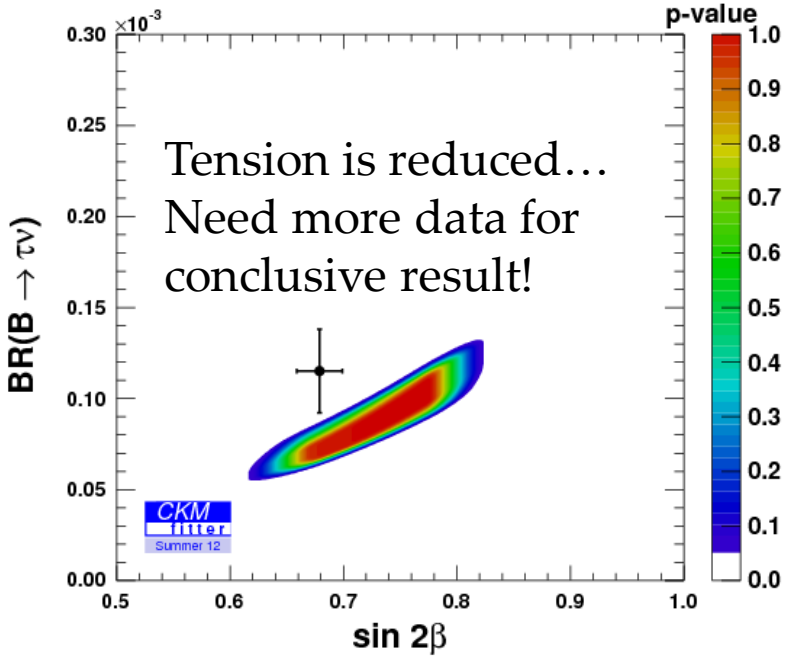
Status of the Belle II GRID environment and performance

Hideki Miyake (KEK)
for Belle II collaboration

DIRAC user workshop
Oct. 31st, 2012, Marseille, France

Goal of Belle II experiment

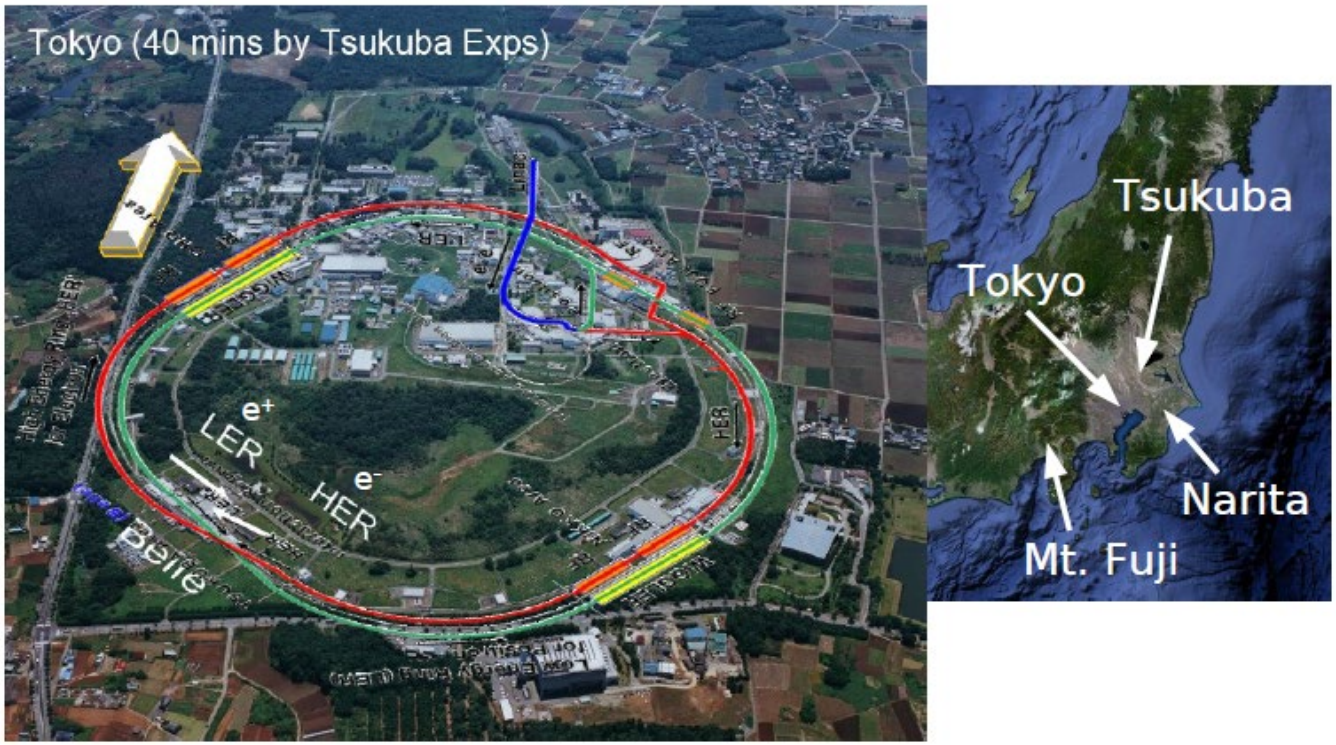
- ✔ Confirmation of KM mechanism of ~~CP~~ in the Standard Model
- ✗ ~~CP~~ in the SM too small (by many orders of magnitude) to generate observed baryon asymmetry in the universe
- → Need sources of ~~CP~~ beyond the SM



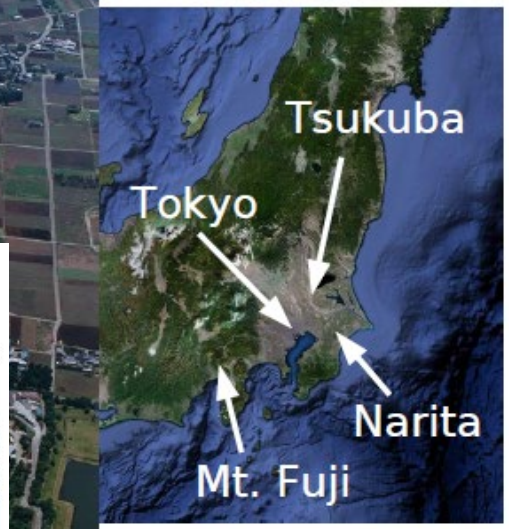
→ Super B factory

Complementary to LHCb

Belle II experiment



Belle II experiment

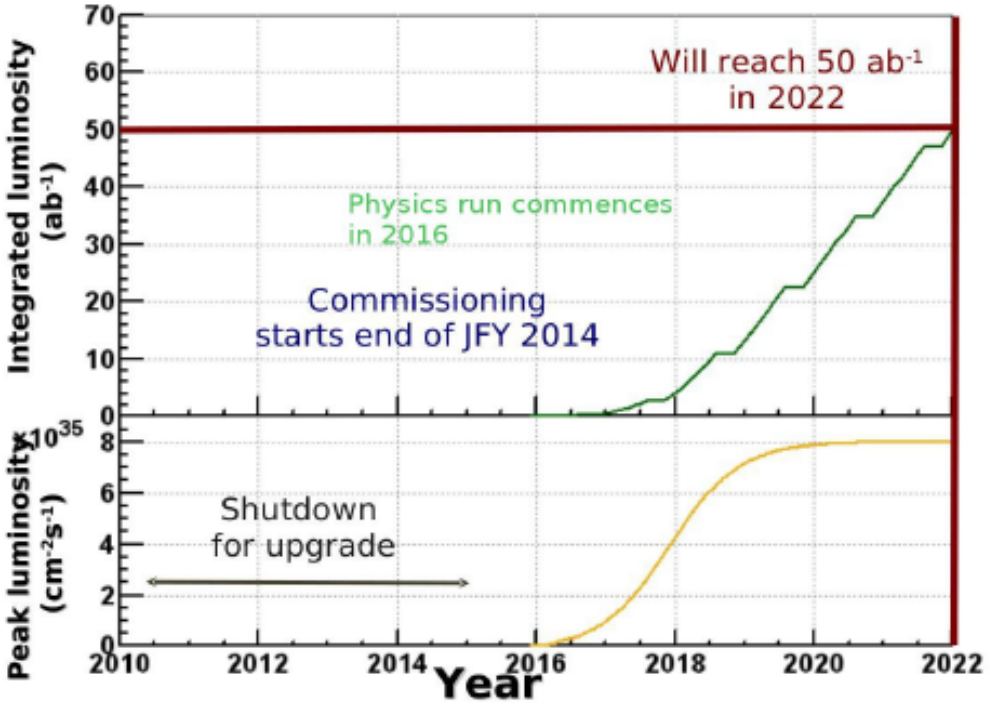


Belle II Collaboration

→ Distributed collaboration

Challenging performance

x50 larger integrated luminosity than Belle



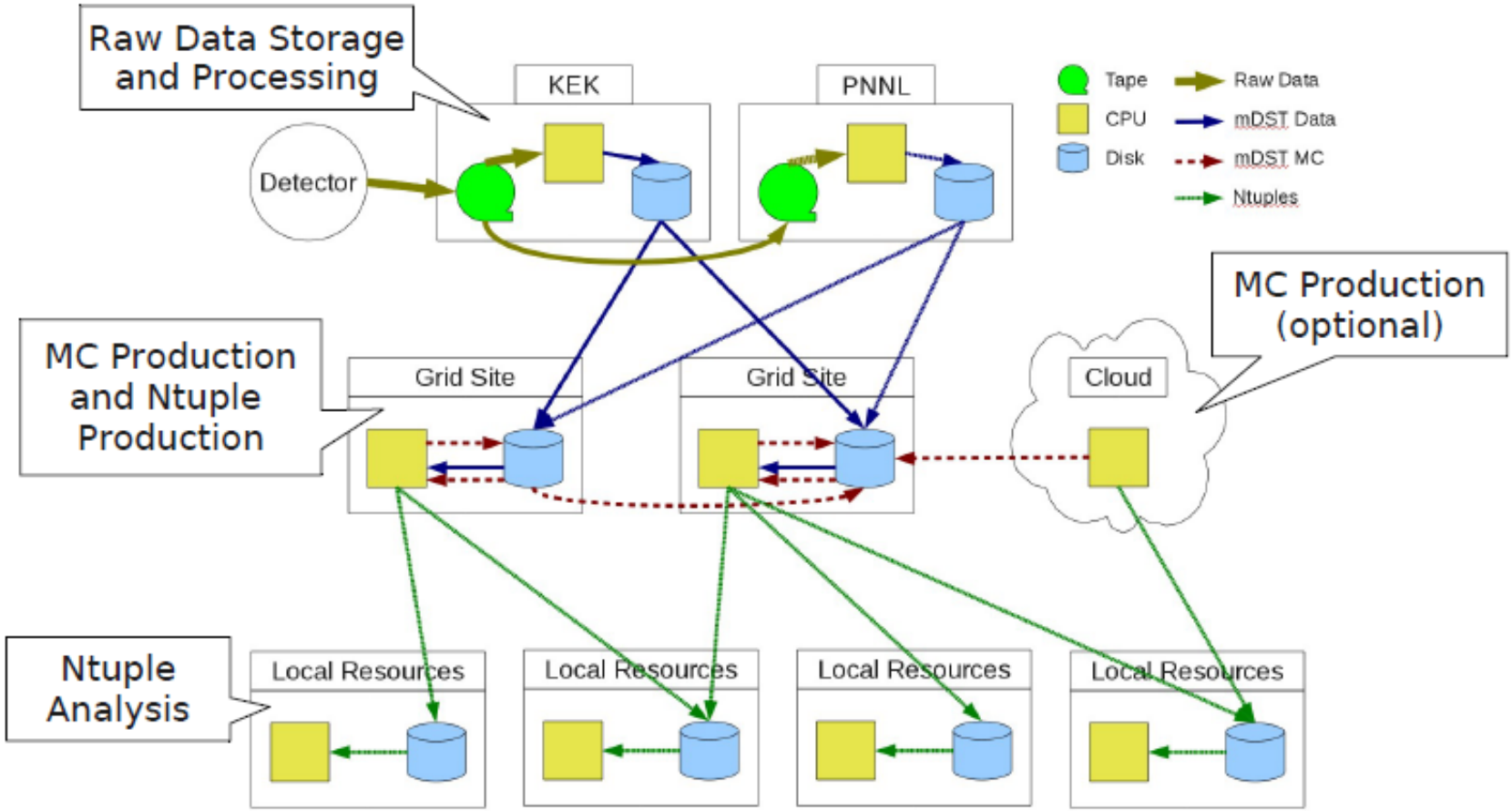
Comparable data rate to all LHC experiments

Experiment	Event Size [kB]	Rate [Hz]	Rate [MB/s]
High rate scenario for Belle II DAQ:			
Belle II	300	6,000	1,800
LCG TDR (2005):			
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 must handle large amount of data

- Tape ~ 200PB
- Disk ~ 200PB (incl. MC)

Belle II computing model



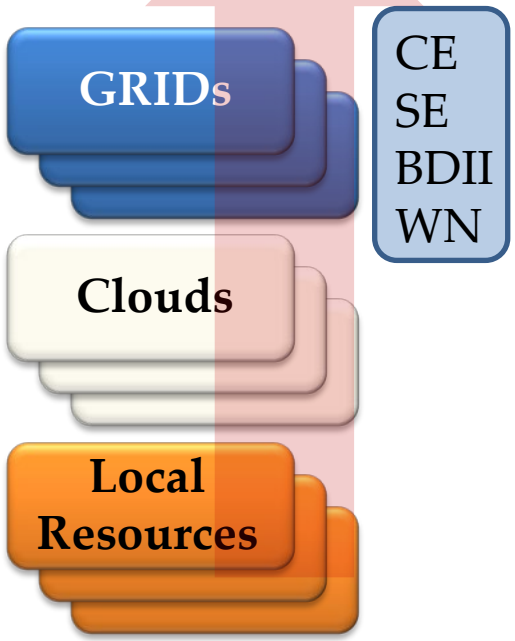
Belle II distributed computing

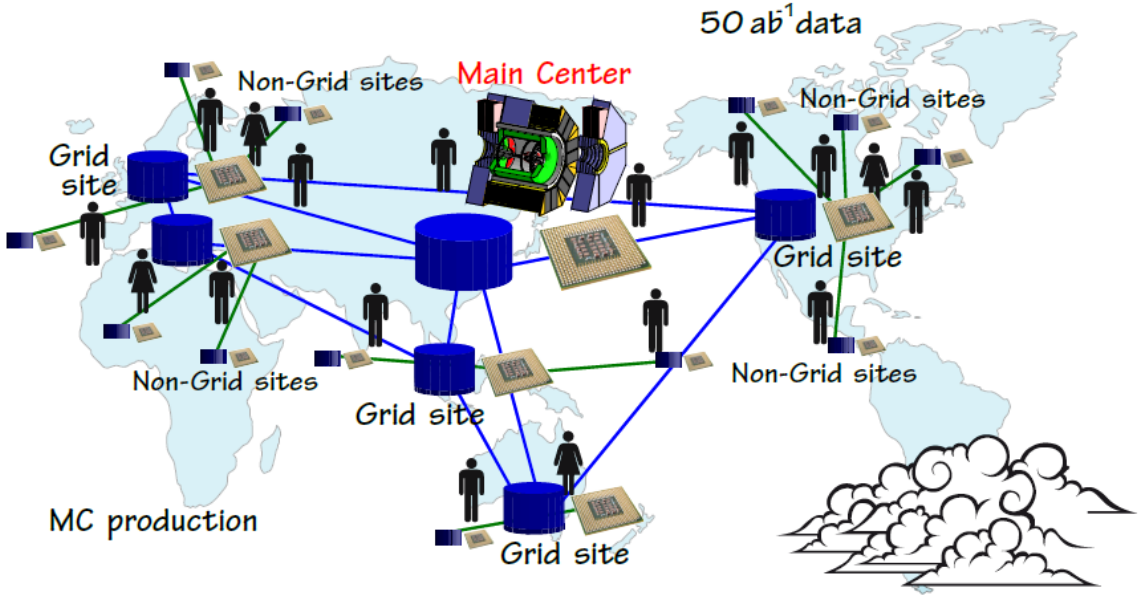
DIRAC
THE INTERWARE

amga ARDA METADATA CATALOGUE PROJECT
overview downloads installation browse source more docs

- Belle II employs DIRAC
- Metadata is managed by AMGA

- Both DIRAC and AMGA main servers are located in KEK

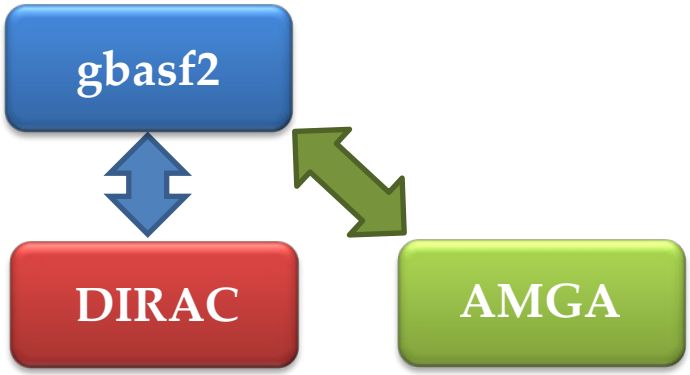




gbasf2

- Gateway to GRID world for Belle II user
  - Wrapper scripts of Belle II analysis framework (basf2)
    - User can run same steering file on both local and GRID
  - Written in Python
  - Provides communication with DIRAC and AMGA via API

% gb2_job_submit –s steeringfile.py



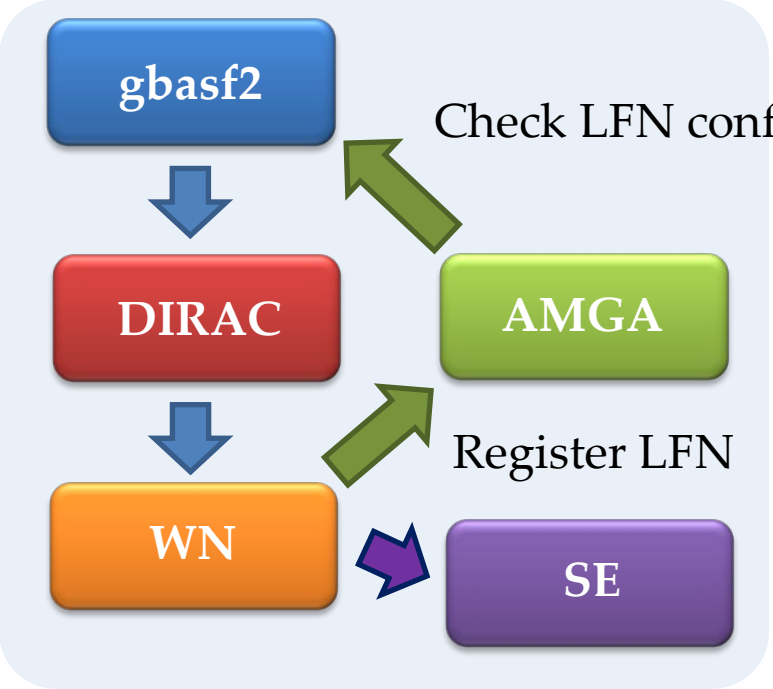
GRID

% basf2 steeringfile.py

local

```
##########################
# Basf2 configuration    #
##########################
# Register modules
evtmetagen = fw.register_module('EvtMetaGen')
evtmetainfo = fw.register_module('EvtMetaInfo')
paramloader = fw.register_module('ParamLoaderXML')
geobuilder = fw.register_module('GeoBuilder')
g4sim = fw.register_module('SimModule')
cdcdigitizer = fw.register_module('CDCDigitizer')
out = fw.register_module('SimpleOutput')

##Set parameters
param_evtmetagen = {
    'ExpList': [1],
    'RunList': [2],
    'EvtStartList': [7001],
    'EvtEndList': [9000],
    }

evtmetagen.param(param_evtmetagen)

param_g4sim = {'RandomSeed': 5486, 'MacroName': 'g4sim.macro'}
g4sim.param(param_g4sim)


##Create paths
main = fw.create_path()

##Add modules to paths
main.add_module(evtmetagen)
main.add_module(evtmetainfo)
main.add_module(paramloader)
main.add_module(geobuilder)
main.add_module(g4sim)
main.add_module(cdcdigitizer)
main.add_module(out)

##Process events
fw.process(main)
```

MC production using gbasf2

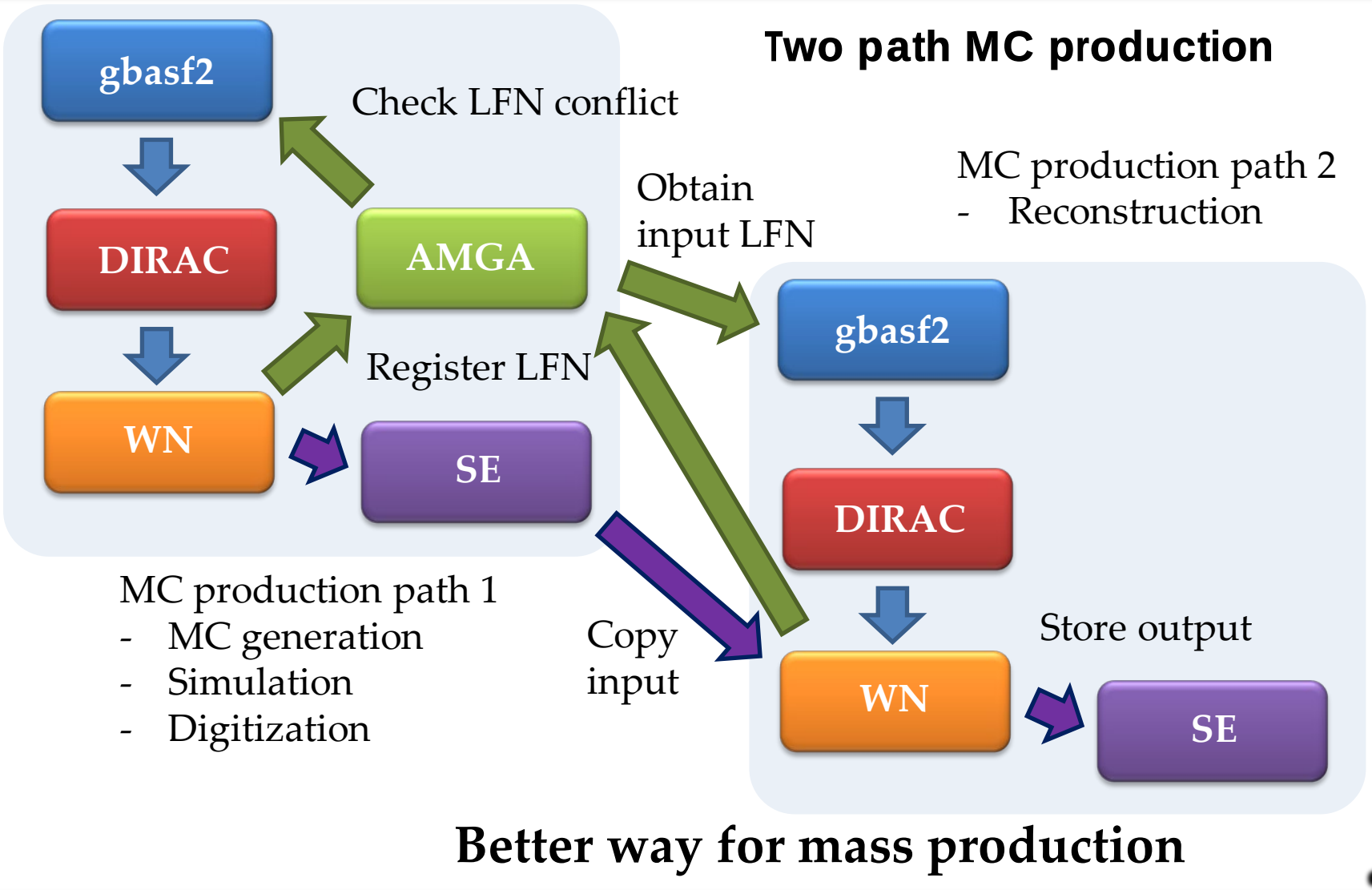


All-in-one
- MC generation
- Simulation
- Digitization
- Reconstruction

One path MC production

- Typical Belle II MC job
  - Generate MC (whole chain)
  - Store data to SE
  - Register metadata to AMGA

More complex case



Better way for mass production

Server status

Recent progress (outline)

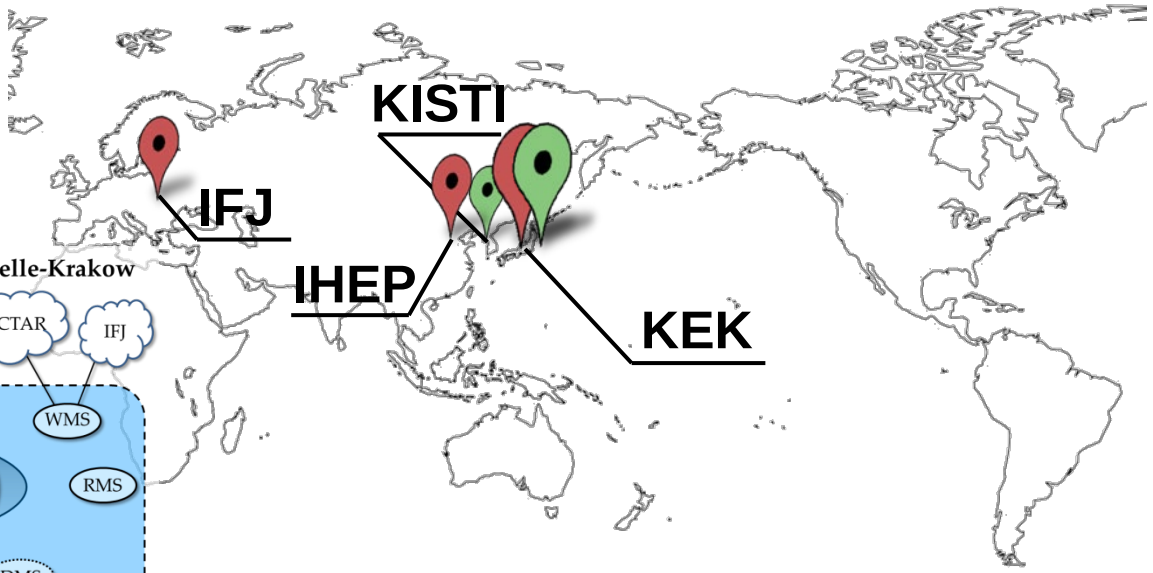
- Before April
  - Neither AMGA nor DIRAC in KEK

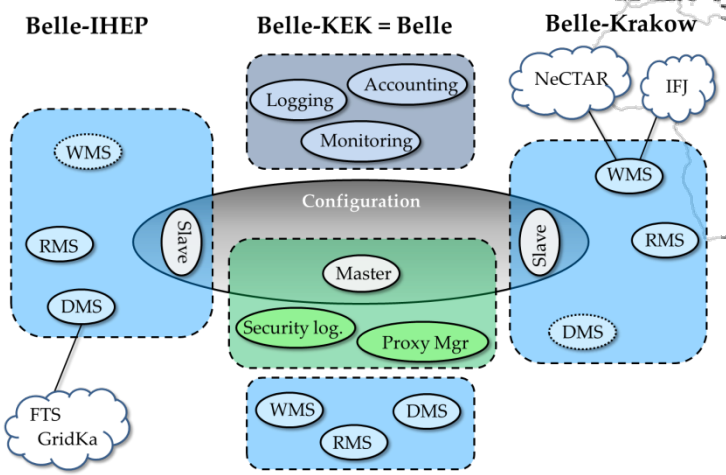
- April
  - Install AMGA and DIRAC on KEKCC
- July
  - Get approval to open service ports from KEK Security Management Committee (took a few months)
  - Reconfigure DIRAC system (by Ricardo Graciani)
- August
  - Distribute Belle II software via CVMFS
  - Run a realistic MC production job using SE
- September
  - Migrate KEK AMGA as a main server

Belle II main servers

DIRAC: master (KEK)
slave (IHEP, IFJ)

AMGA: master (KEK)
slave (KISTI)





Currently Belle II has three DIRAC servers

Current GRID sites

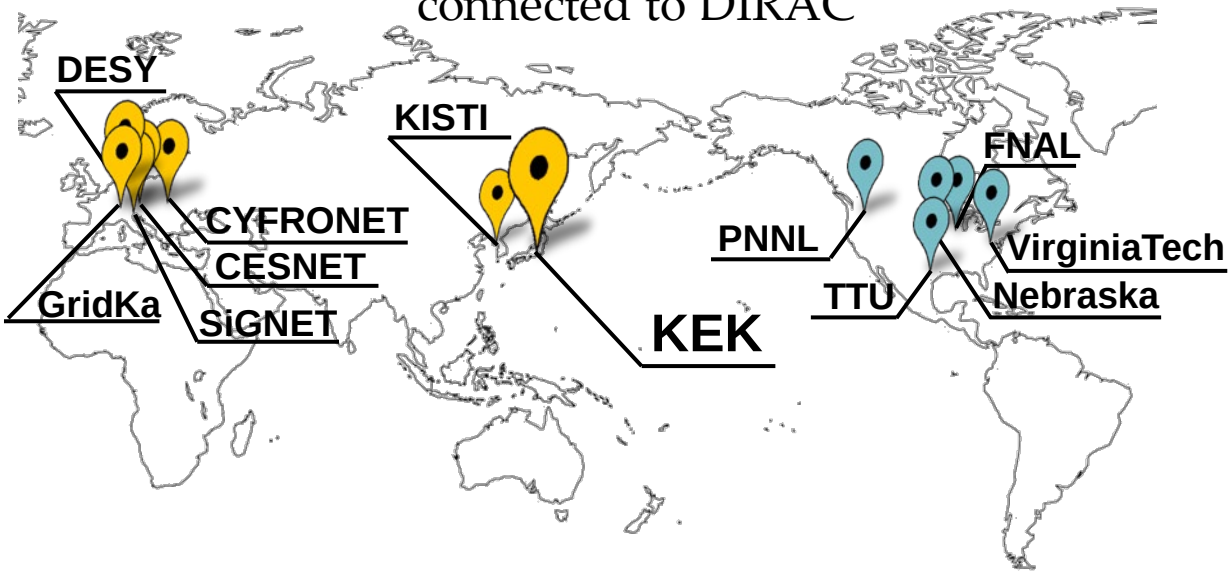
LCG sites

OSG sites

GRID sites currently connected to DIRAC

GRIDs

CE
SE
BDII
WN

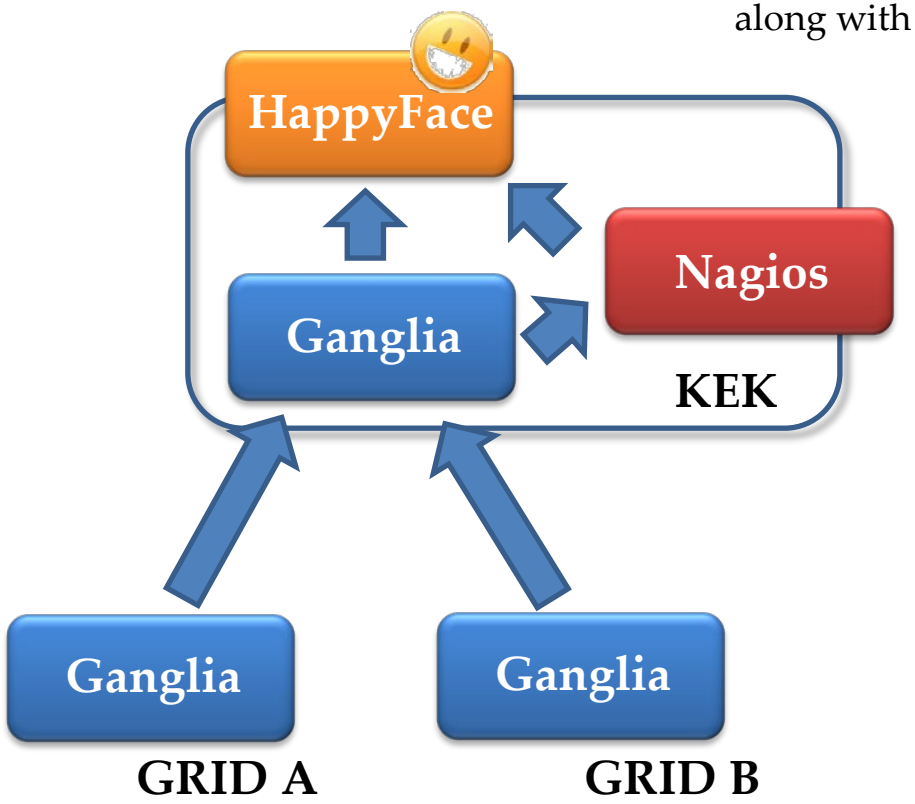


More resources to be incorporated
(including clouds and local clusters)

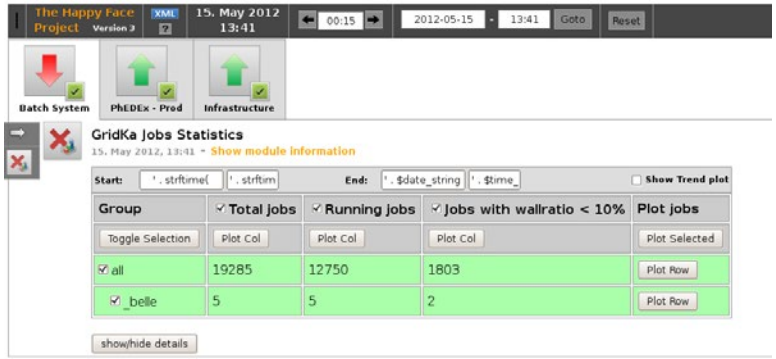
Monitoring

- Ganglia+Nagios monitor AMGA/DIRAC
- Plan to integrate to HappyFace

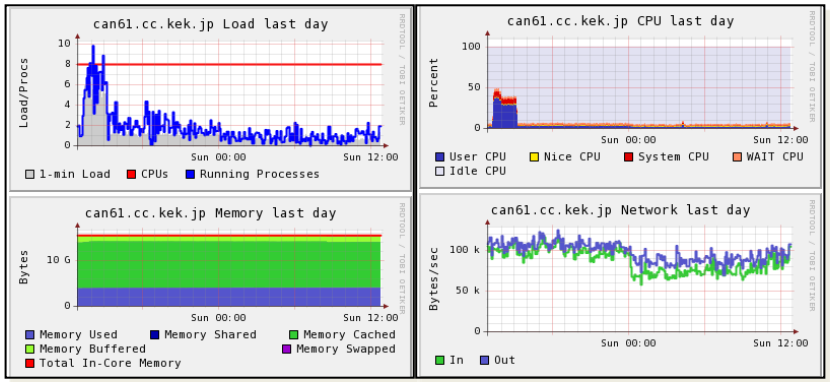
along with perfSONAR?



HappyFace



Ganglia



CVMFS

- Current size of Belle II software tarball ~ 10GB
  - Brings heavy traffic when one run Belle II jobs on environment that doesn't install Belle II distributions

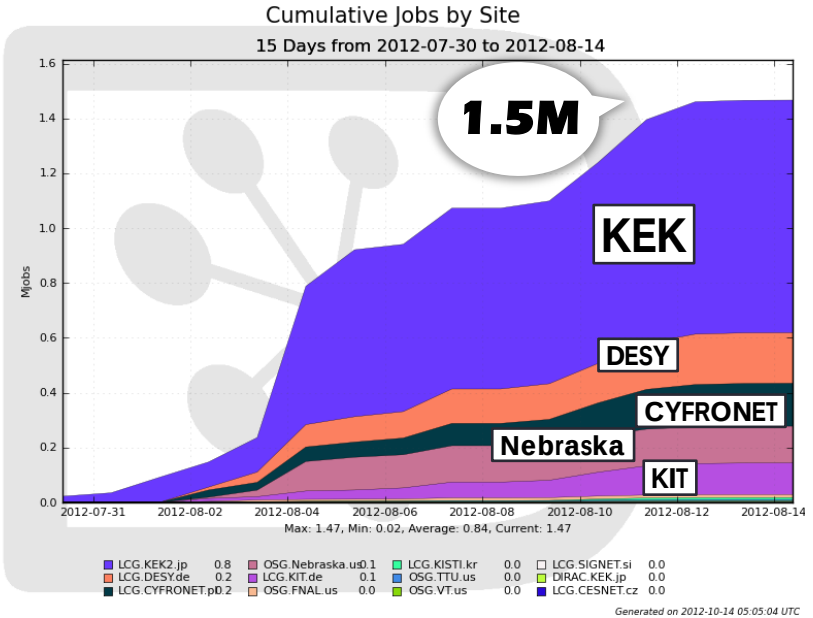
- CERN Virtual Machine File System
- Main feature
  - Based on FUSE
  - mount web directory as local disk
  - Local cache
  - SQUID proxy

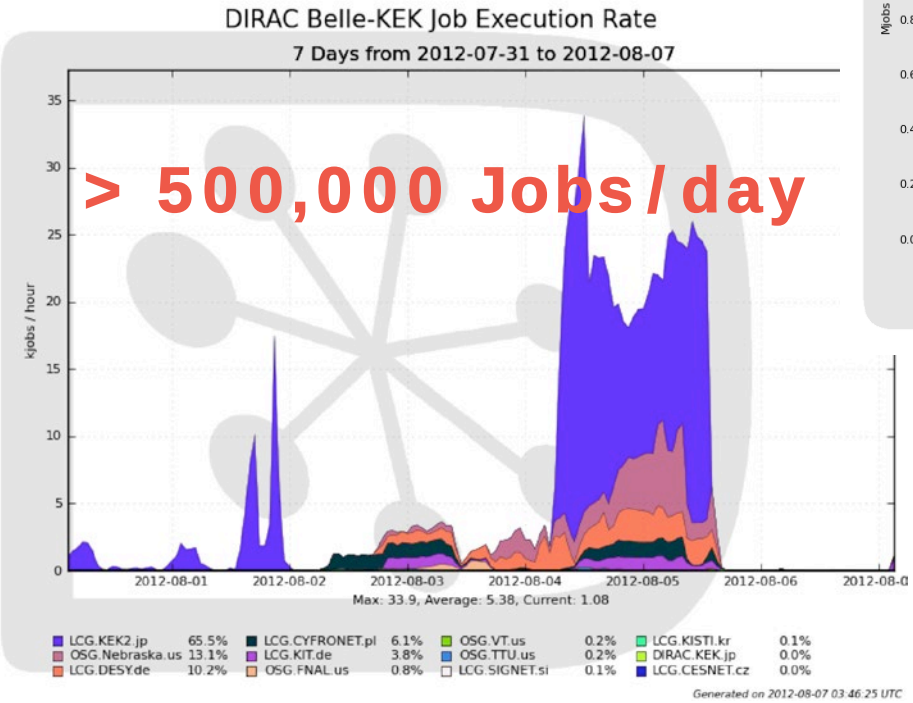
- Many Belle II GRID sites have installed Belle II CVMFS directory: /cvmfs/belle.cern.ch (typical path)
- gbasf2 modified to deal with CVMFS
  - Works pretty well

Current performance

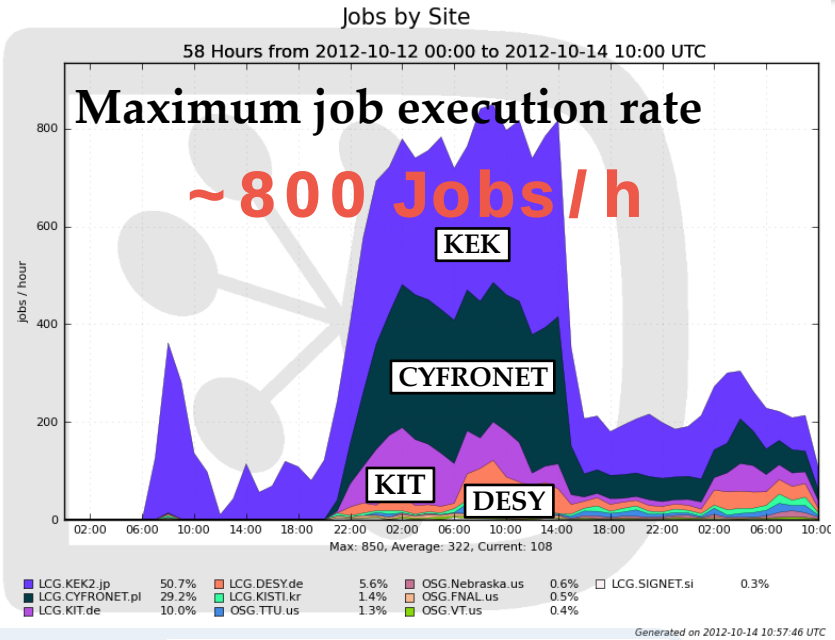
Simple job

- Random number generation (500/job) or just filling pilot job
  →no SE/AMGA used
- Good performance
  - Even saturated KEKCC GRID
- DIRAC itself was stable





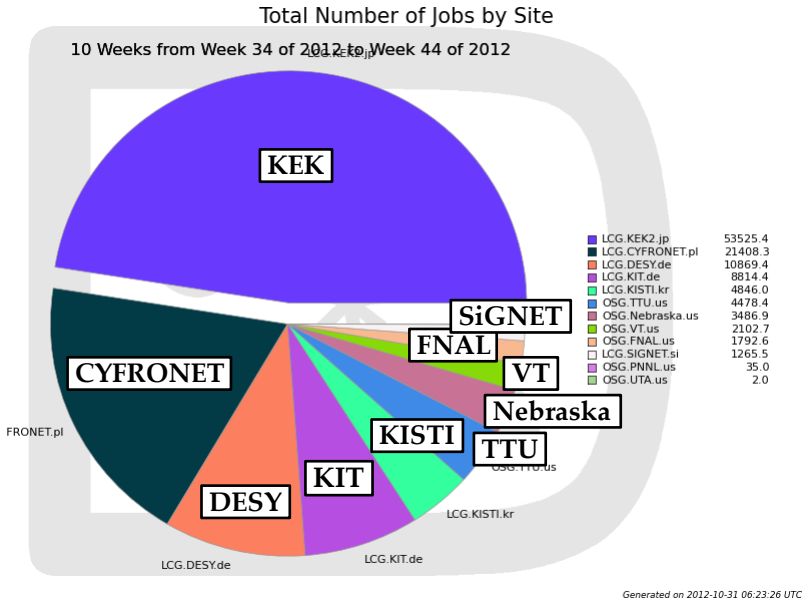
MC production job



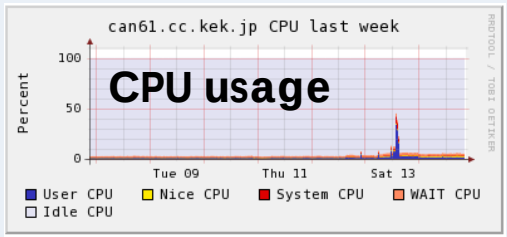
1000 events / job
(in total 200K jobs submitted at the same time)

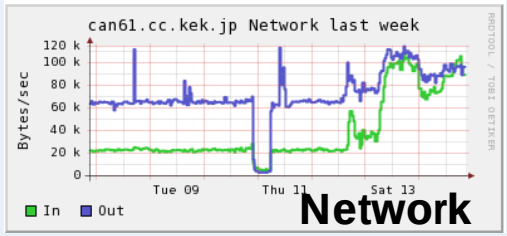
113K jobs/10 weeks

Almost all are MC jobs



Monitor





Plan

- More realistic test (MC mass production)
  - More jobs, more events, more CEs and SEs
  - Find out bottle-neck before physics run
- More realistic workflow
  - Multiple inputs/outputs
  - Develop Belle II specific Workflow extension
- More resources
  - Local resources and clouds
  - Especially resources on each University

Summary

- Belle II GRID environment is getting advanced
  - Main servers centralized to KEK (DIRAC/AMGA)
    - Along with slave servers around the world
  - Successfully introduced CVMFS/monitoring service
  - Performed realistic MC production using SE
    - Need much more heavy load to find out bottle-neck
- Plan to evolve aiming to physics run

Belle II GRID works pretty well!

Backup

DIRAC service ports

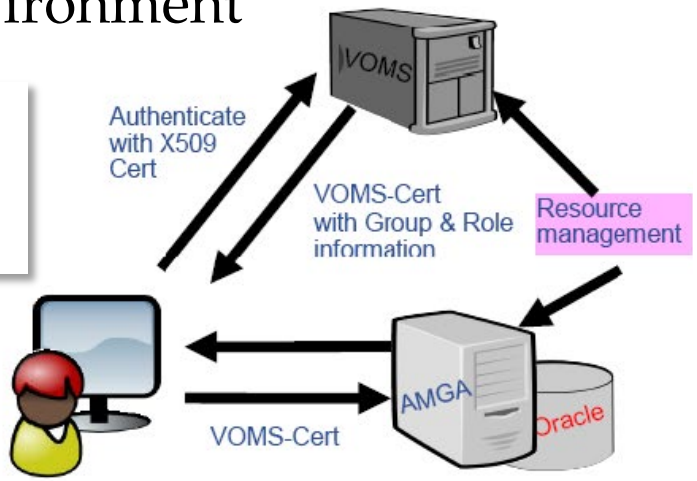
Port	System	Service
80	http	
8080	http redirect	For web portal
443	https	
8443	https redirect	
9130	WorkloadManagement	JobMonitoring
9132	WorkloadManagement	JobManager
9133	Accounting	DataStore
9134	Accounting	ReportGenerator
9135	Configuration	Server
9136	WorkloadManagement	JobStateUpdate
9141	Framework	SystemLogging
9142	Framework	Monitoring
9143	RequestManagement	RequestManager
9144	Framework	SystemLoggingReport
9145	WorkloadManagement	WMSAdministrator
9148	DataManagement	StorageElement
9149	DataManagement	StorageElementProxy
9152	Framework	ProxyManager
9153	Framework	SecurityLogging
9154	Framework	Notification
9155	Framework	UserProfileManager
9157	Framework	Plotting
9158	Framework	BundleDelivery
9162	Framework	SystemAdministrator
9170	WorkloadManagement	Matcher
9196	WorkloadManagement	SandboxStore
9197	DataManagement	FileCatalog

AMGA

- ARDA Metadata Grid Application
  - Metadata server for GRID environment

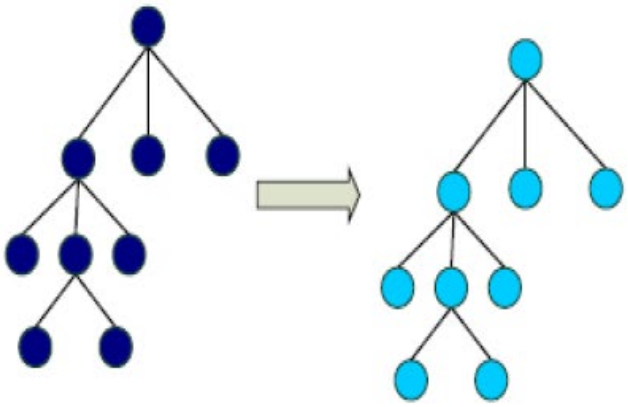
Metadata: data of data

LFN, run range, software version…



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

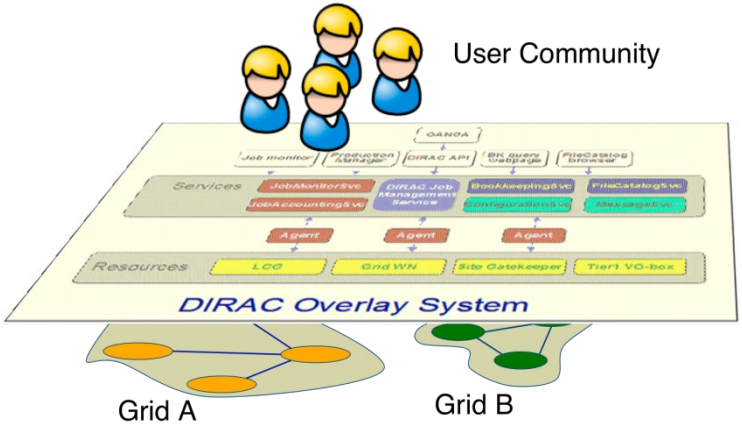
e.g. replication of specific data set or run period for a GRID site



DIRAC

- Distributed Infrastructure with Remote Agent Control
  - Software framework for distributed computing

Developed by LHCb
→ Independent project since 2010 :

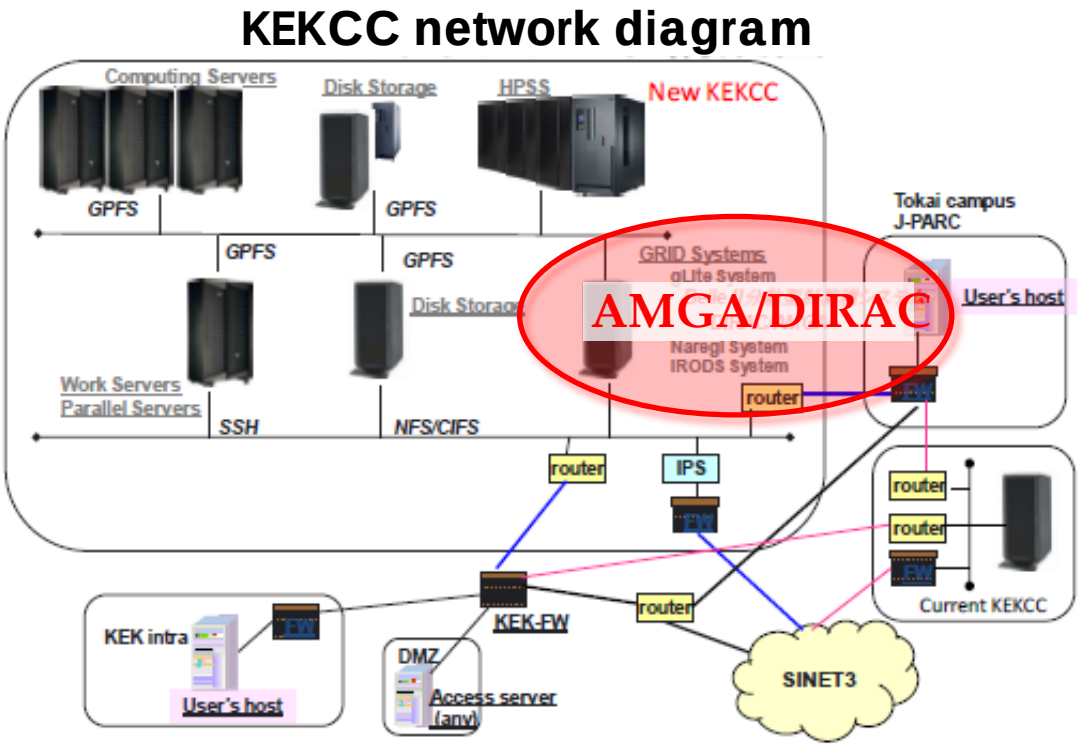


- Main feature
  - Pilot jobs
  - Workload management overall VO
  - Extendible (modular structure)

→ GRID, cloud, and local cluster

LCG OSG

AMGA/DIRAC@KEKCC



- Both AMGA/DIRAC main servers are located inside KEKCC network
- Independent from KEKCC GRID itself

AMGA server status

- Main server: KEK
- Slave server: KISTI

Data migration: copied all of KISTI server data to KEK

- KEK server detail
  - One of KEKCC nodes
    - Xeon 2.93GHz 6*2 threads
    - Memory 48G
    - SL5.7 (x86_64)
    - Storage 1TB (LVM/no-RAID)
    - Database PostgreSQL

Service ports
- 8822 (usual business)
- 8823 (replication)
- 80 (web monitor)

- GUI access (AMGA Manager)
- CUI access (mdclient)

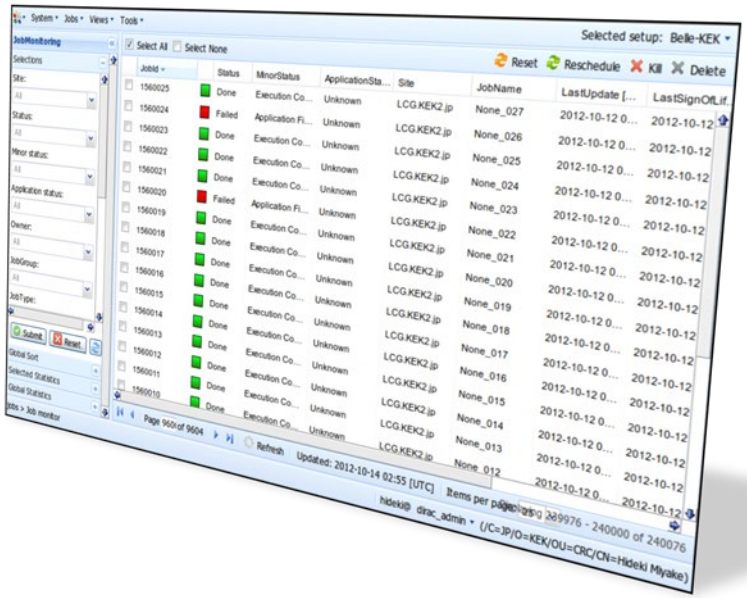
- Store dummy metadata + user metadata ~38M records

DIRAC server status

- Main server: KEK
- Slave server: IHEP, IFJ

- KEK server detail
  - One of KEKCC nodes
    - Xeon 2.67GHz 4*2 threads
    - Memory 16G
    - SL5.7 (x86_64)
    - Storage 600GB (LVM/no-RAID)
    - Single server

Service ports
- Standard DIRAC ports
- 80, 443 (web portal)



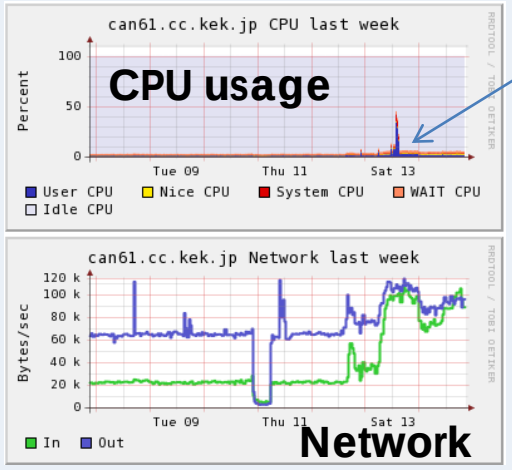
- GUI access (Web portal)
- CUI access (dirac_client, gbasf2)

- KEK server has processed ~1.6M jobs so far

(No) stress test

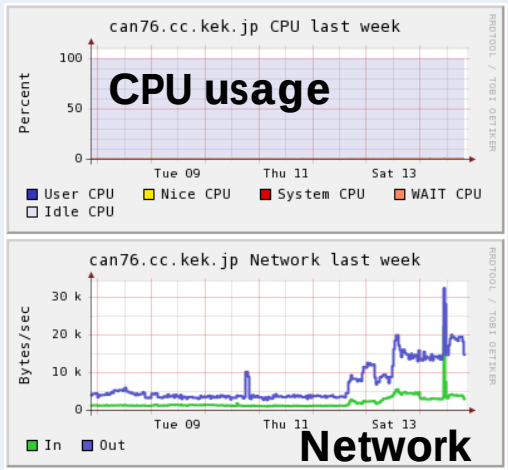
- Submit 200K MC jobs (one half to KEKCC, the other to ANY sites)

DIRAC



Submission

AMGA



- Number of concurrent jobs already saturated (~800/h)
- Generally each load is mild

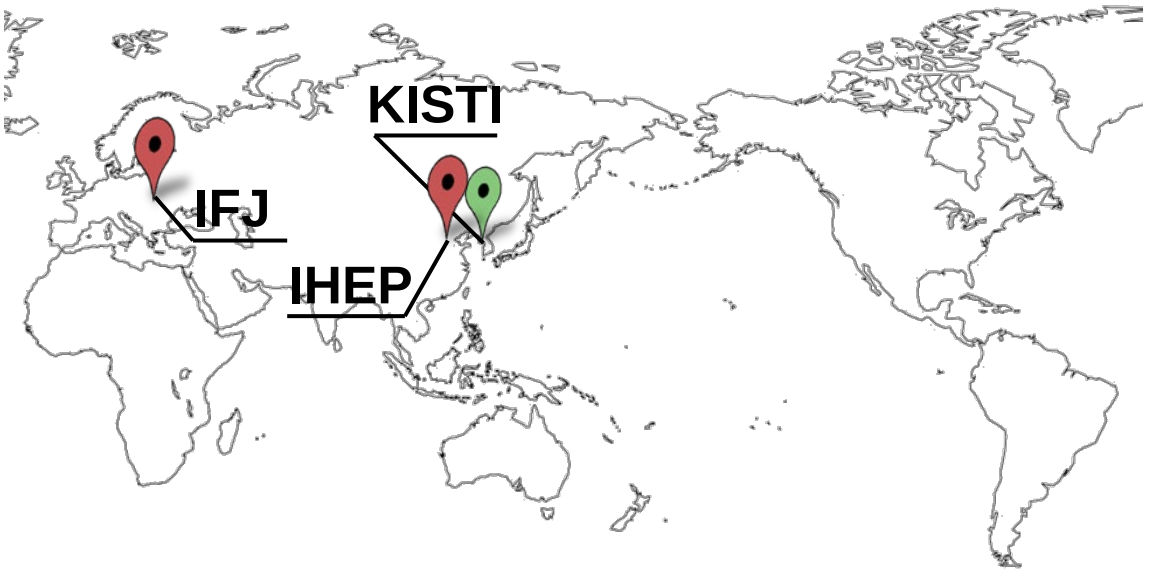
- DIRAC CPU load increased at the job submission but not so high (keeps ~40%)
- AMGA network usage is being increased by returning job, but mild, too

- Current system can handle jobs much more
  - One reason: our sandbox is small

Main servers before April

AMGA: master (KISTI)

DIRAC: master (IFJ)
slave (IHEP)



No main servers at KEK

Shooting! Shooting! Shooting!

We have experienced several issues during main server construction/operation:

- Conflict of name resolving
  - E.g. local name is not registered in DNS
  - Solved by suitable modification of /etc/hosts
- Tight limit of system configuration
  - max number of user processes

Local issues

- Incorrect Site-BDII information
  - Queue announce (DIRAC only refers the information at that time)
  - Number of free job slot
- GRID jobs on GPFS disk
  - Working directory of GRID job was mounted on GPFS

Site wide issues

Thanks to KEK-CRC people, we could resolve many (but minor) issues in a short time so far

DIRAC configuration