

# DIRAC @ Belle II

↑  
KEKB:  $1 \text{ ab}^{-1}$



↑  
SuperKEKB:  $50 \text{ ab}^{-1}$



Thomas Kuhr  
for the Belle II  
Computing Group

DIRAC Users  
Meeting

12.05.2011

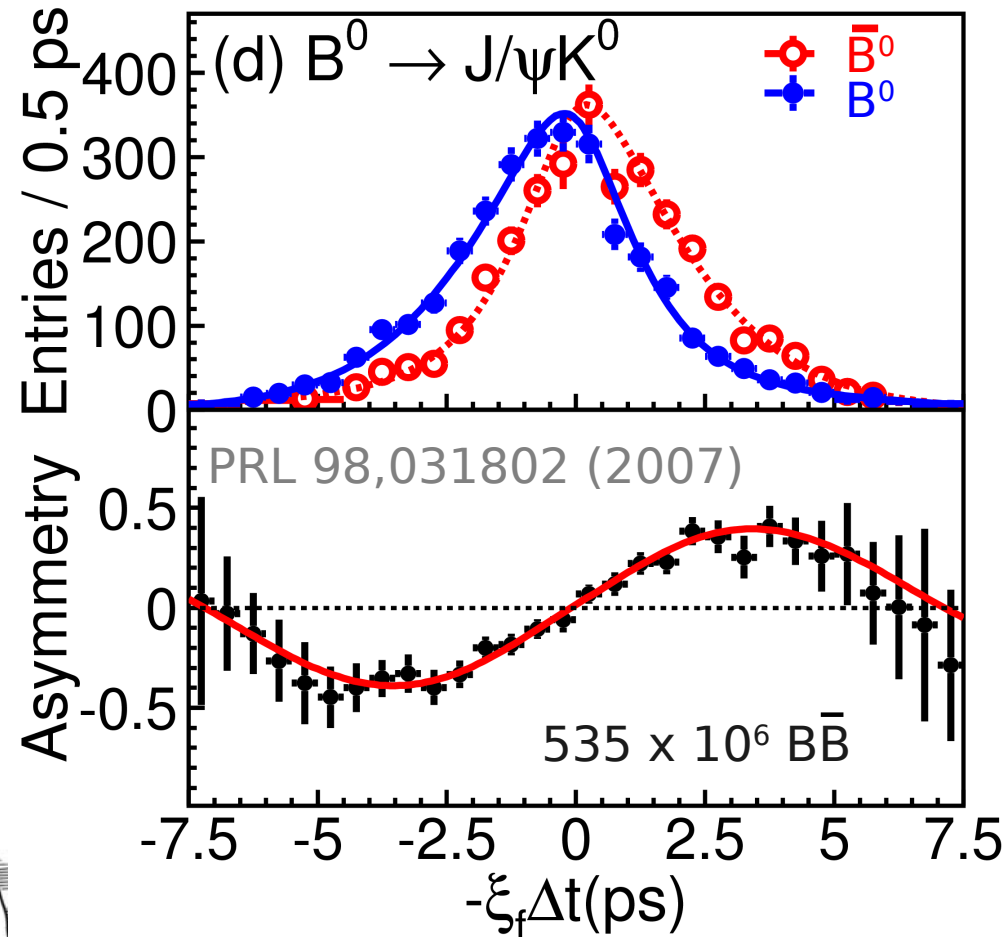


# Physics Objective of Belle and Belle II



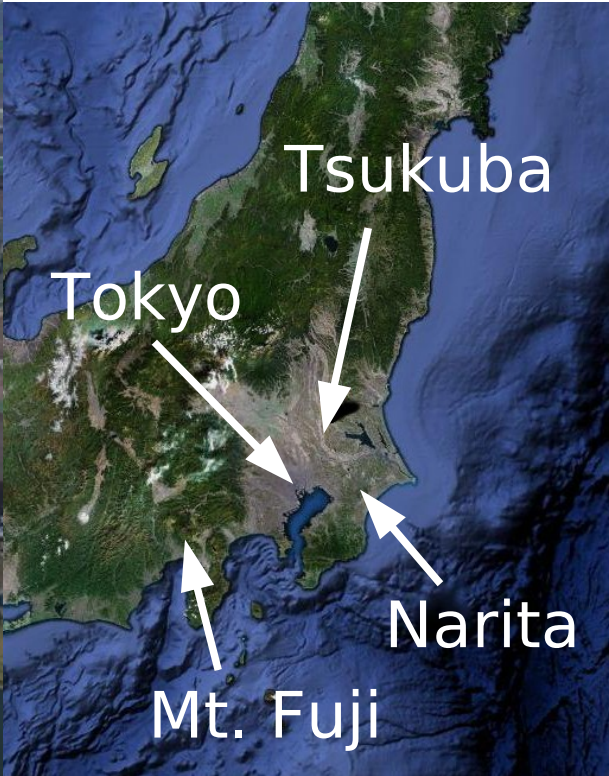
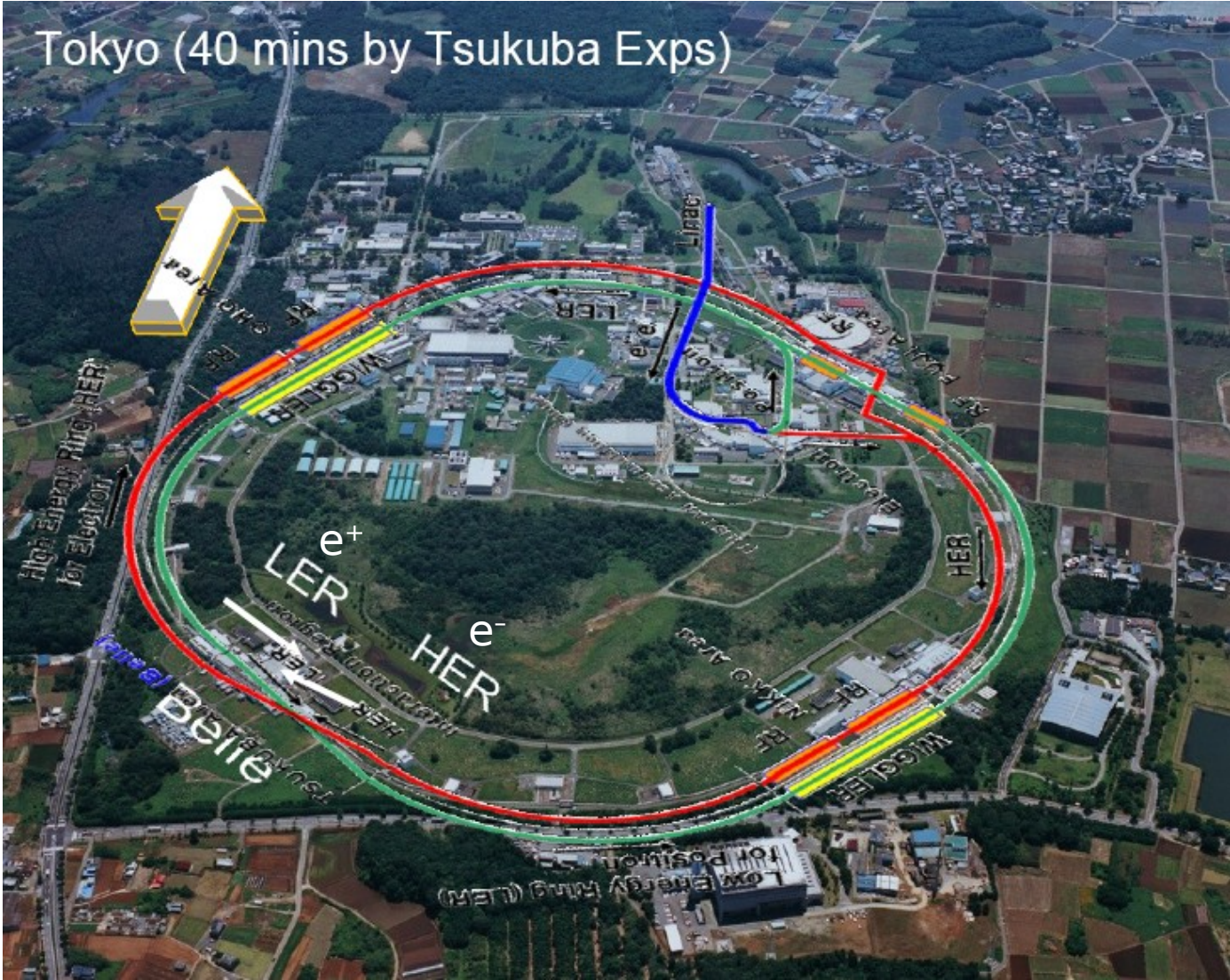
- ✓ Confirmation of KM mechanism of  $\mathcal{CP}$  in the Standard Model
- x  $\mathcal{CP}$  in the SM too small (by many orders of magnitude) to generate observed baryon asymmetry in the universe

→ Need sources of  $\mathcal{CP}$  beyond the SM



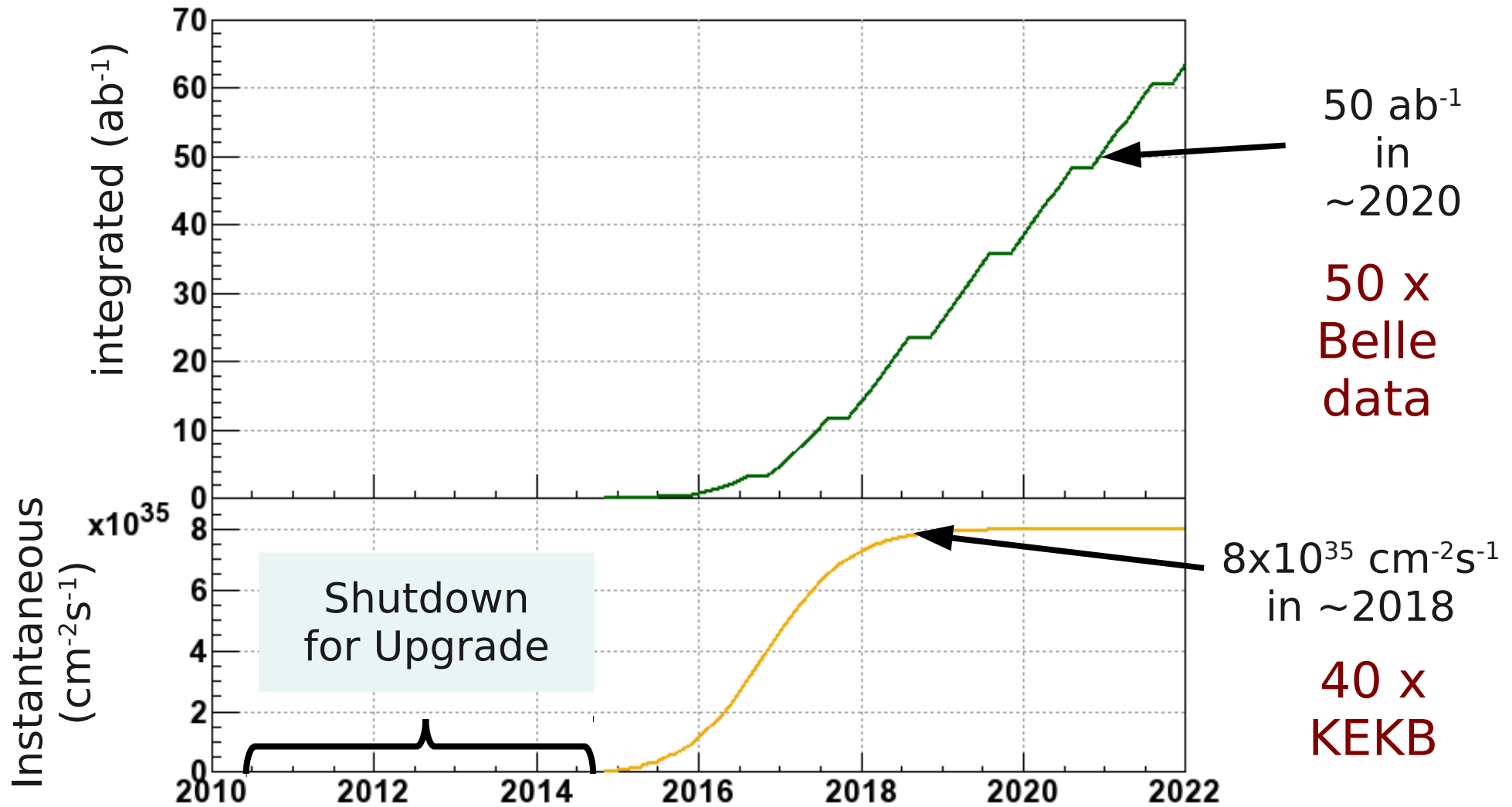
→ Super B factory

# KEK Site



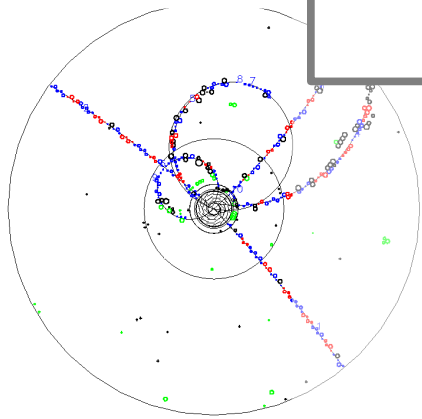
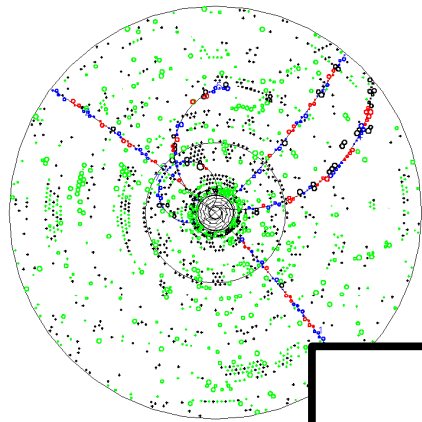


# Projection of Luminosity at SuperKEKB



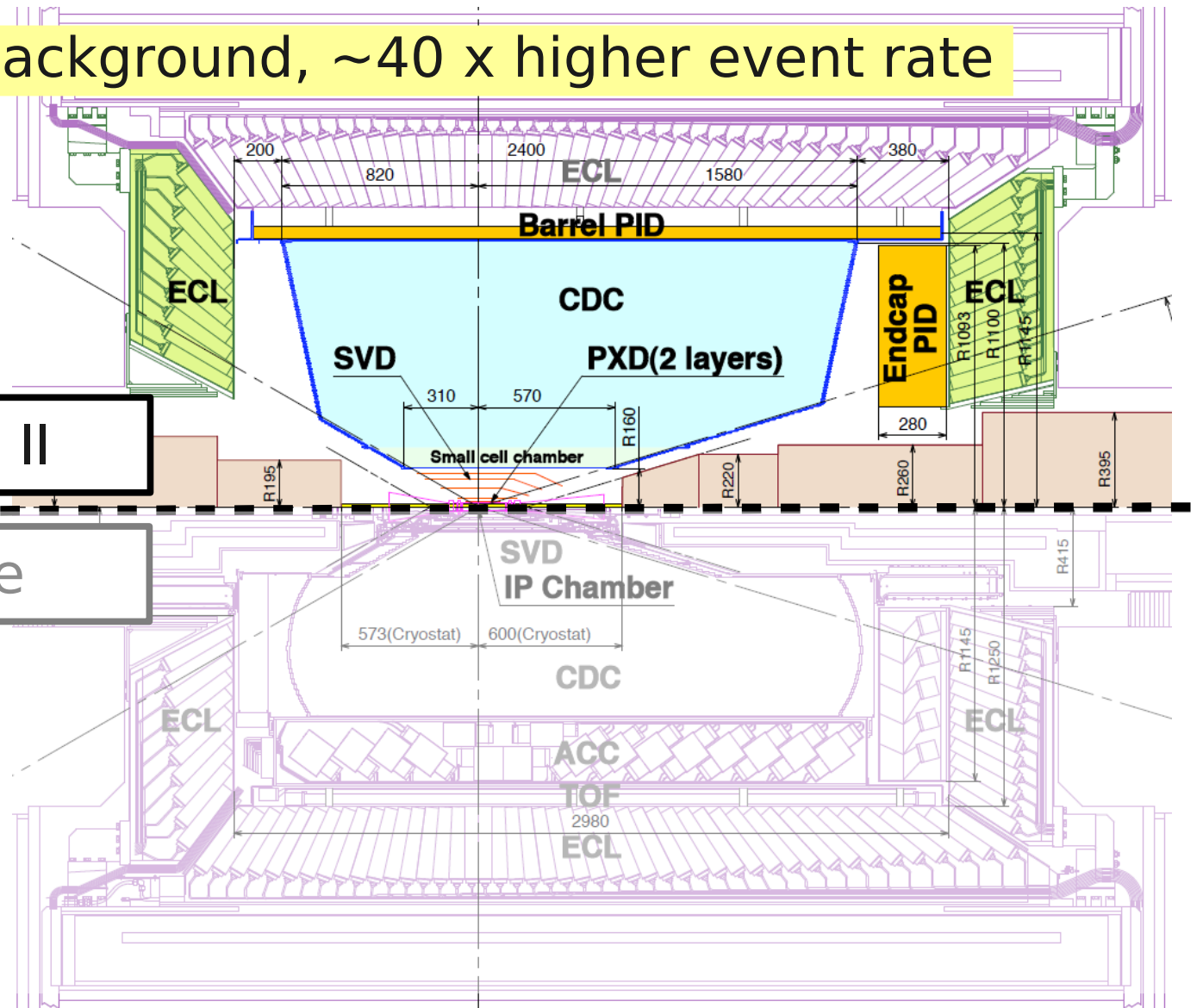
# Belle II Detector

→ 10-20 x higher background, ~40 x higher event rate



Belle II

Belle



# Estimated Data Rates

Experiment	Event Size [kB]	Rate [Hz]	Rate [MB/s]
<i>High rate scenario for Belle II DAQ:</i>			
Belle II	300	6,000	1,800
<i>LCG TDR (2005):</i>			
ALICE (HI)	12,500	100	1,250
ALICE (pp)	1,000	100	100
ATLAS	1,600	200	320
CMS	1,500	150	225
LHCb	25	2,000	50

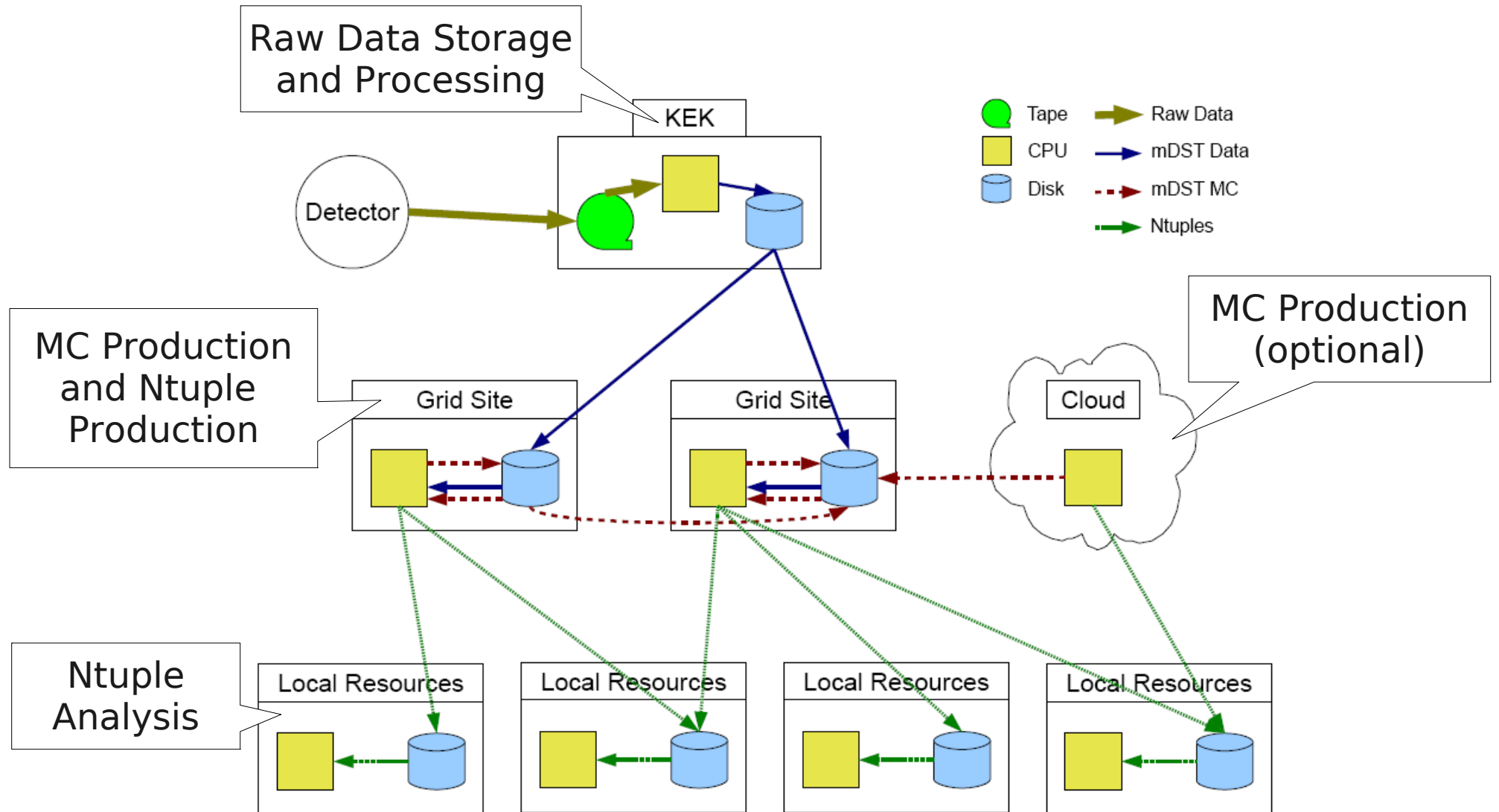
# Belle II Collaboration



~400 members  
58 institutions  
from 13 countries

→ Distributed  
collaboration

# Belle II Computing Model





# DIRAC @ Belle II

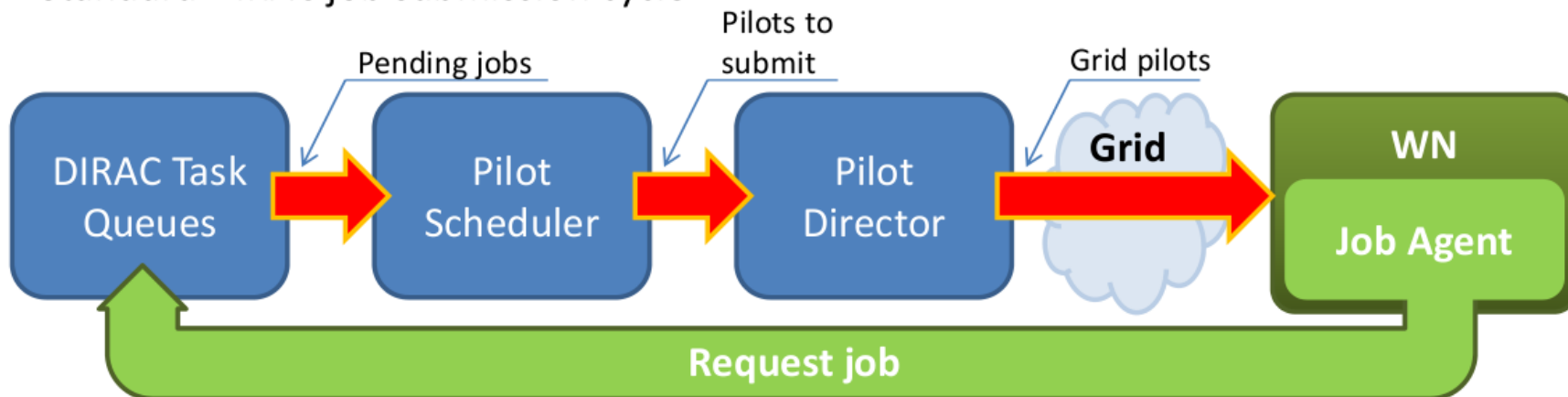
---

Belle II distributed computing system is based on DIRAC

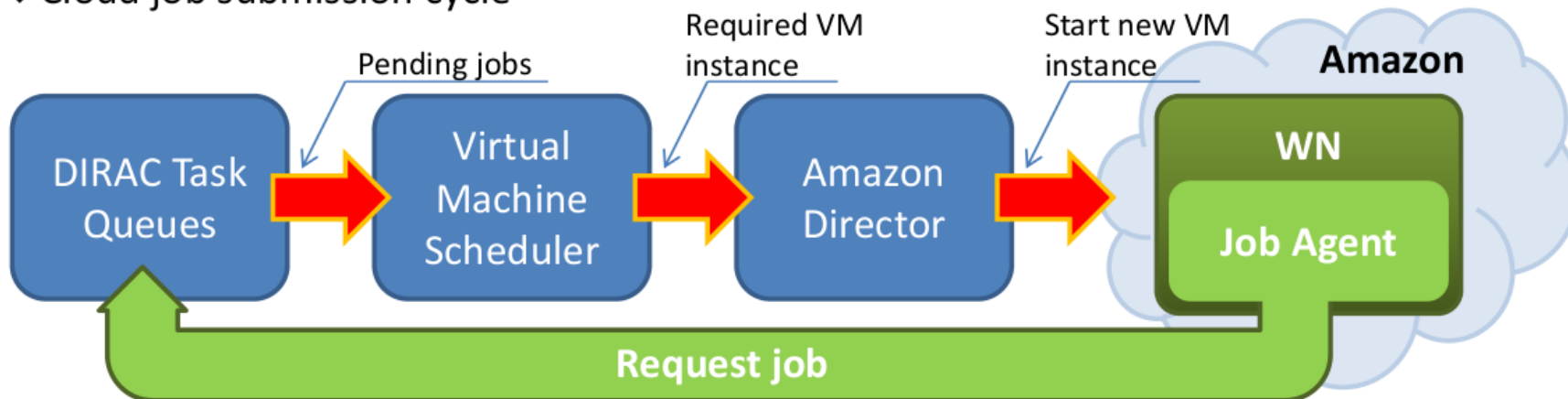
- It provides many features that we would have to develop on our own otherwise, like
  - Pilot jobs, web frontend, ...
- Modular design allows to easily extend and adjust the system for Belle II
  - Integration of cloud resources (Amazon EC2)
    - Already used for Belle MC production
  - Integration of OSG middleware
  - Management of job collections (project)

# Extension for Virtual Machines

## ❖ Standard DIRAC job submission cycle

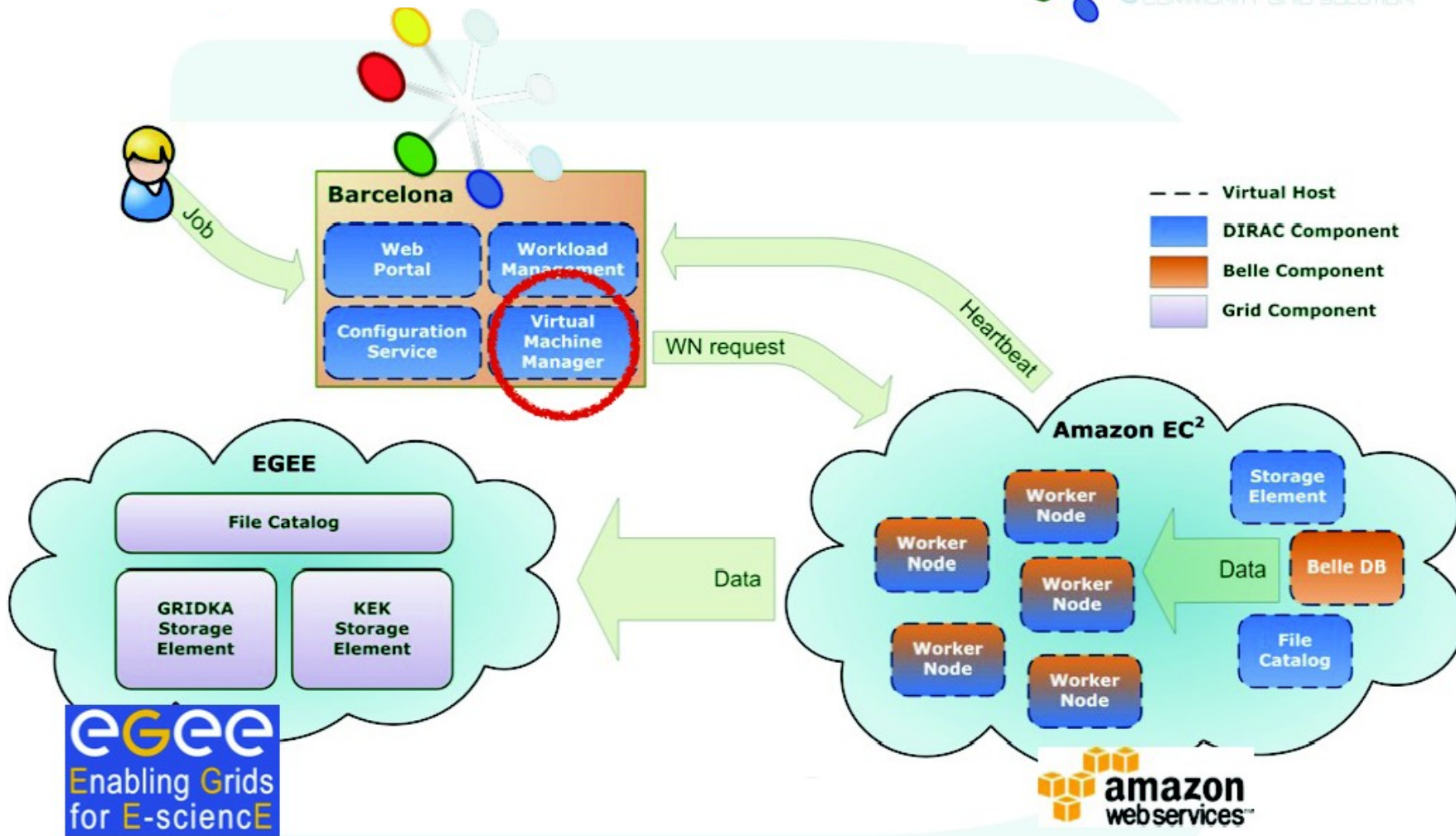


## ❖ Cloud job submission cycle



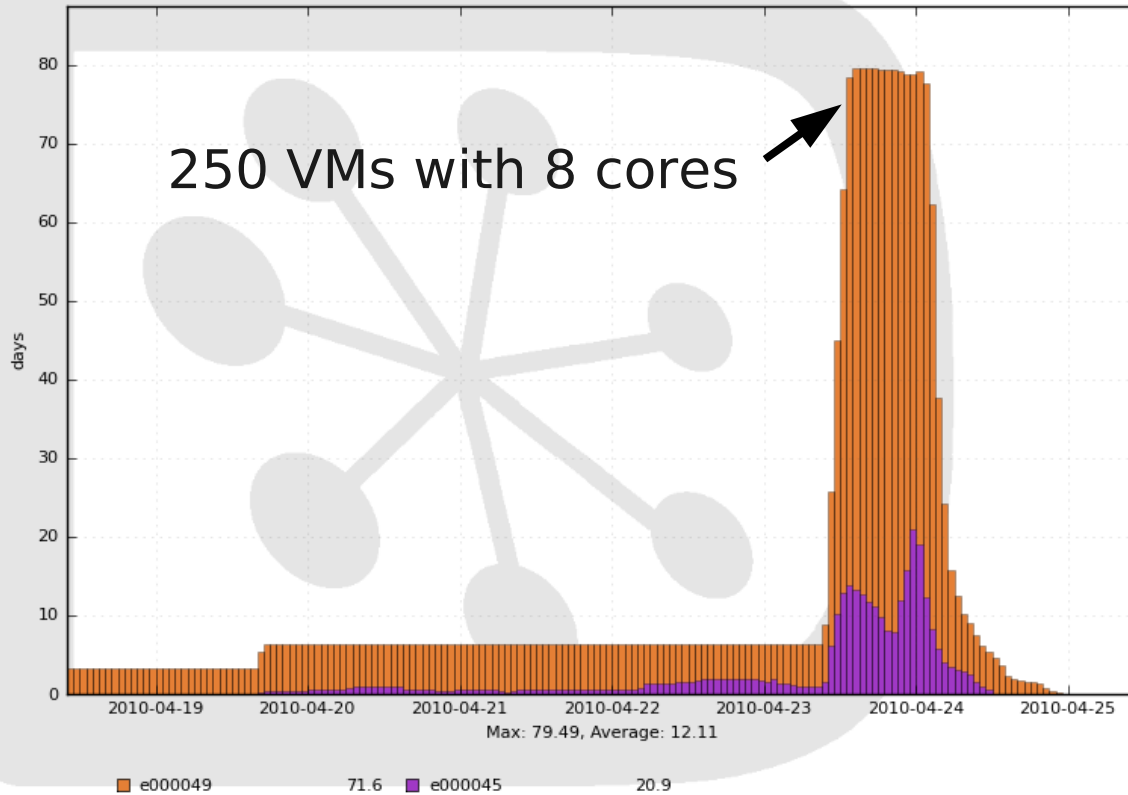


# DIRAC for VM management

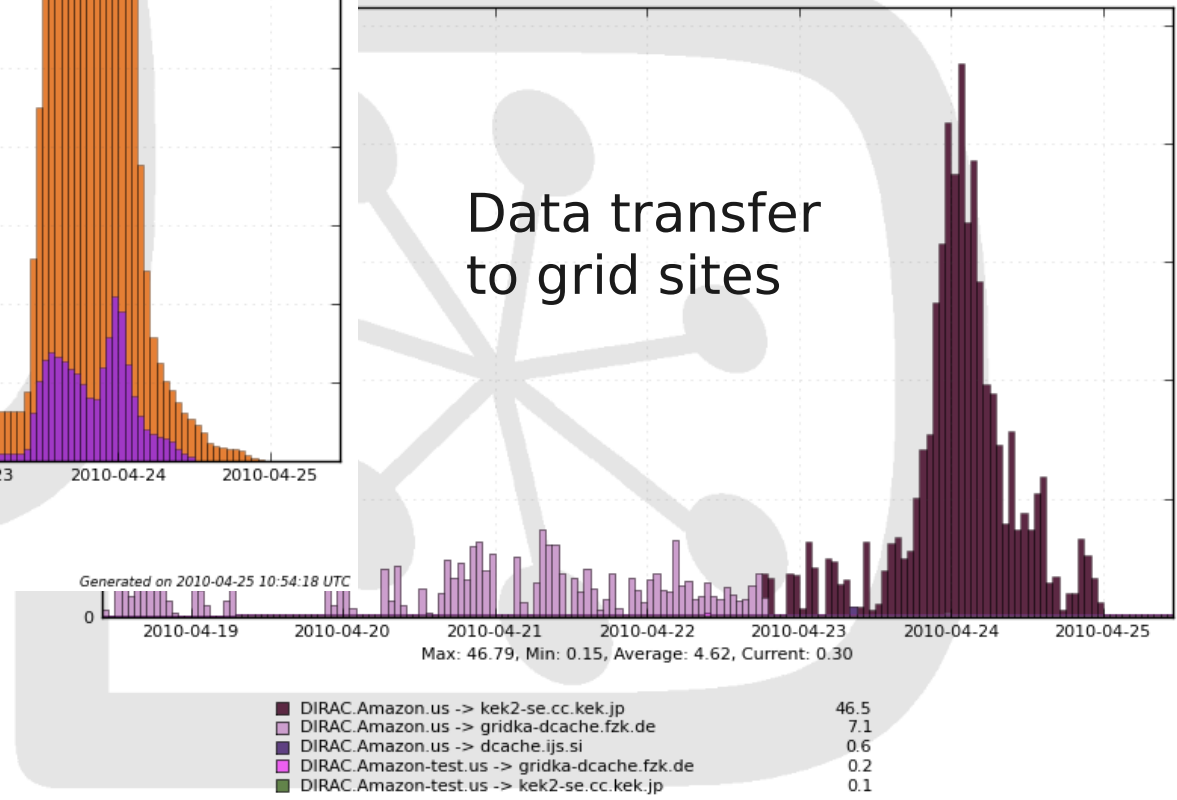


# DIRAC Test Phase I: Only Cloud

CPU usage by JobGroup  
7 Days from 2010-04-18 to 2010-04-25



Throughput by Channel  
7 Days from 2010-04-18 to 2010-04-25

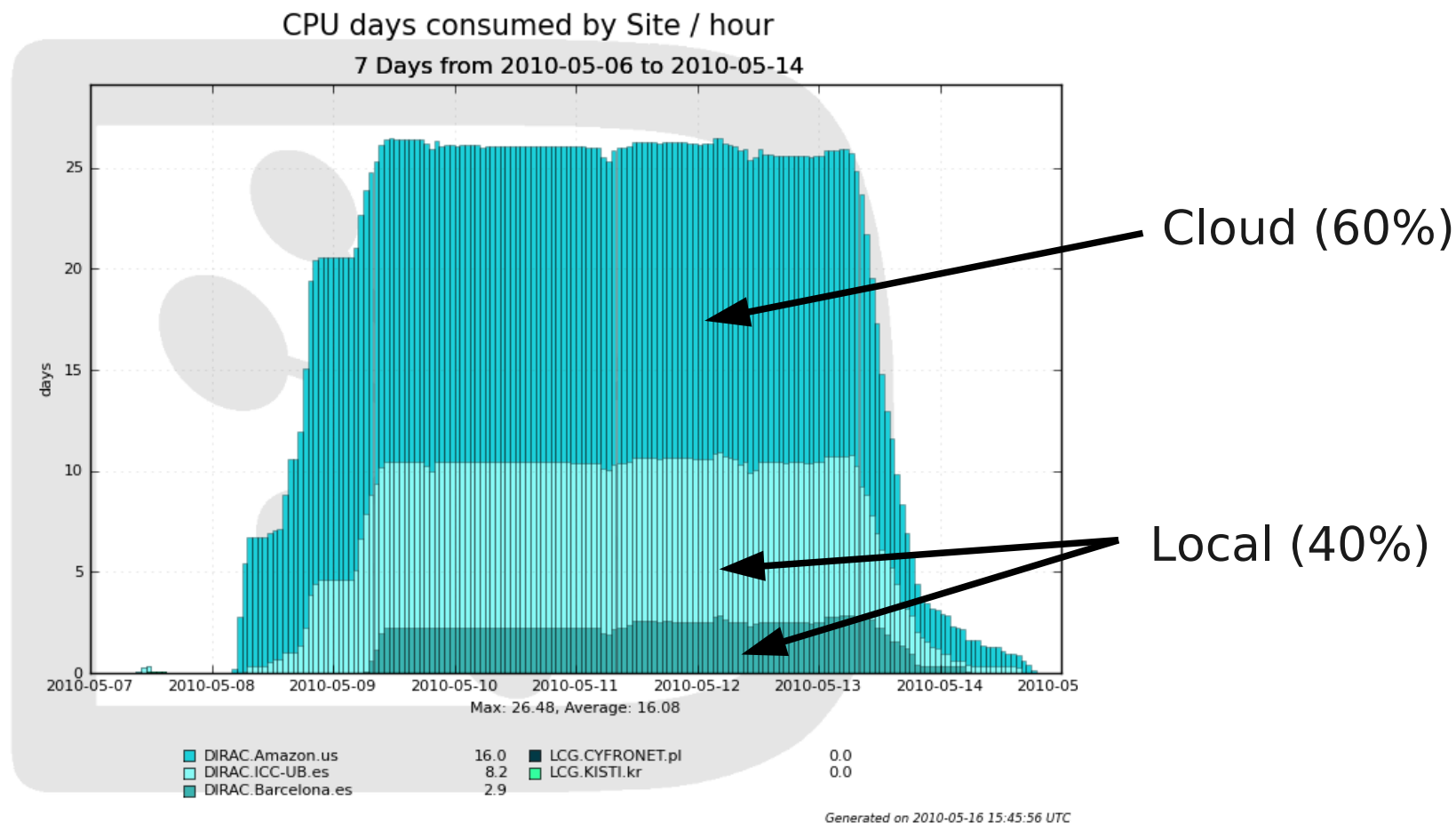


- 120M events (2.7 TB) produced in 10 days
- 0.46 USD / 10k events

Generated on 2010-04-25 10:49:36 UTC

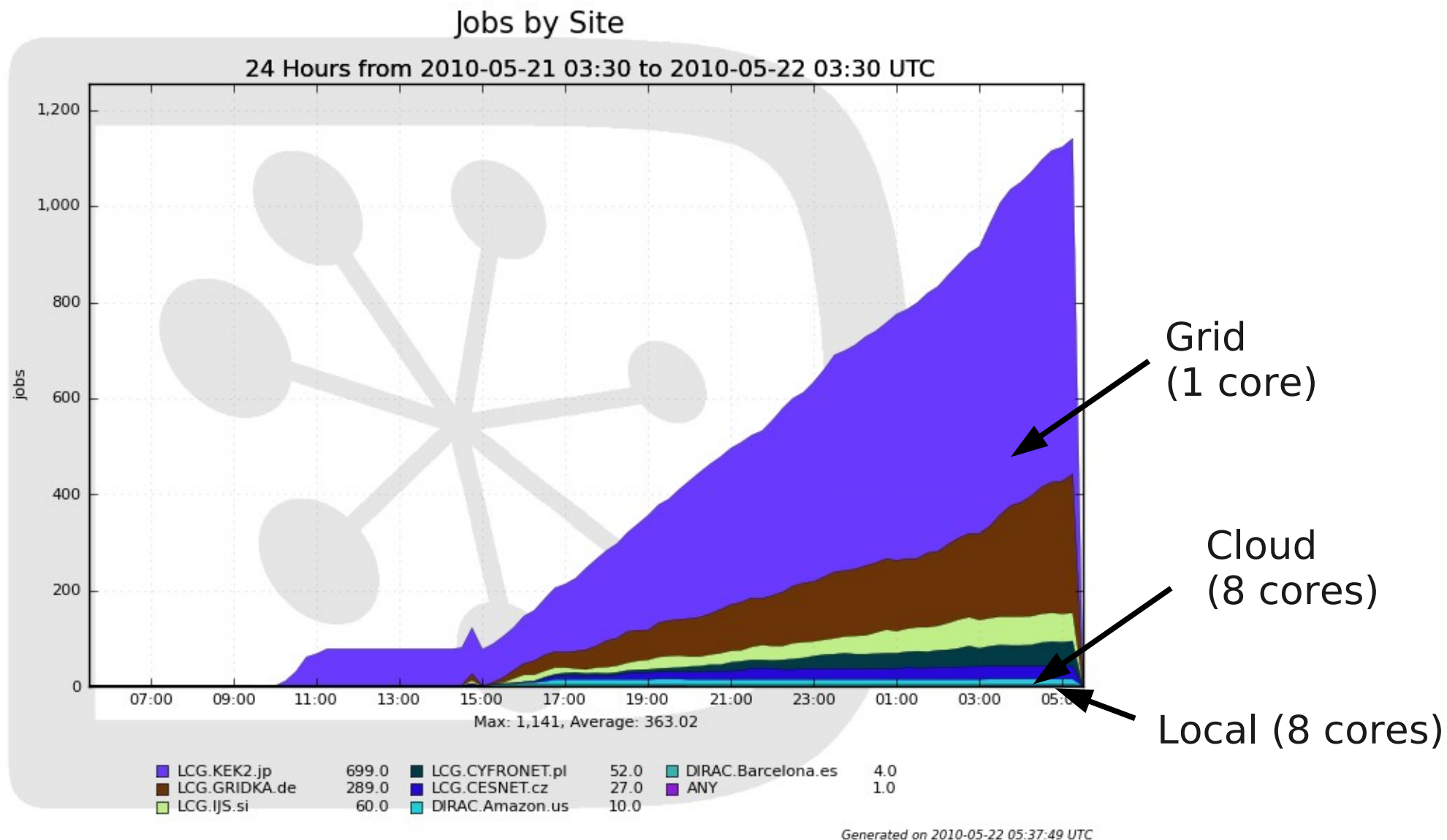


# DIRAC Test Phase II: Cloud + Local



- 170M events (3.6 TB) produced in 6 days
- Amazon Spot Instances → 0.20 USD / 10k events

# DIRAC Test Phase III: Cloud + Grid + Local





# User Interface: gbasf2

---

## Basf2 Steering File options

The default configuration option for gBasf2 is to set a number of variables in your normal basf2 steering file:

```
#####  
# gBASF2 configuration      #  
#####  
#Name for project  
project='e055-test'  
# (optional) Job priority [0-10]  
priority='1'  
#Experiments (comma separated list)  
experiments='13,57'  
#Metadata query  
query='id > 10 and id < 15'  
#Type of Data ('data' or 'MC')  
type='data'  
#estimated Average Events per Minute (eg Mcprod = 40)  
evtpermin='45'  
# (optional) Files to be sent with the job  
inputsandboxfiles = 'file1.txt,file2.txt'  
# (optional) max events - the maximum number of events to use  
maxevents = '100000'
```

You can then invoke gBasf2 using the steering file and it will do the rest:

```
./gbasf2.py -s steering_file.py
```

Same python steering file as for offline basf2 job, but with additional parameters for the grid job

# Metadata Catalog (AMGA)

Applications Places System 13°C Mon Dec 27, 12:33 AM fifieldt

Dataset Browser as belle@Dirac-Test - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://kek2-uidev.cc.kek.jp:15043/DIRAC/Dirac-Test/belle/systems/datasets/browse

Most Visited EPP Nagios PWLCG Tier2 Tools Investigate EC2 Belle Dev iCMS DIRAC@JP ROC DIRAC

Dataset Browser as belle@Dirac-Test

Systems Jobs Virtual machines Projects Datasets Help Selected setup: Dirac-Test

- belle
  - MC
  - config
  - dataset
    - dataset001
    - dataset002
    - dataset003
    - dataset004
    - dataset005
    - dataset006
  - site
  - skim
  - software
  - user
- belle2
  - MC
  - data
    - E17
    - E33
    - E59
    - E61
    - E23
    - E25
    - E27
    - E29
    - E31
    - E35
    - E37

➤ Input data selected by query on file level metadata attributes

➤ “Datasets” are named queries

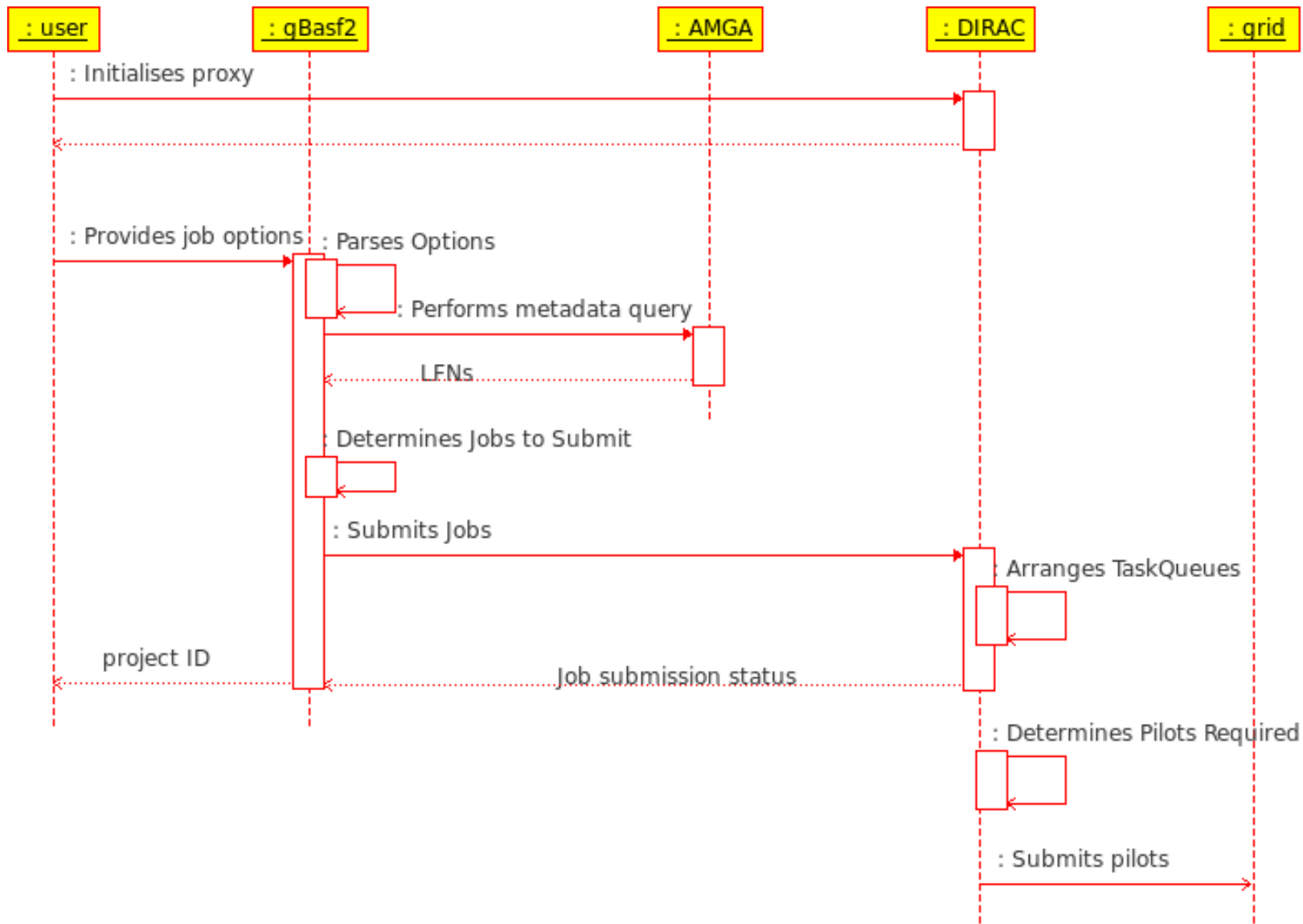
dirac@ belle (/C=AU/O=APACGrid/OU=The University of Melbourne/CN=Tom Ffield)

https://kek2-uidev.cc.kek.jp:15043/DIRAC/Dirac-Test/belle/systems/datasets/browse

[fifieldt@nomad3010:~] Dataset Browser as b... Simon Bian [fifieldt]

Tor Disabled

# gbasf2 Control Flow





# Analysis Projects

The screenshot shows a web browser window titled "Project Overview as belle@Dirac-Test - Mozilla Firefox". The address bar contains the URL "https://kek2-uidev.cc.kek.jp:15043/DIRAC/Dirac-Test/belle/systems/projects/overview". The browser's toolbar shows various icons and a search bar with "javascript replace". The page content includes a navigation menu with "Systems", "Jobs", "Virtual machines", "Projects", "Datasets", and "Help". Below the menu, there are action buttons: "Select all", "Select none", "Reschedule Project", and "Terminate Project". The main content is a table with the following columns: "Project", "Progress", "Status", "LastUpdate", "Submission Time", and "Owner".

Project	Progress	Status	LastUpdate	Submission Time	Owner
<input type="checkbox"/> test3	100%	Done - with failures	seven days ago	2010-12-28 17:17:5	tkuhr
<input type="checkbox"/> test1	100%	Done - with failures	seven days ago	2010-12-28 14:21:5	tkuhr
<input type="checkbox"/> loadstorm-kek2	60%	Running	two weeks ago	2010-12-22 06:38:4	dirac
<input type="checkbox"/> loadstorm-all	100%	Done	two weeks ago	2010-12-22 05:35:5	dirac
<input type="checkbox"/> installation	100%	Running	seven days ago	2010-12-26 00:58:1	dirac
<input type="checkbox"/> e055-test2	100%	Done - with failures	two weeks ago	2010-12-21 08:39:1	dirac
<input type="checkbox"/> e055-test	100%	Done - with failures	22 minutes ago	2011-01-04 06:47:4	dirac
<input type="checkbox"/> NoGroup	100%	Done	two weeks ago	2010-12-22 05:50:3	dirac

At the bottom of the browser window, the console shows several network requests:

- POST https://kek2-uidev.cc.kek.jp:15043/DIRAC/Dirac-Test/belle/systems/projects/getProjectsList Aborted
- POST https://kek2-uidev.cc.kek.jp:15043/DIRAC/Dirac-Test/belle/systems/projects/getProjectsList 200 OK 4.01s
- POST https://kek2-uidev.cc.kek.jp:15043/DIRAC/Dirac-Test/belle/web/notifications/getUserStats 200 OK 898ms

- Analysis projects provide high level user interface
- ➔ Bookkeeping of jobs

# People

---

- Tom Fifield (Melbourne)
- Yanliang Han, Wenjing Wu (Beijing)
- Miłosz Zdybał, Rafal Grzymkowski (Krakow)
- Sunil Ahn, Jung-Hyun Kim (KISTI)
- Martin Sevier (Melbourne)

# Experience with DIRAC

---

- “I find it's very convenient to manage the grid proxy and then submit jobs to the grid with Dirac.”
- “I think the biggest problem with Dirac is that there are no manuals to introduce the Dirac functions and classes, so it's difficult to do development (We have to understand the code).”
- Dirac team provided great support for Belle MC production campaign

## Wish list:

- Automatized client installation on any linux system
- Output merging
- Data transfer tools



# Summary

---

- Belle II is developing a distributed computing system
  - ➔ It is based on DIRAC
- ➔ DIRAC was already successfully used for Belle MC production on grid, cloud, and local clusters
- Extension for analysis projects:  
Bookkeeping of jobs in job collections
- Interface to metadata catalog
- ➔ Looking forward to continue good collaboration between



and

