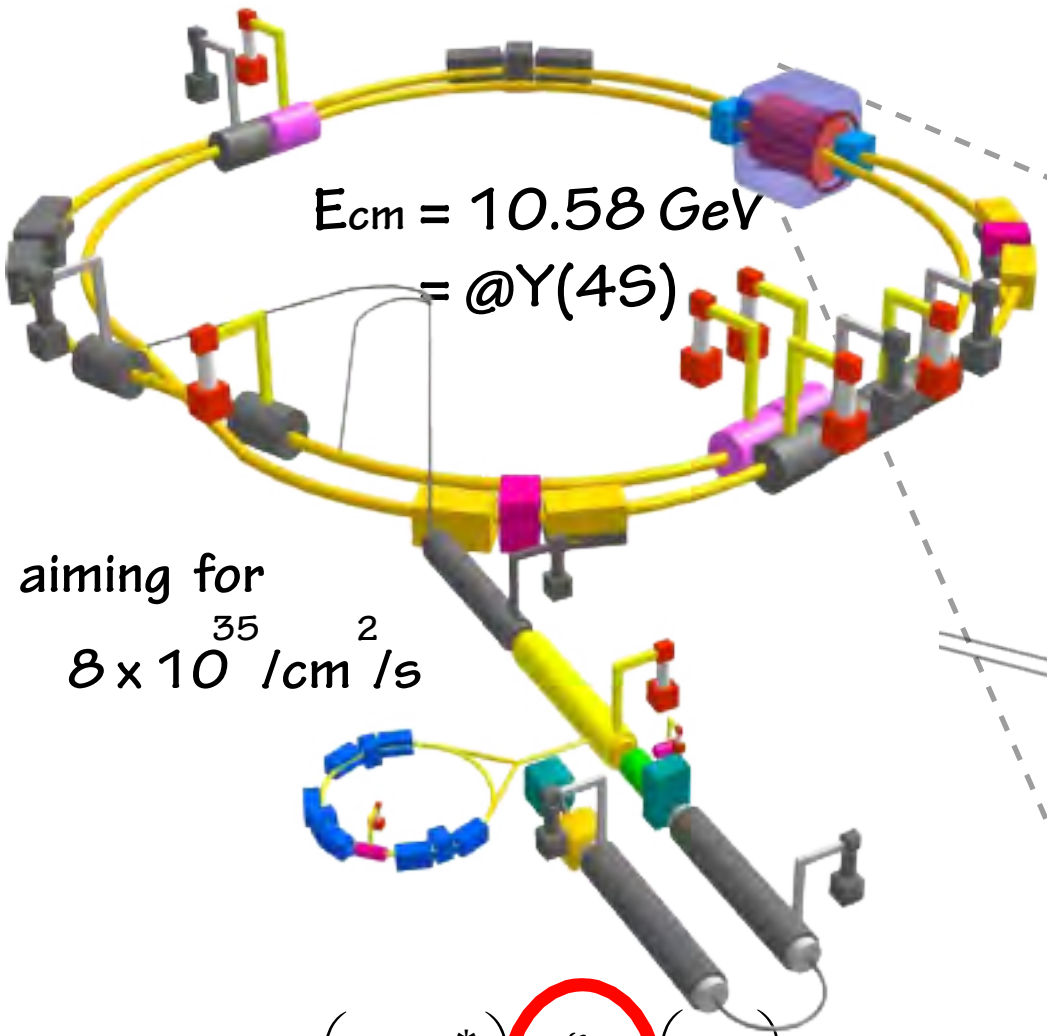
There are similarities between Belle II and ILC
+ software concepts
+ DIRAC
some chances to work together ?

Belle II top page http://belle2.kek.jp/

Now in the 16th collaobration meeting !!
http://kds.kek.jp/conferenceDisplay.py?ovw=True&confId=13911

I cordially appreciate the opportunity to report
the Belle II software and computing at LCWS !!!

SuperKEKB / Belle II



Ecm = 10.58 GeV
= @Y(4S)

aiming for
8×10^{35} /cm^2/s

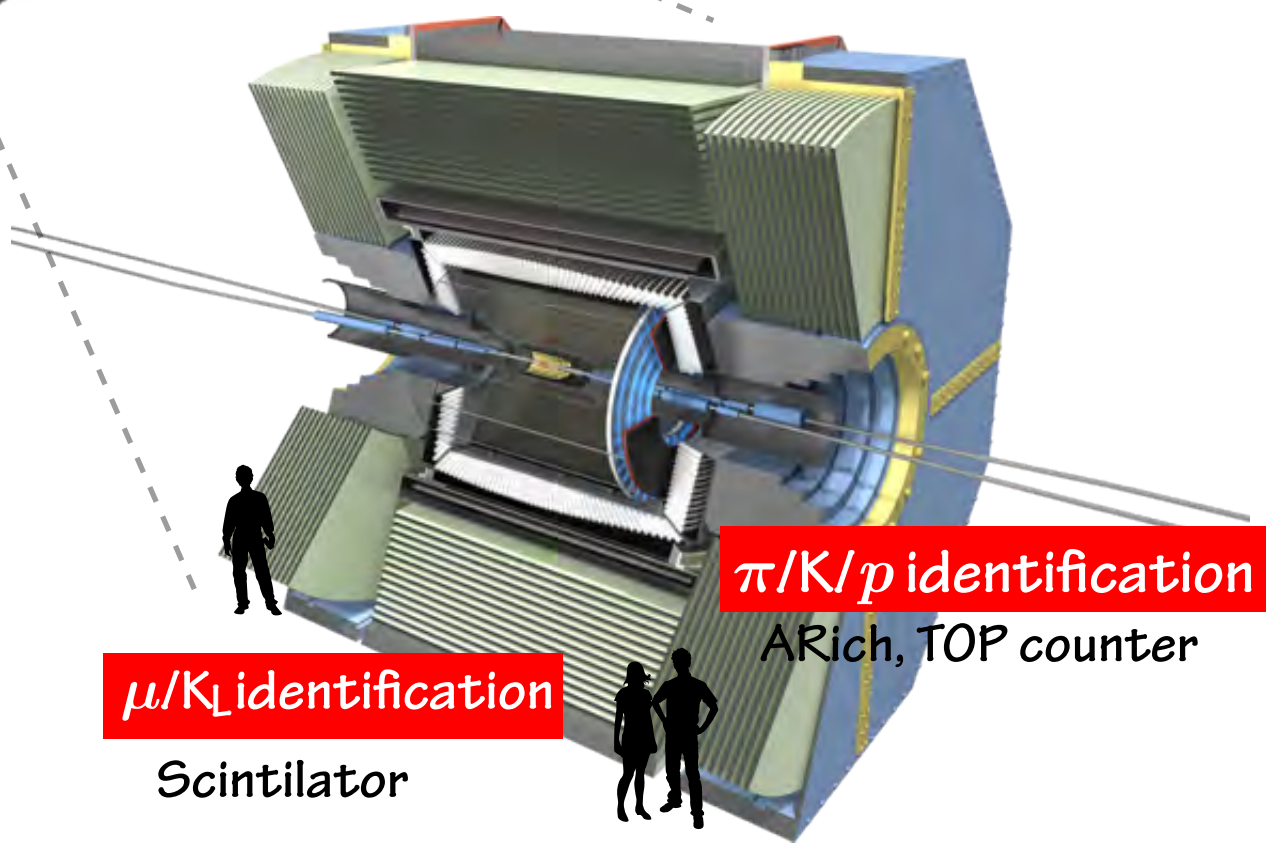
$$L = \frac{\gamma_\pm}{2er_e}\left(1+\frac{\sigma_y^*}{\sigma_x^*}\right)\frac{I_\pm \xi_{\pm y}}{\beta_y^*}\left(\frac{R_L}{R_y}\right)$$

tracking/vertexing

small-cell Drift Chamber
Silicon Strip det.
longer lever arm
Pixel detector

e/γ detection

pure-CsI calorimeter (end-cap)



π/K/p identification

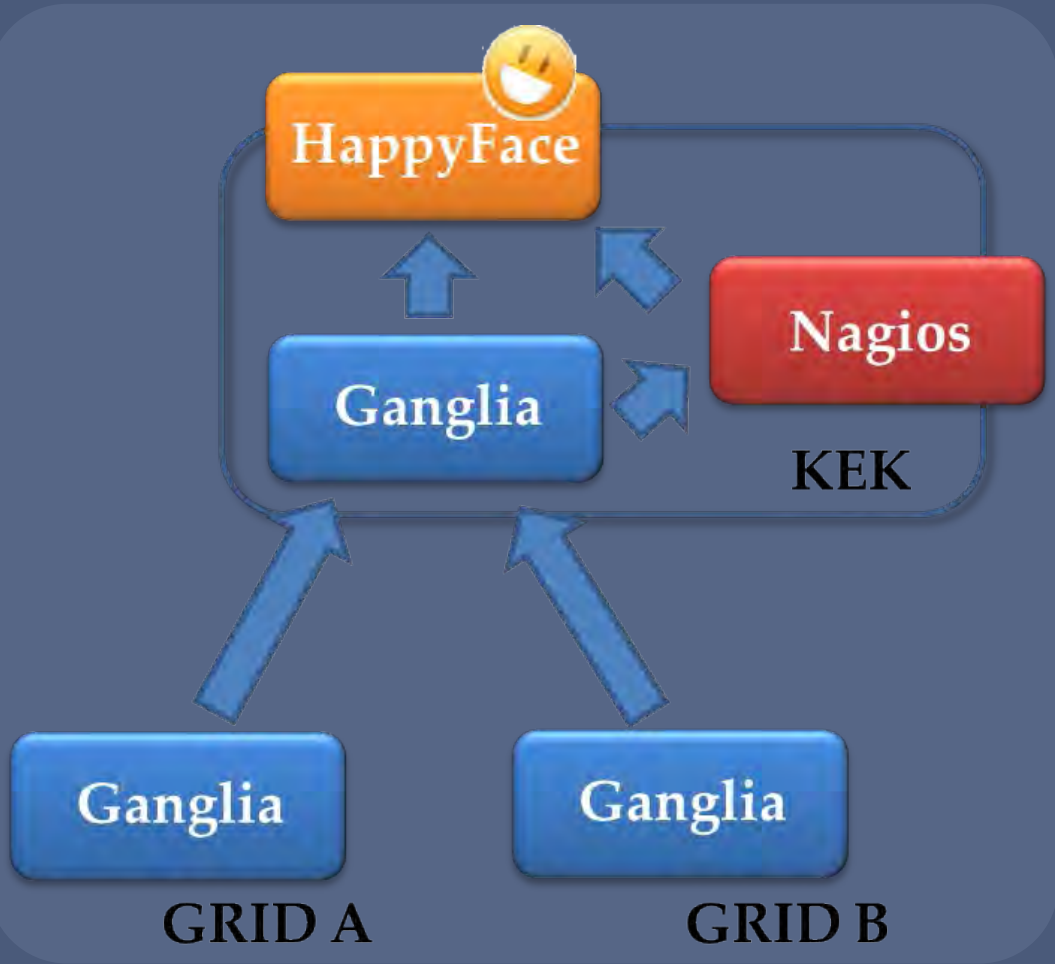
ARich, TOP counter

μ/K$_L$ identification

Scintilator

Monitoring

- Ganglia+Nagios installed in KEK and HappyFace in KIT
- HappyFace collects information from Ganglia/Nagios and directly DIRAC



HappyFace

Ganglia