Computing at Belle II

Dr. Silvio Pardi (INFN-Napoli)
on the behalf of Belle II collaboration

SUMMARY

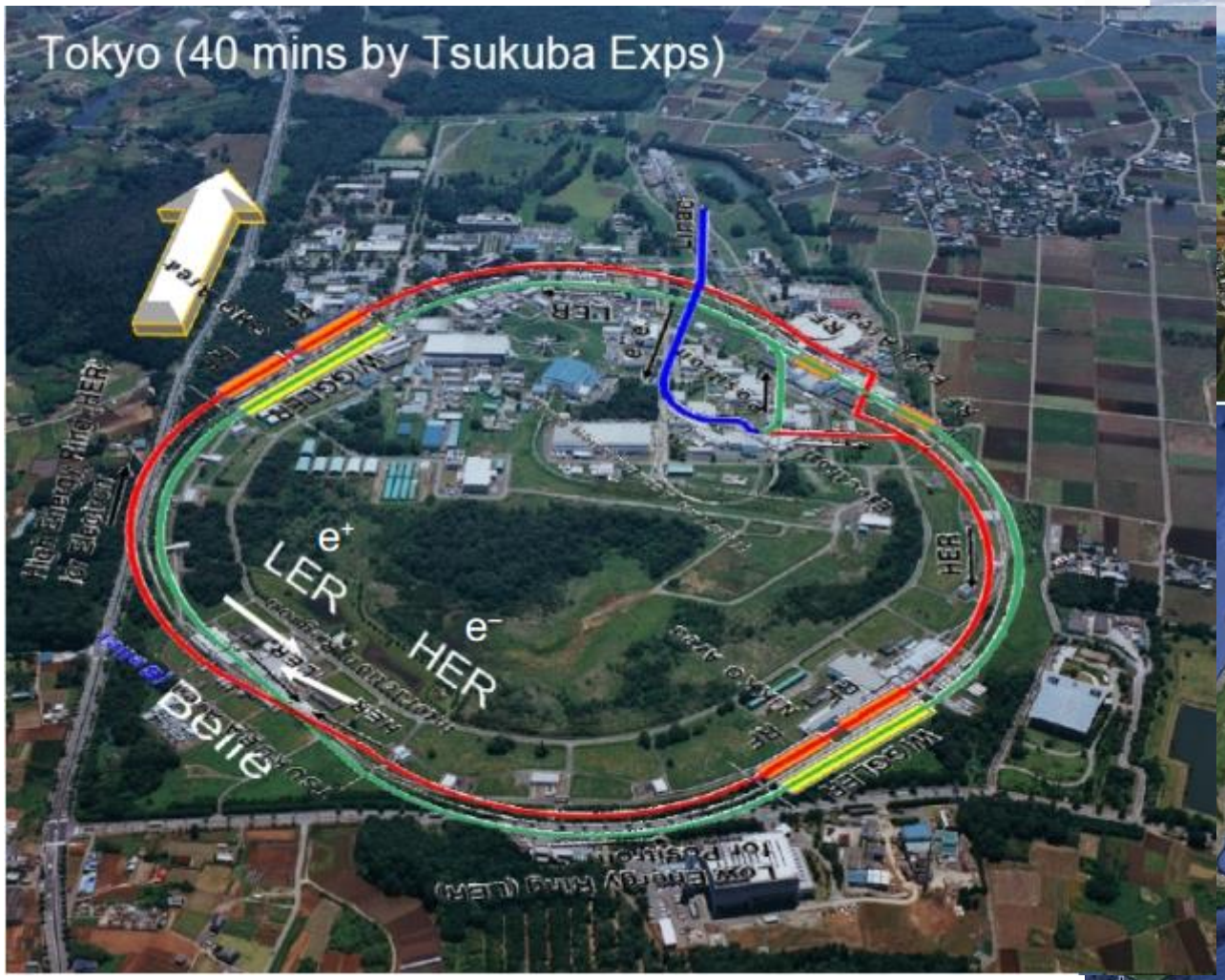
- BELLE II introduction
- Estimation of CPU and Storage needs
- The Computing Model
- Network needs Estimation
- The Future of Network infrastructure
- The production system: DIRAC
- Monte Carlo Campaign
- Reference and Acknowledgements

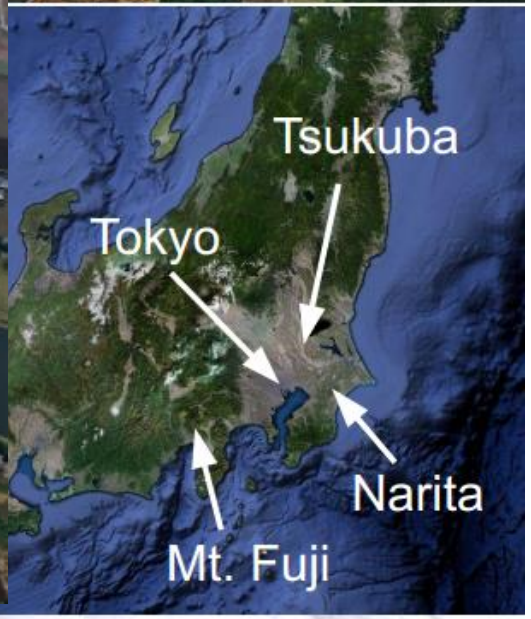
THE BELLE II COLLABORATION

BELLE II is an international collaboration that aims at develop a large scientific program on the Physics of flavor through a new e+e- collider named SuperKEKB, in development at the Japanese laboratory KEK (Tsukuba).

SuperKEKB , is an upgrade of KEKB, it plans to improve of a factor 50 the quantity of data to respect the previous facility.

BELLE II


Tokyo (40 mins by Tsukuba Exps)
e+
LER
e−
HER
Belle


Tsukuba
Tokyo
Narita
Mt. Fuji

THE BELLE II COLLABORATION

Belle II

23 countries/regions
97 institutes
577 colleagues

as of June 30, 2014

Asia : ~45%	N. America : ~15%	Europe : ~40%
Japan : 137	US : 63	Germany : 83
Korea : 34	Canada : 17	Italy : 59
Taiwan : 22		Russia : 37
India : 20		Slovenie : 14
China : 15		Austria : 14
Australia : 18		Poland : 11

BELLE II COMPUTING MODEL

The BELLE II Computing model has to accomplish, in a geographically distributed environment, the following main tasks:

- RAW data processing
- Monte Carlo Production
- Physics analysis
- Data Storage and Data Archiving

On going activities

- Resource Estimation
- Define strategy for analysis and data distribution
- Individuating technologies

STORAGE & TAPE

Storage estimation

- RAW Data
- mDST (data set produced from the raw data with saving the reconstructed results.)
  - mDST after data taking
  - mDST during data reprocessing
  - mDST-Monte Carlo related the data
  - mDST-Monte Carlo related data reprocessing

The current parameters for data estimation are
Event Size x RAW Data: 300Kb
Event Size x mDST : 40Kb

X10^9	Year1	Year2	Year3	Yera4	Year5	Year6	Year7
Event/year	1.2	3.1	29.6	70.7	87.8	89.3	93.5
Integrated	1.2	4.3	33.9	104.6	192.4	281.7	375.2

CPU NEEDS

The needs of CPU in term of kHepSPEC, based on the currently available Belle II software.
The CPU ratio for the MC and RAW data processing is extrapolate from the Belle experience.

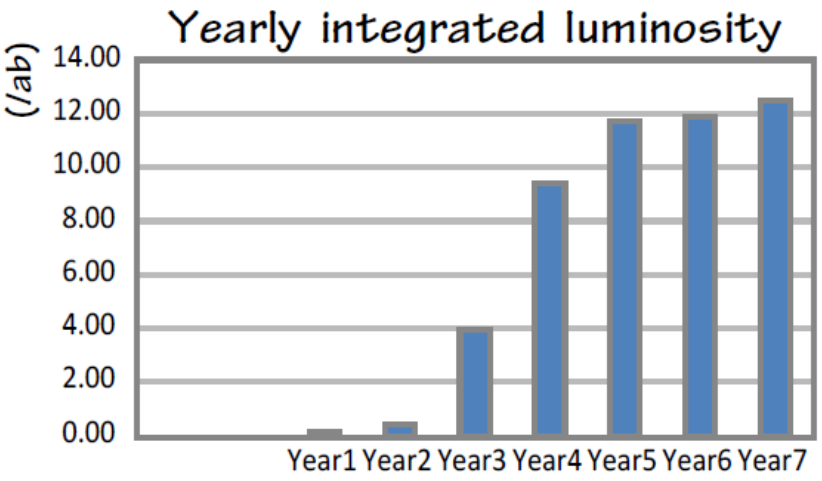
- RAW Data processing: 45HepSPEC*s/ev
- Monte Carlo processing: 89HepSPEC*s/ev

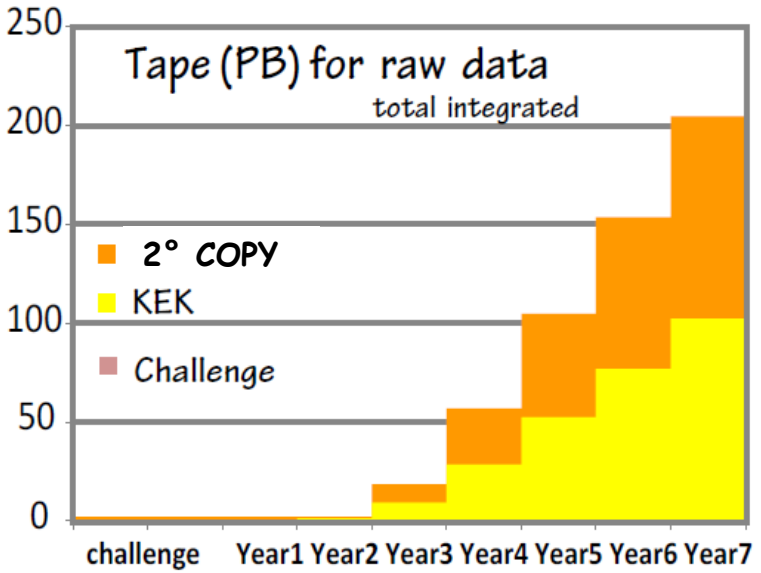
In current hypothesis is estimated an high data reprocessing activity in the first three years of data taking, and then one reprocessing every 2 years in the second phase.

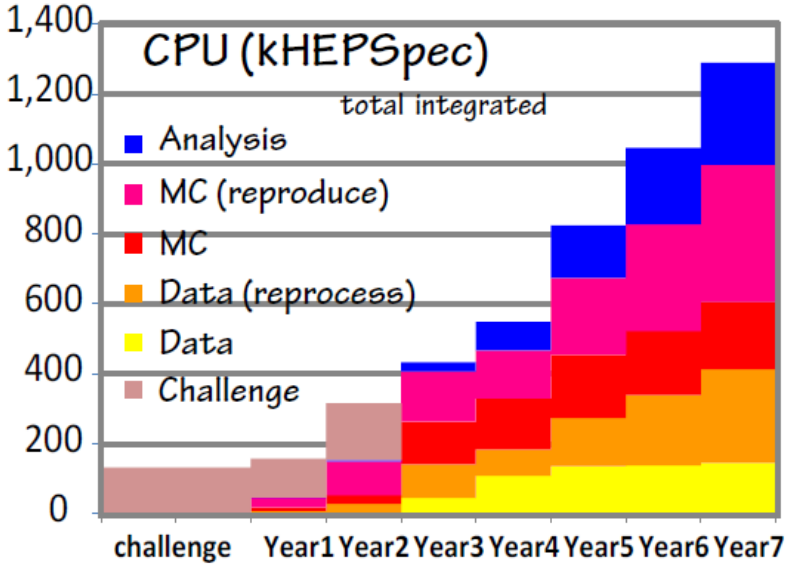
	Year1	Year2	Year3	Year4	Year5	Year6	Year7
#repro	4	4	2	0.5	0.5	0.5	0.5

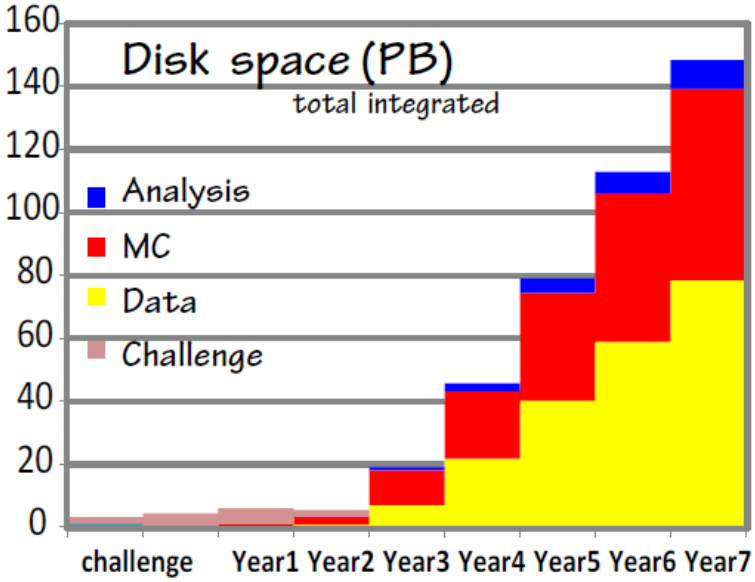
BELLE II RESOURCE ESTIMATION

Belle II

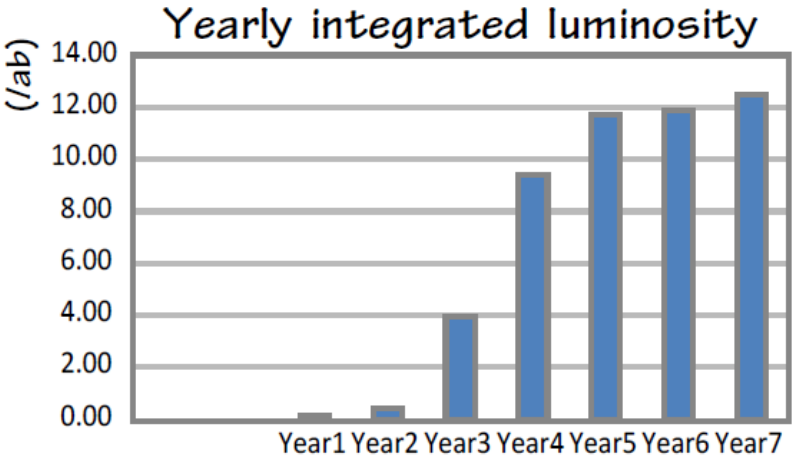


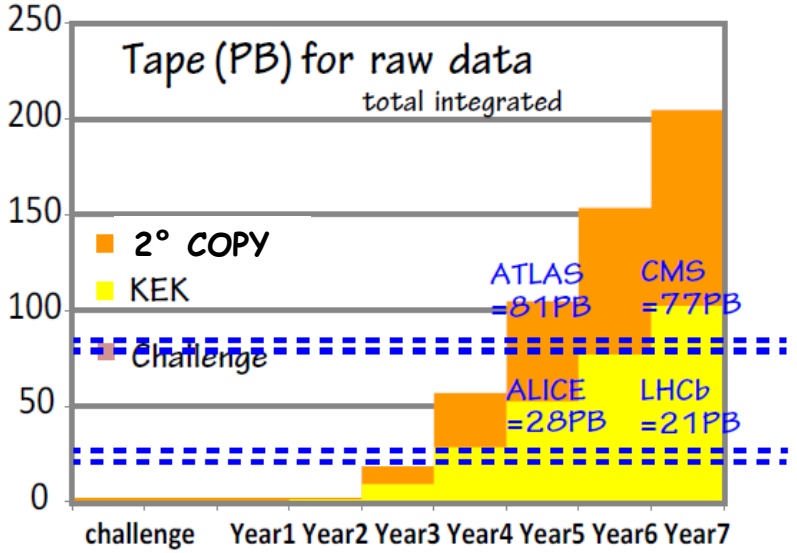


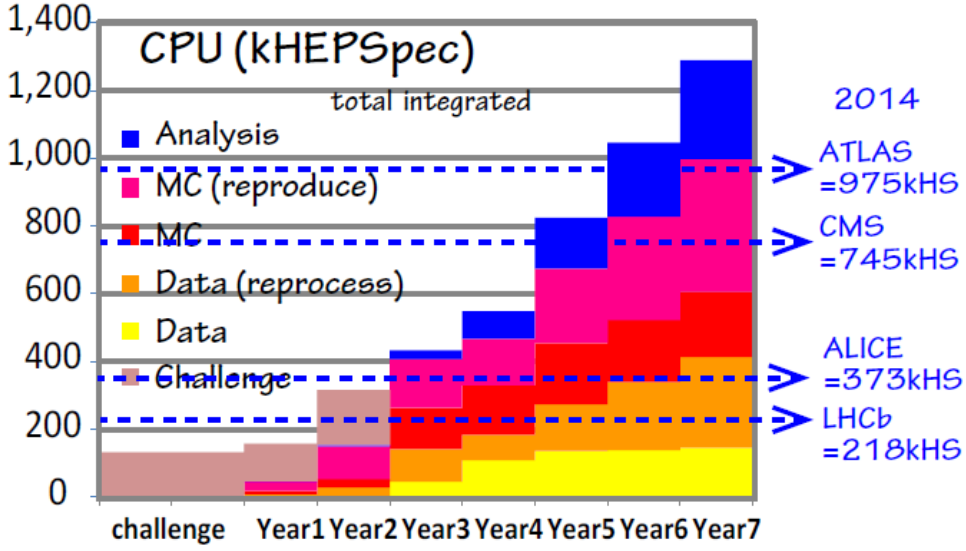


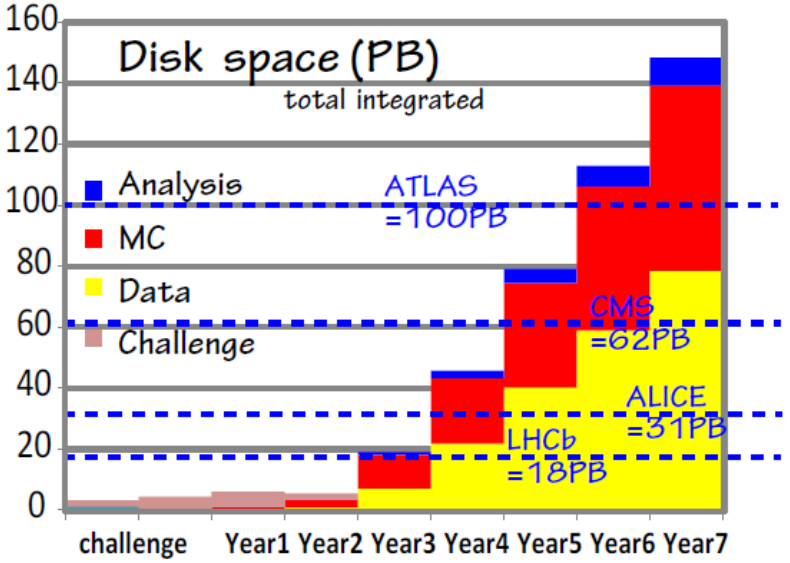


BELLE II RESOURCE ESTIMATION





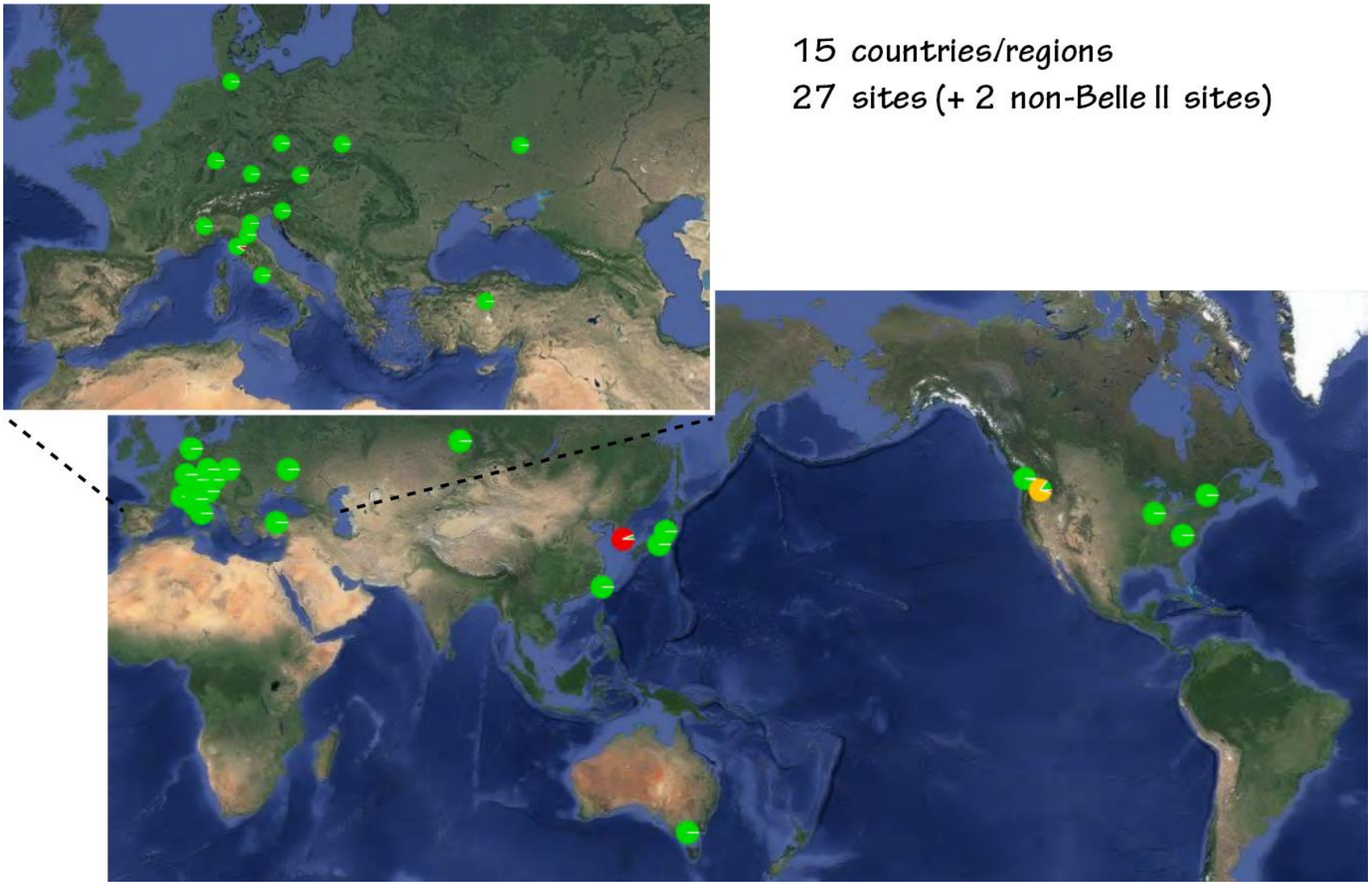




Pledge summary of LHC experiments : http://wlcg-rebus.cern.ch/apps/pledges/summary/

BELLE II COMPUTING SITES

15 countries/regions
27 sites (+ 2 non-Belle II sites)



SITE CLASSITICATION

The BELLE II Computing Sites are classified as follow:

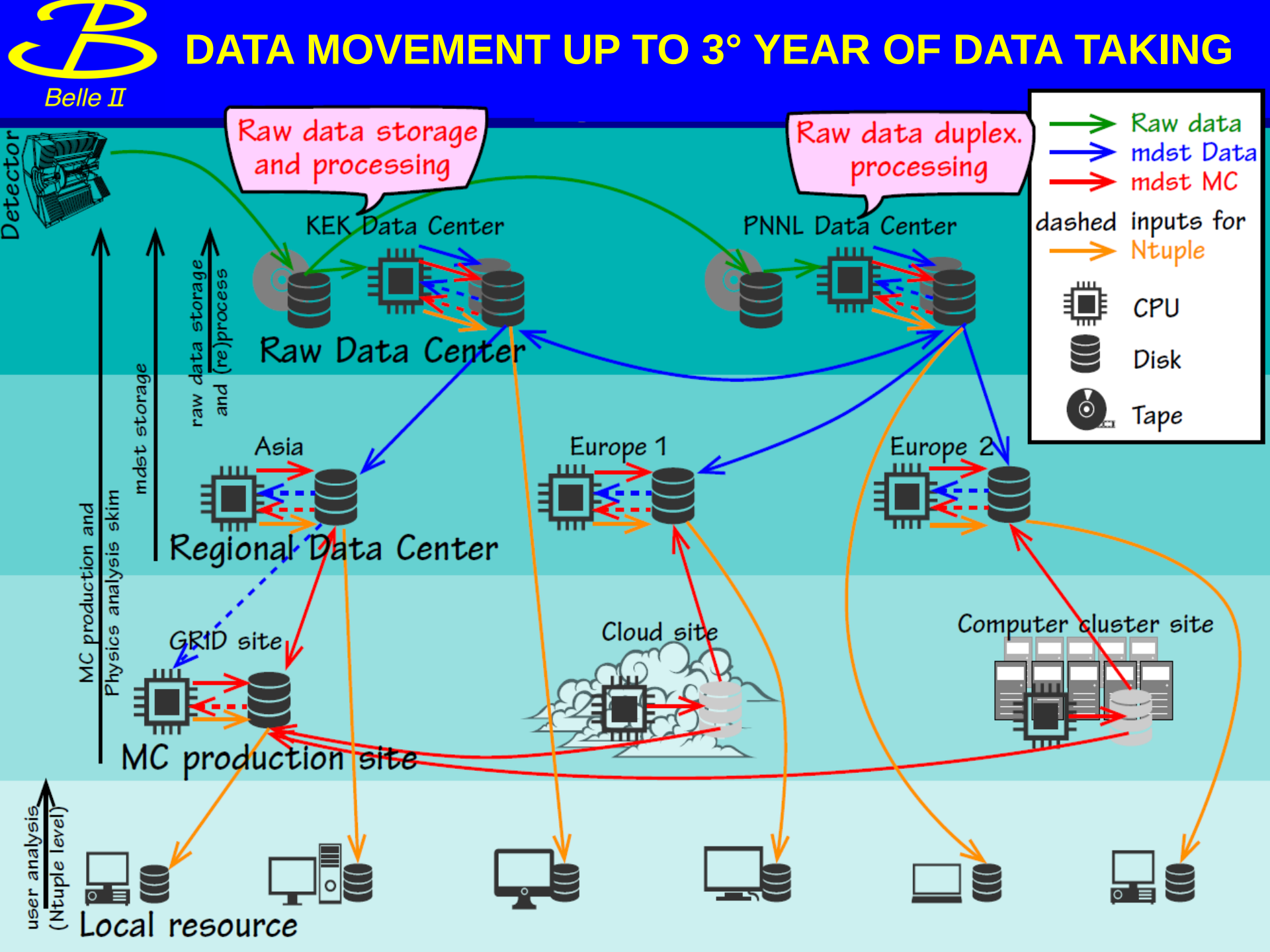
- **Raw Data Center:** Who store the RAW Data and made data processing and/or data reprocessing.

- **Regional Data Center**: Large data center that stores mDST and participates at the Monte Carlo production

- **MC Production site**: Data Center that produces and stores Monte Carlo simulations, that included:
  - Grid Site
  - Cloud Site
  - Computing Cluster Site

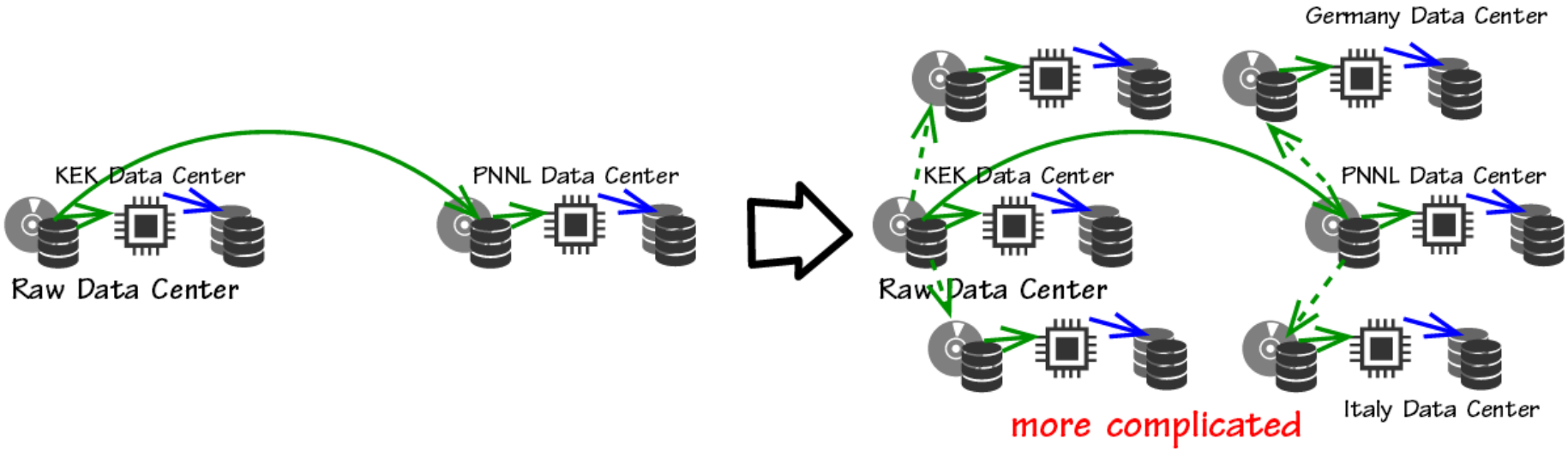
DATA MOVEMENT UP TO 3° YEAR OF DATA TAKING

Belle II



NEW RAW DATA MOVEMENT

Schema of the 2nd raw data copy distribution starting from the 4° year of data acquisition.



NETWORK REQUIREMENTS

Starting from the data distribution schema is possible to estimate the network requirements up to the maximum luminosity.

The Main Issue are:

- Individuate all the data flows
- Deal with high latency infrastructure:
  - 200ms EU/USA
  - 450ms EU/JAPAN

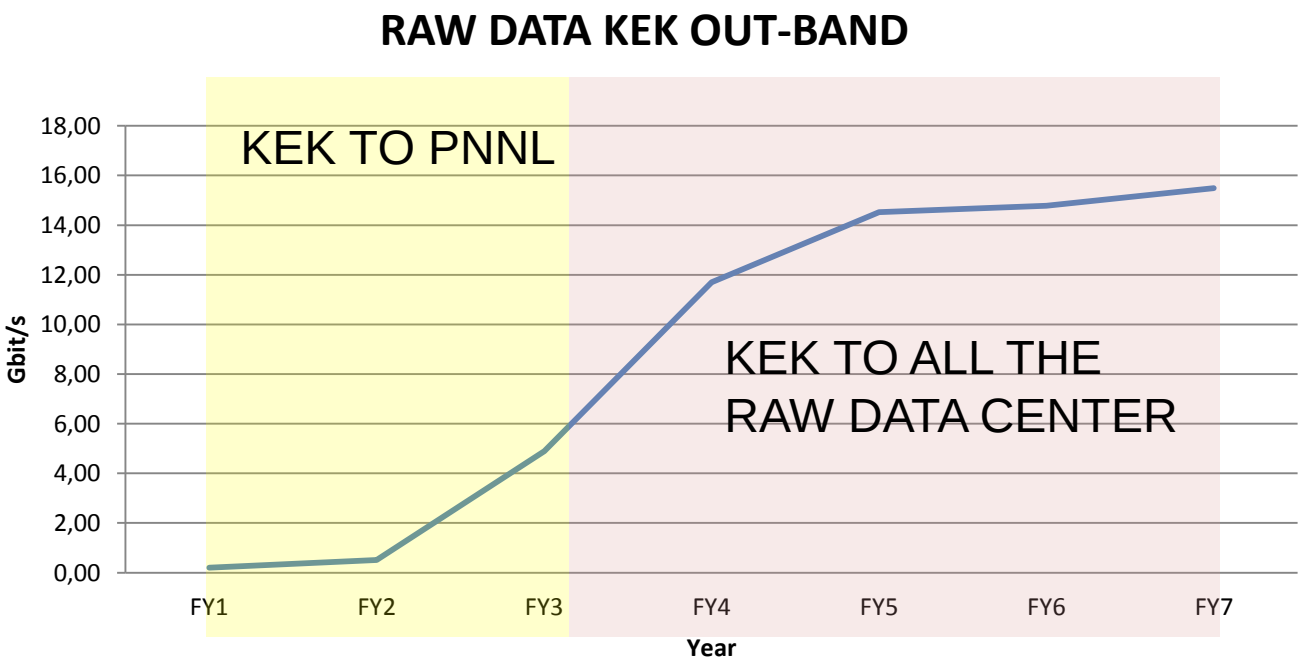
RAW DATA

RAW data produced at KEK and then replicate in other sites

Event size= 300k,Month=11, Tollerance:50%

	FY1	FY2	FY3	FY4	FY5	FY6	FY7
Event Rate	1,20E+09	3,08E+09	2,96E+10	7,07E+10	8,78E+10	8,93E+10	9,35E+10
RAW Data (PB)	0,33	0,84	8,06	19,28	23,94	24,35	25,52
Network Req+toll (Gbit/s)	**0,20**	**0,51**	**4,89**	**11,70**	**14,53**	**14,78**	**15,48**

Network peak
From 200Mbit/s
Up to 15,48
Gbit/s



mDST from raw data

mDST from data taking (60% produced at KEK and 40% in other RAW Data Centers):

Event size= 40k,Months=11, Tollerance:50%

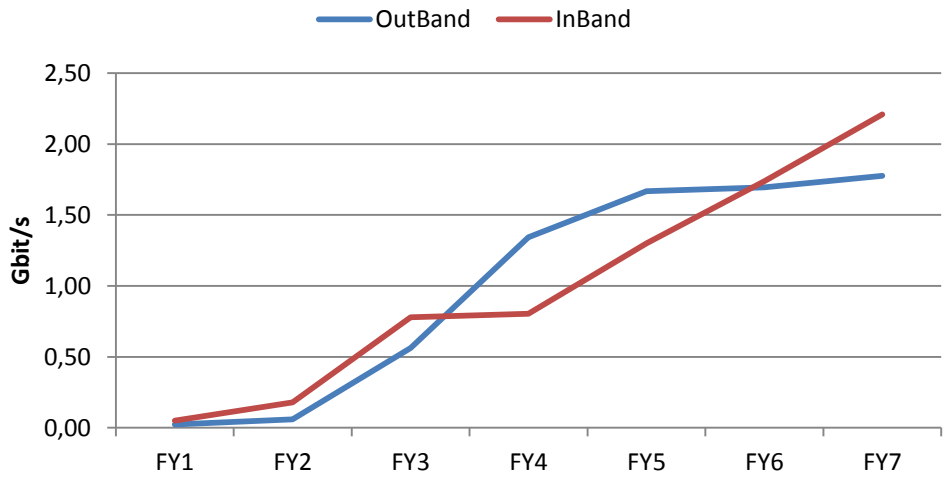
mDST from data reprocessing:

Event size=40k, Months=12, Tollerance:50%

mDST to distribute in Time scale:

	FY1	FY2	FY3	FY4	FY5	FY6	FY7
Event Rate	8,00E+08	2,05E+09	1,97E+10	4,71E+10	5,85E+10	5,95E+10	6,24E+10
MDST - Data (PB)	0,03	0,07	0,72	1,71	2,13	2,16	2,27
MDST-SUM	0,03	0,10	0,82	2,53	4,66	6,83	9,09
# Reprocessing	4	4	2	0,5	0,5	0,5	0,5
MDST-Repro.(PB)	0,12	0,41	1,64	1,27	2,33	3,41	4,55
Gbit/s REPRO	0,05	0,17	0,65	0,51	0,93	1,36	1,81
Gbit/s mDST	0,01	0,03	0,31	0,75	0,93	0,94	0,99

For the KEK Datacenter
Network peak
OutBand
From 20Mbit/s Up to 1.78Gbit/s
InBand
From 50Mbit/s Up to 2.21 Gbit/s



MONTECARLO

Two flows of mDST from MC production

- **mDST-MC produced during data taking**
- **mDST-MC produced during data reprocessing**

All the produced mDST-MC are distributed in that way:

Each MC Production Site sent the produced mDST-MC data to 2 other sites. (The topology is not defined at this stage)

Is a many-to-may network connection

MONTECARLO

mDST-MC - data taking:

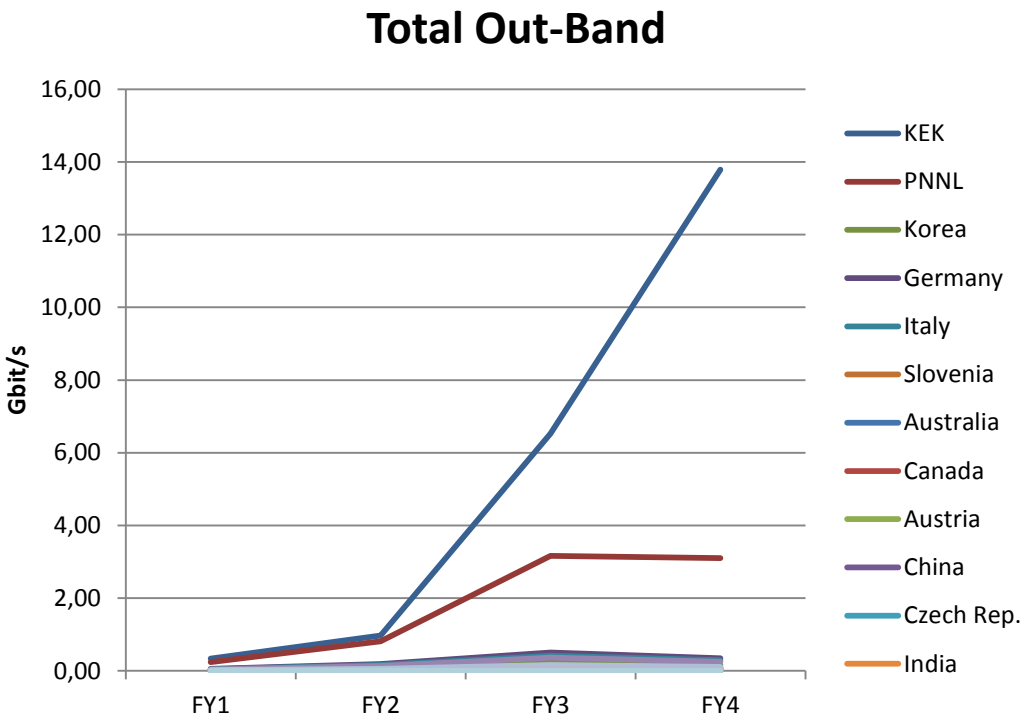
Event size= 40k,Months=11, Tollerance:50%

mDST-MC - data reprocessing:

Event size=40k, Months=12, Tollerance:50%

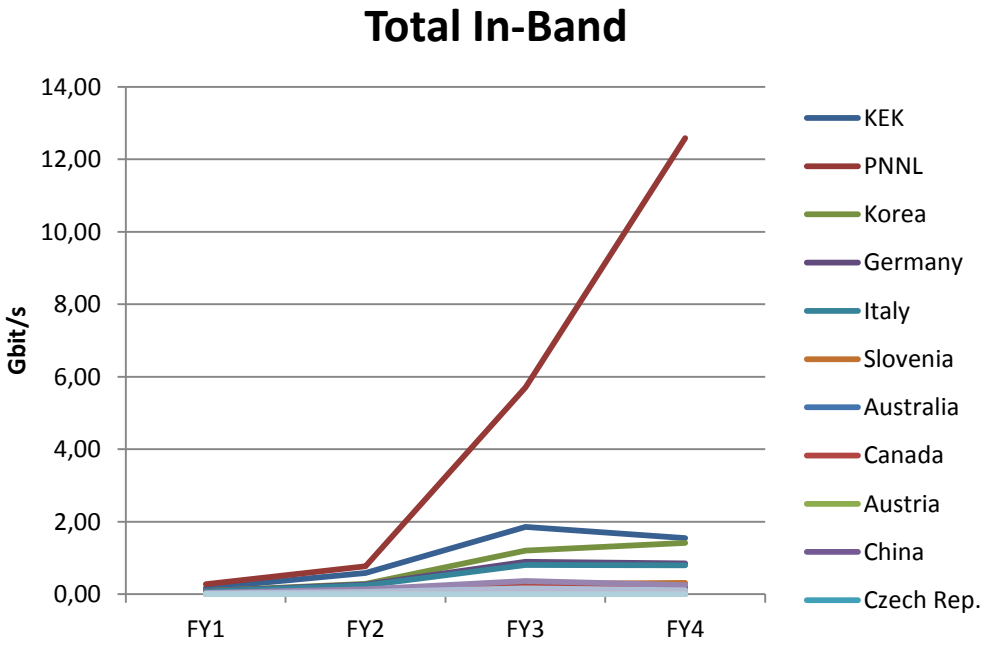
	FY1	FY2	FY3	FY4	FY5	FY6	FY7
Event Rate	1,60E+08	4,10E+08	3,94E+09	9,42E+09	1,17E+10	1,19E+10	1,25E+10
MDST - Data (PB)	0,01	0,01	0,14	0,34	0,43	0,43	0,45
# Reprocessing	4	4	2	0,5	0,5	0,5	0,5
MC-Stream	20	20	10	5	5	5	5
MC-MDST(PB) Per Year	0,12	0,30	1,43	1,71	2,13	2,16	2,27
MC-MDST-TOT(PB)	0,12	0,41	1,85	3,56	5,69	7,85	10,12
MC-MDST-Reprocessing	0,47	1,66	3,70	1,78	2,84	3,93	5,06
MC-Ch(PB)	5,00	2,00	0,00	0,00	0,00	0,00	0,00
MC-MDST	0,58	1,96	5,13	3,49	4,97	6,09	7,33
BW MC Gbit/S	0,05	0,13	0,62	0,75	0,93	0,94	0,99
BW Repro Gbit/s	0,19	0,66	1,47	0,71	1,14	1,57	2,02
BW MC-Ch Gbit/s	1,62	0,55	0	0	0	0	0
Total	0,19	0,66	1,47	0,71	1,14	1,57	2,02

TOTAL NETWORK UP TO THE 4° YEAR



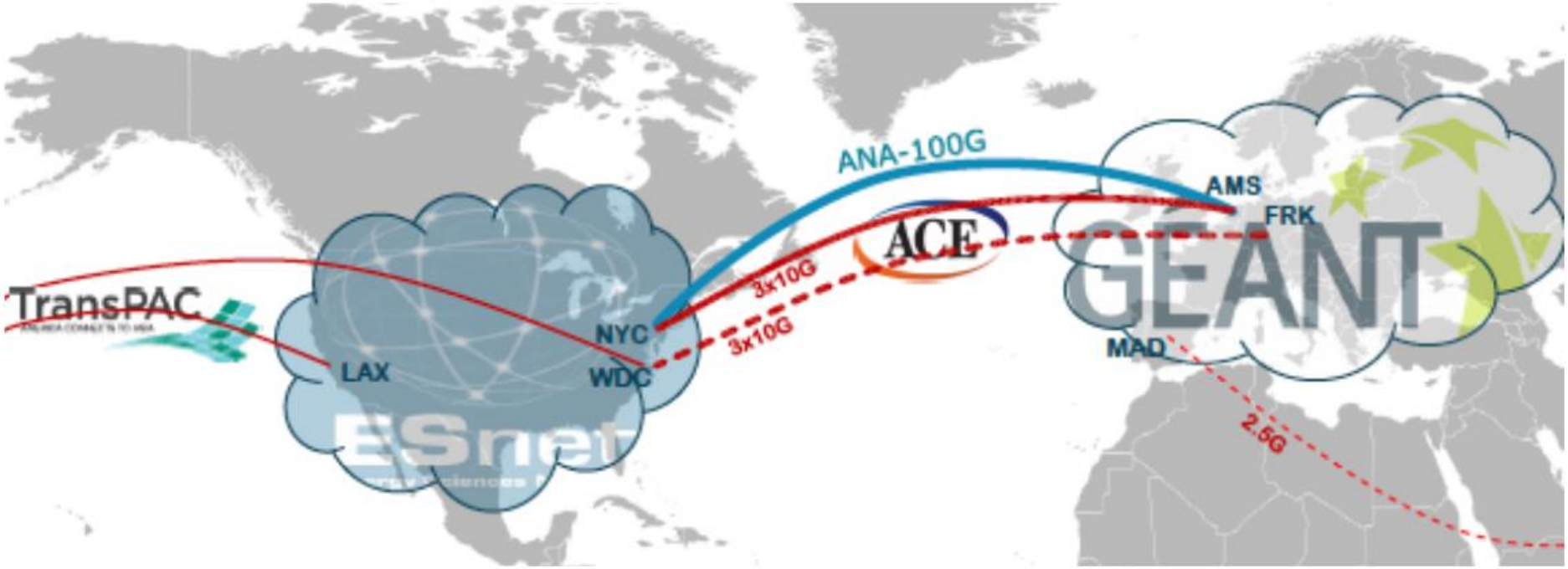
Peak of ~ 14 Gbit/s for KEK

Peak of ~ 12 Gbit/S for PNNL



CURRENT NETWORK

- Trans-Asia
  - 2x10G Tokyo-LA
  - 2.5G Madrid-Mumbai
- Trans-Atlantic
  - 3x10G Amsterdam-NY
  - 3x10G Frankfurt-Washington
  - ANA-100G Amsterdam-NY (temporary)



TEST LINK ANA-100

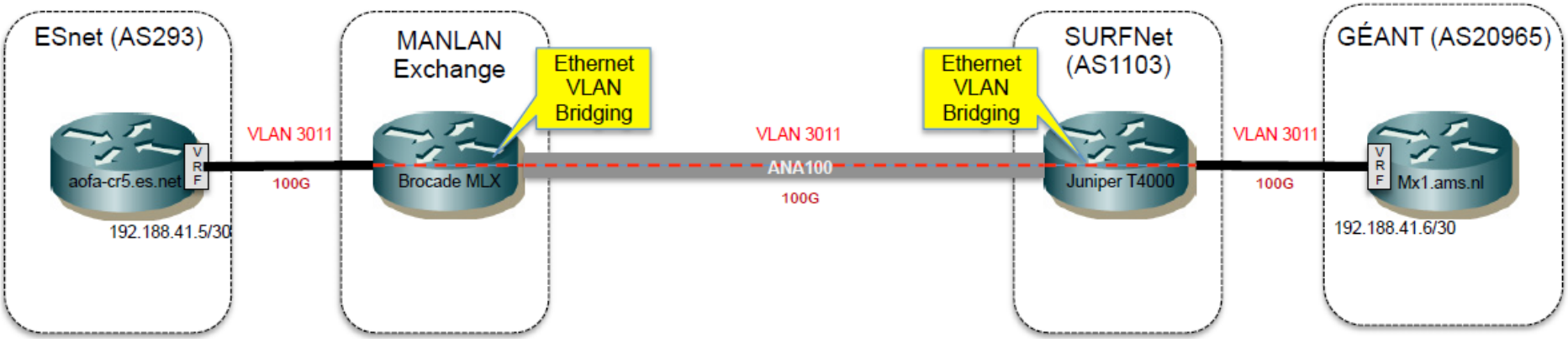
Vincenzo Capone, Aleksandr Kurbatov, Mian Usman

Chin Guok

Marco Marletta

Thomas Schmid, Hubert Weibel

In june 2014 BELLE II tested the link ANA-100
Belle II sites: INFN Napoli, CNAF, KIT, and PNNL



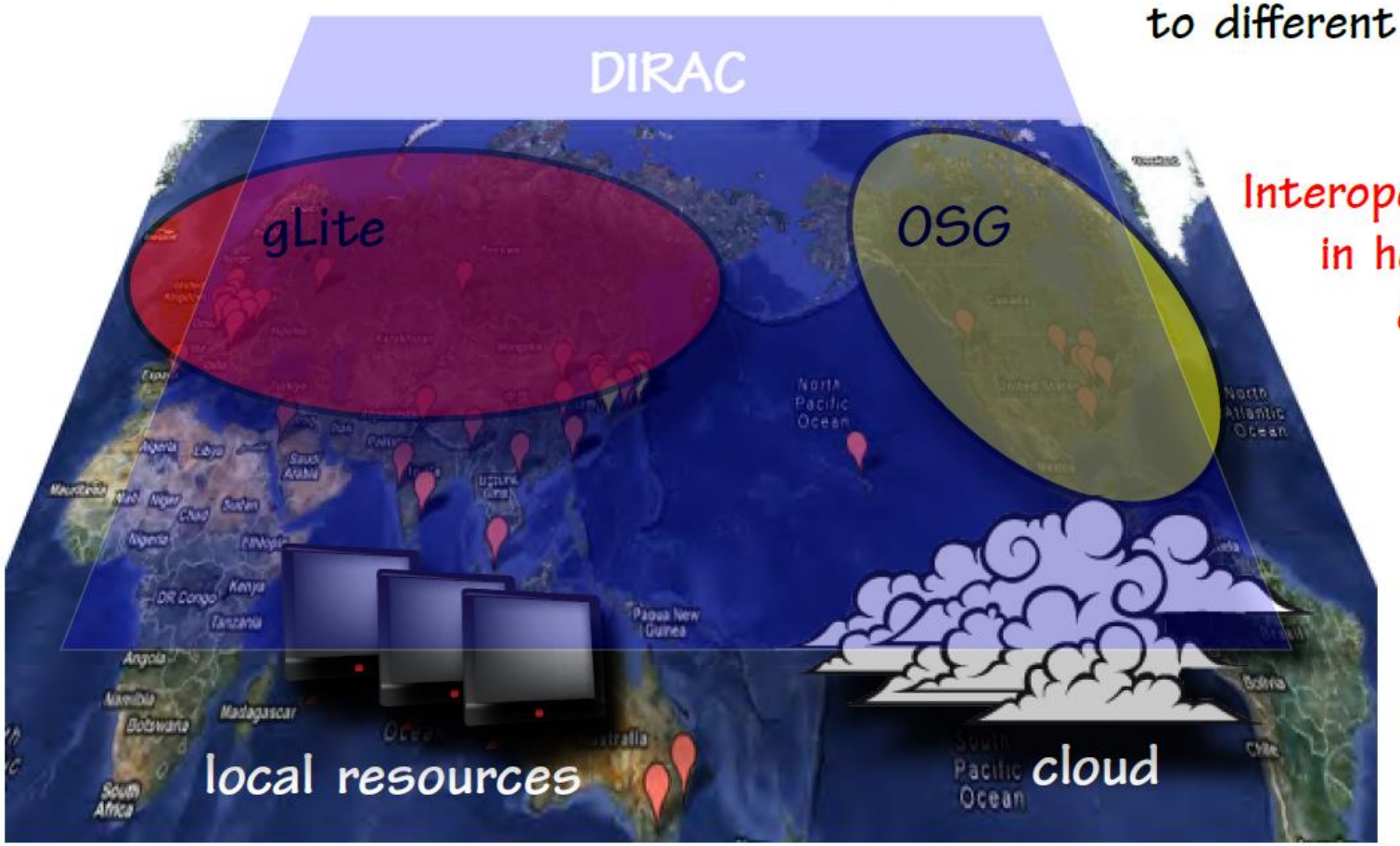
FUTURE UPGRADE

- EEX (ESNet Extension to Europe)
  - 2x100G New York – London
  - 100G Washington – Geneva
  - 40G Boston - Amsterdam
- 10G Mumbai – GEANT

GEANT is studying an additional link from Japan to Europe

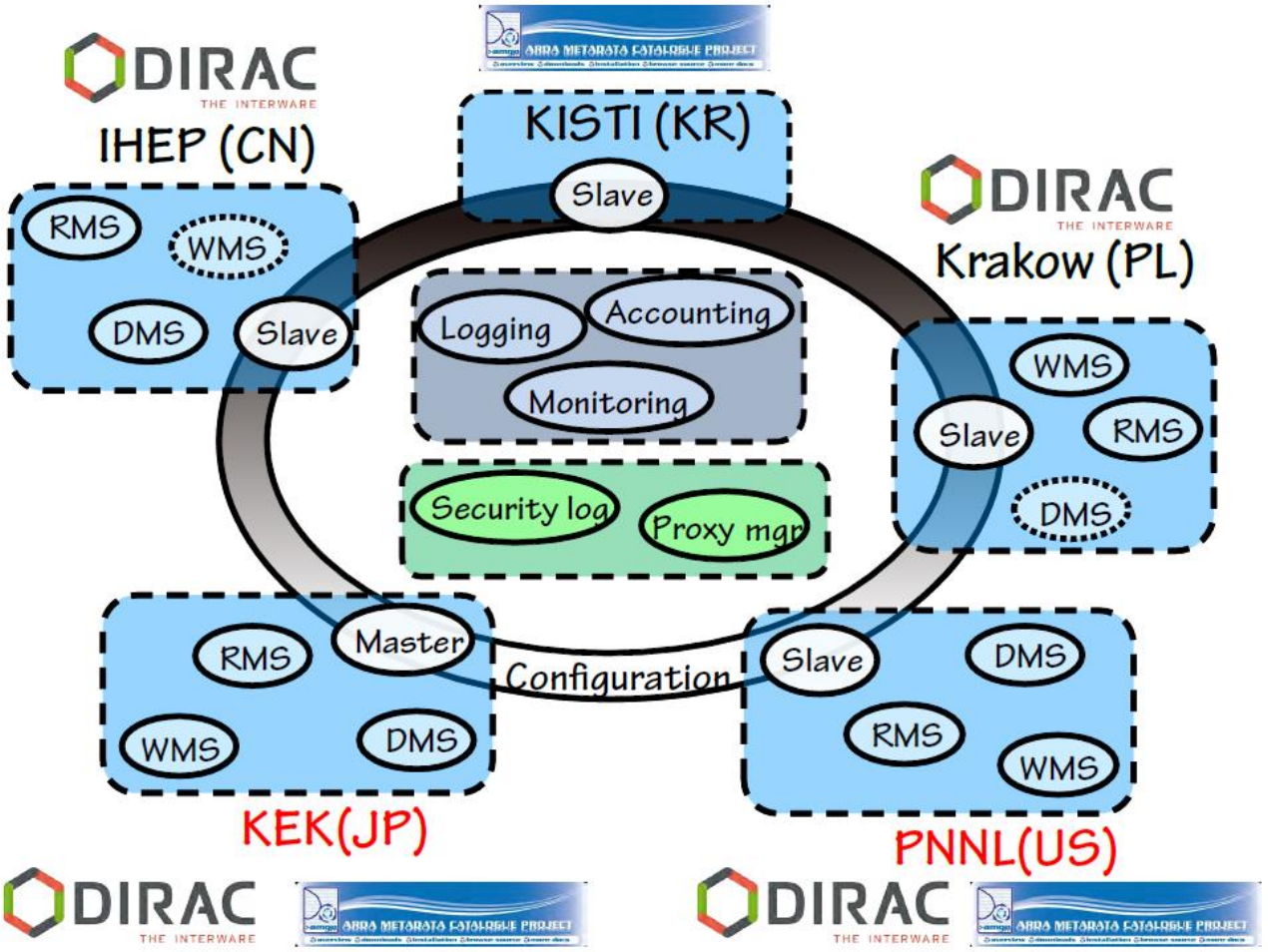
PRODUCTION SYSTEM

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



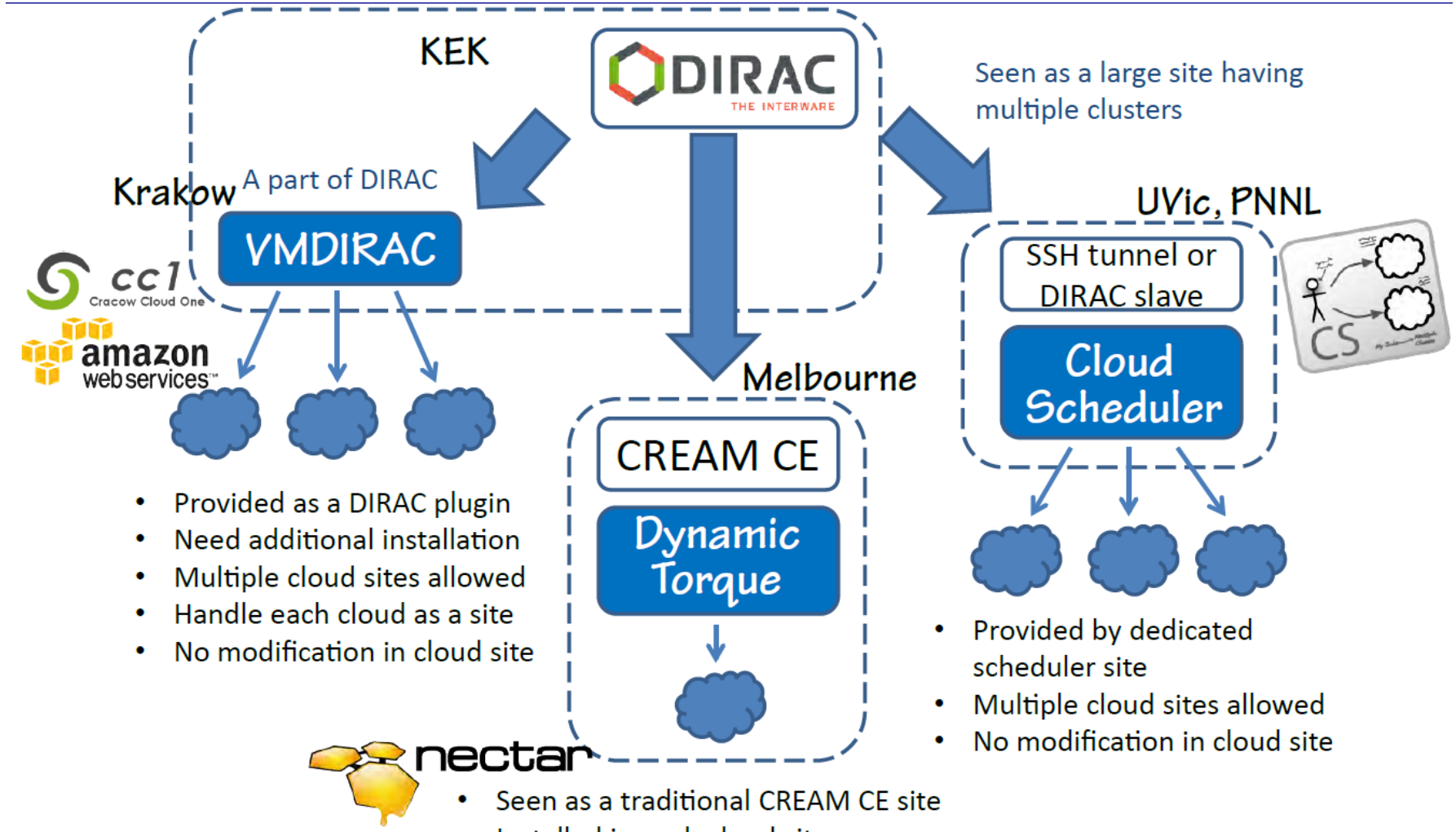
Interoperability in heterogeneous computings

PRODUCTION SYSTEM

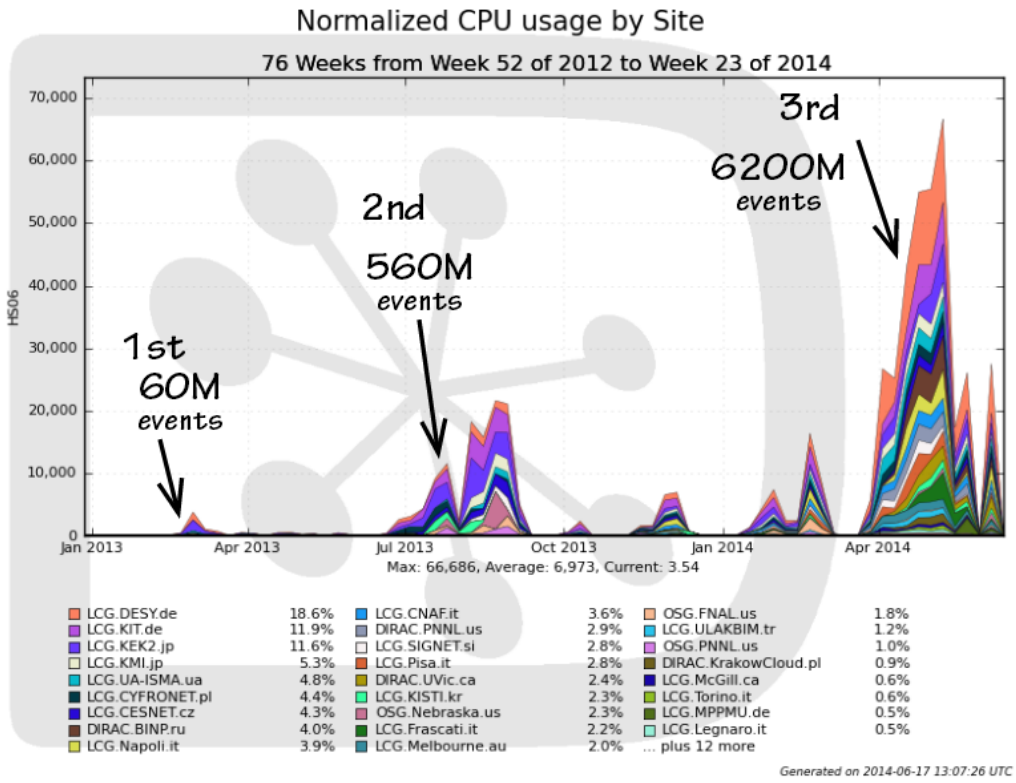


CernVM File system cvmfs is used for software installation for most of sites

PRODUCTION SYSTEM



MC Campaign



15 countries/regions

27 sites (+ 2 non-Belle II sites)

HEPHY (Vienna) and MPPMU (Munich) joined recently

GRID, Cloud, local cluster is available

First official release of MC samples

BB generic decay/continuum

tau pair

(corresponding to 100fb^{-1} w/ and w/o BG)

Trans-pacific / trans-atlantic network data tranfer challenge

About 6 billion MC samples within roughly 2 months (April and May) in this using the DIRAC framework.

INSTITUTES

Budker Institute of Nuclear Physics, Novosibirsk, Russian Federation
Novosibirsk State University, Novosibirsk, Russian Federation
CESNET, Prague, Czeck Republic
Deutsches Elektronen-Synchrotron, Hamburg, Germany
University of Hawai`i at M\={a}noa, Honolulu, USA
Institute of High Energy Physics, Austrian Academy of Sciences, Vienna, Austria
International Center for Elementary Particle Physics, Tokyo, Japan
Institute of High Energy Physics, Chinese Academy of Sciences, Beijing,China
Indian Institute of Technology Guwahati, Assam, India
INFN Laboratori Nazionali di Frascati, Roma, Italy
INFN and Federico II Universita di Napoli, Napoli, Italy
INFN and Universita di Pisa, Pisa, Italy
INFN and Universita di Torino, Torino, Italy
INFN/CNAF, Italy
INFN/Legnaro, Legnaro, Italy
National Taiwan Univ.(NTU), Taipei, Taiwan
Pacific Northwest National Laboratory, Richland, Washington, USA
Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
Faculty of Information Technology, Czech Technical University, Prague, Czech Republic
University of Victoria, Victoria, Canada
CNP, Virginia Polytechnic Institute and State University, Blacksburg, USA
Institute for Scintillation Materials, Kharkov, Ukraine
Institute for Theoretical and Experimental Physics, Moscow, Russia
High Energy Accelerator Research Organization (KEK), Tsukuba, Japan
Korea Institute of Science and Technology Information, Daejeon, Korea
Institut f\"ur Experimentelle Kernphysik, Karlsruhe Institute of Technology, Karlsruhe, Germany
H. Niewodniczanski Institute of Nuclear Physics, Krakow, Poland
J. Stefan Institute, Ljubljana, Slovenia
University of Melbourne, School of Physics, Victoria, Australia
Middle East Technical University, Ankara, Turkey
McGill University, Department of Physics, Montreal, Canada
LRZ/RZG,Munich, Germany
Kobayashi-Maskawa Institute for the Origin of Particles and the Universe, Nagoya University, Nagoya, Japan

ACKNOWLEDGEMENT

We are grateful for the support and the provision of computing resources by

CoEPP in Australia,
HEPHY in Austria,
McGill HPC in Canada,
CESNET in Czech Republich,
DESY, GridKa, LRZ/RZG in Germany,
INFN/CNAF in Italy,
KEK-CRC, KMI in Japan,
KISTI GSDC in Korea,
Cyfronet, CC1 in Poland,
NUSC, SSCC in Russia,
SiGNET in Slovenia,
ULAKBIM in Turkey,
UA-ISMA in Ukraine,
and
OSG, PNNL in USA.

We acknowledge the service provided by

CANARIE, Dante, ESnet, GARR, GEANT, and NII.

We thank the DIRAC and AMGA teams for their assistance
and CERN for the operation of a CVMFS server for Belle II.

REFERNCES

[1] “BelleII software and computing” Takanori Hara - LCWS13 – Tokyo
http://agenda.linearcollider.org/getFile.py/access?contribId=74&sessionId=35&resId=0&materialId=slides&confId=6000

[2] “Computing for Belle-2 “Takanori Hara at ALICE Physics Analysis and Tier-1/2 Workshop 3-8 March 2014
http://indico.cern.ch/event/274974/contribution/73/material/slides/0.pdf

[3] “Belle II Networking Overview” MALACHI SCHRAM et al the 18th B2GM (June 18-21 (Wed-Sat), 2014, KEK, Japan.
http://kds.kek.jp/getFile.py/access?contribId=35&sessionId=83&resId=0&materialId=slides&confId=15329

[4] “MC CAMPAIGN SUMMARY Hideki Miyake“ at 10th Belle II Computing/Software Workshop
http://kds.kek.jp/getFile.py/access?contribId=39&sessionId=12&resId=5&materialId=slides&confId=14705

THANK YOU

BACKUP

CPU time for MC production

Belle II

Name	Calls	Memory(MB)	Time(s)	Time(ms)/Call
EventInfoSetter	1001	0	0.028	0.028
EvtGenInput	1000	0	4.289	4.289
Gearbox	1000	0	0.021	0.021
Geometry	1000	0	0.018	0.018
FullSim	1000	10	2213.03	2213.032
PXDDigitizer	1000	0	18.988	18.988
PXDClusterizer	1000	0	2.700	2.700
SVDDigitizer	1000	0	82.309	82.309
SVDClusterizer	1000	0	9.767	9.767
CDCDigitizer	1000	0	14.618	14.618
TOPDigitizer	1000	0	0.679	0.679
ARICHDigitizer	1000	0	0.085	0.085
ECLDigitizer	1000	1	5.414	5.414
BKLMDigitizer	1000	0	96.934	96.934
EKLMDigitizer	1000	0	47.138	47.138
Level3	1000	0	1.683	1.683
Trasan	1000	0	616.060	616.060
VXDTF	1000	15	20.836	20.836
MCTrackCandCombine	1000	0	5.756	5.756
GenFitter	1000	80	696.626	696.626
DedxPID	1000	0	37.631	37.631
Ext	1000	4	95.747	95.747
TOPReconstructor	1000	0	1676.62	1676.622
ARICHReconstructor	1000	0	10.170	10.170
ECLReconstructor	1000	0	6.781	6.781
ECLGammaReconstru	1000	0	0.361	0.361
ECLPi0Reconstructo	1000	0	14.947	14.947
ECLTrackShowerMatc	1000	0	2.125	2.125
ECLElectronId	1000	0	0.260	0.260
ECLMCMatching	1000	0	5.112	5.112
EKLMReconstructor	1000	0	0.126	0.126
EKLMKOLReconstruc	1000	0	0.088	0.088
BKLMReconstructor	1000	0	0.368	0.368
MuId	1000	0	30.507	30.507
MdstPID	1000	0	0.560	0.560
RootOutput	1000	5	1.510	1.510
Total	1001	119	5732.00	5726.273

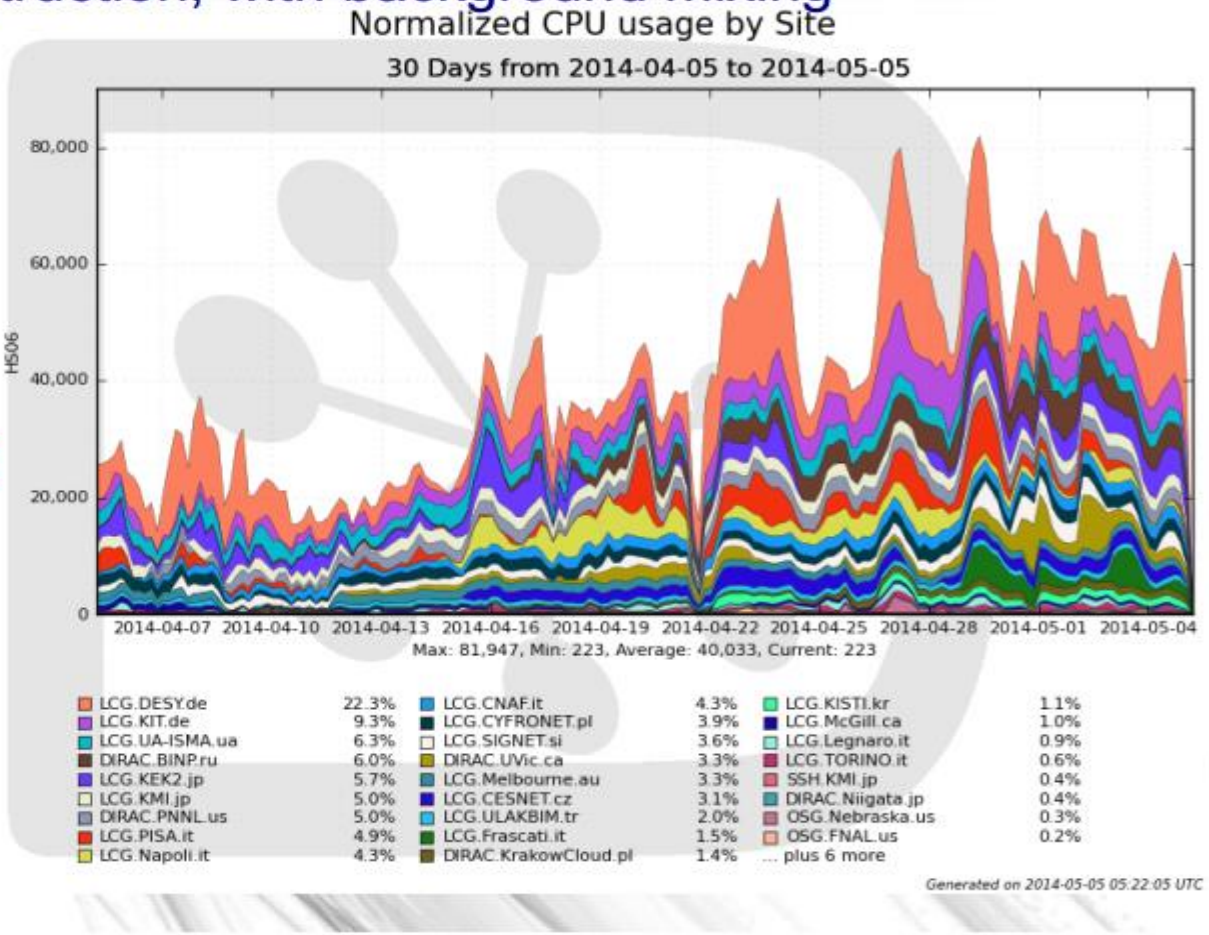
Breakdown of CPU usage
of typical hadronic MC events

- Event generator : 0.1 %
- Geometry setting : ~0 %
- Det. simulation : 39 %
- Digitizer : ~5 %
- Reconstruction : 56 %

MC Campaign

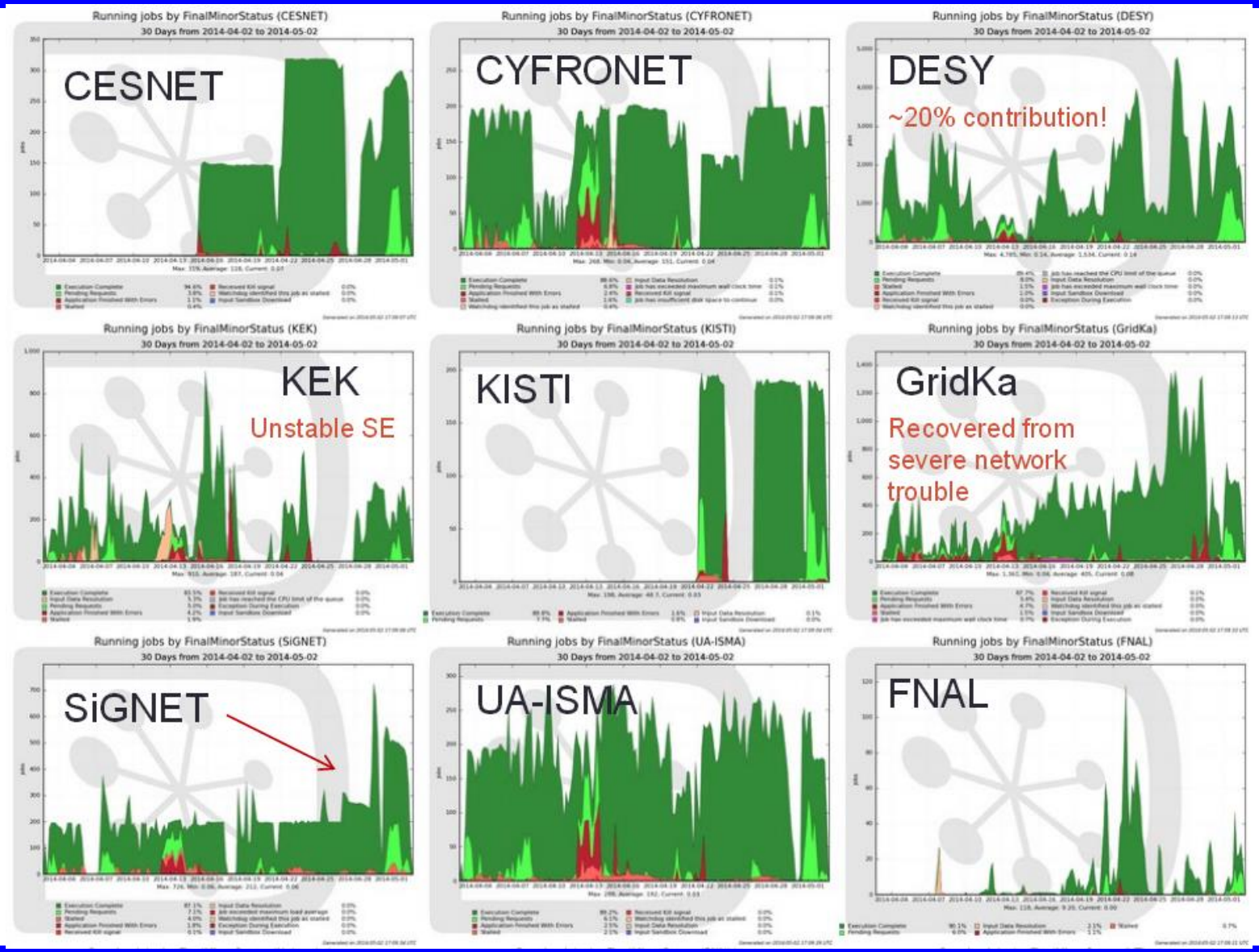
- 3rd MC Campaign: ~ April 1 – May 15, 2014
- Simulation and reconstruction, with background mixing
  → mdst data
- 2x previous CPU#:
  11k concurrent jobs;
  > 80 kHS max
- ~30 sites contributing
- 4.2G events produced

→ Very successful;
also updated analysis
and grid software
→ To obtain useful data
for physics studies **new extensive MC production started this week**

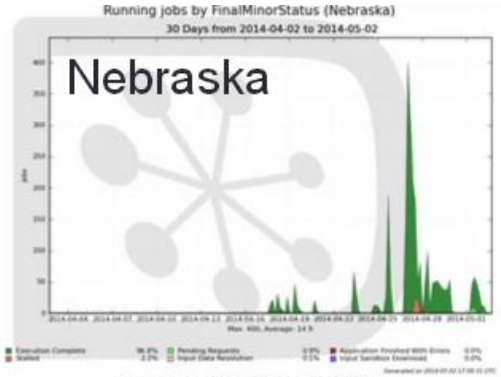


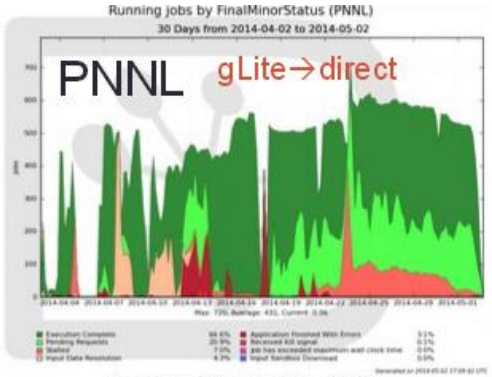
SITE STATUS DURING THE MC Campaign 3.0

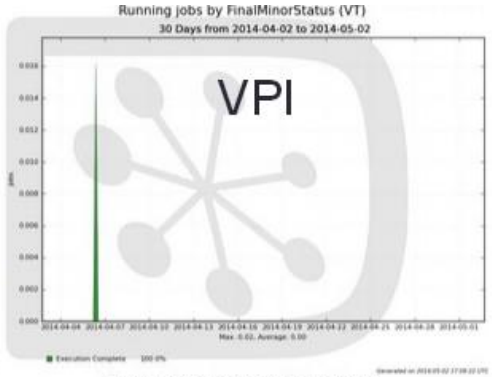
Belle II

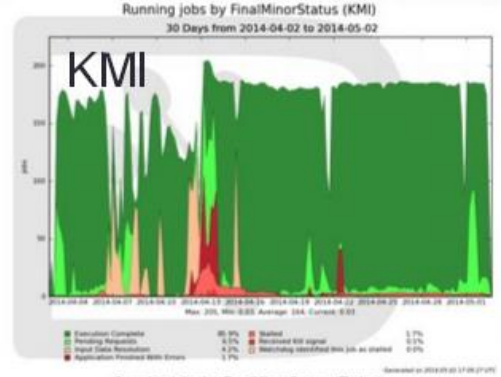


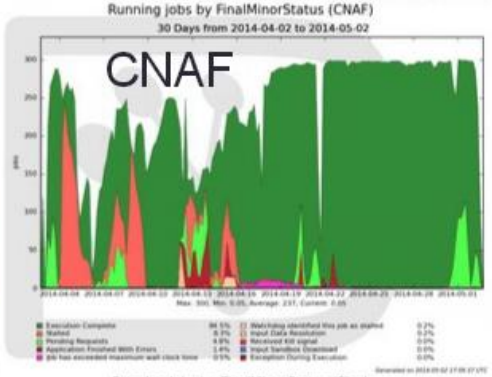
SITE STATUS DURING THE MC Campaign 3.0

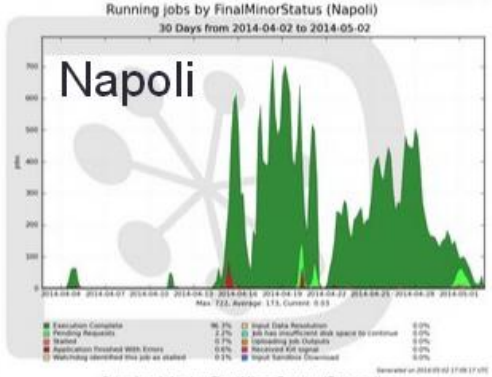


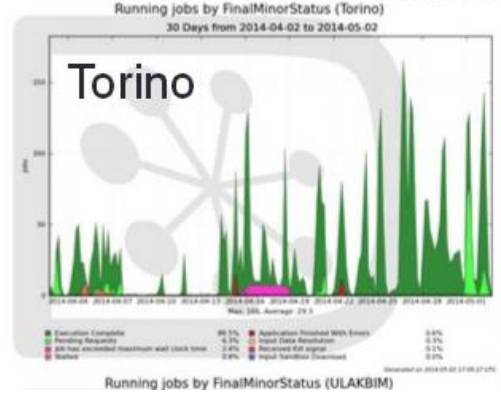


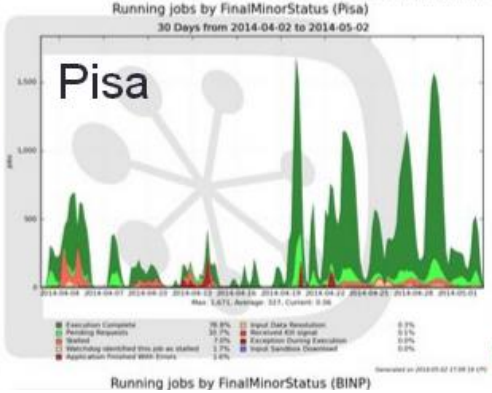


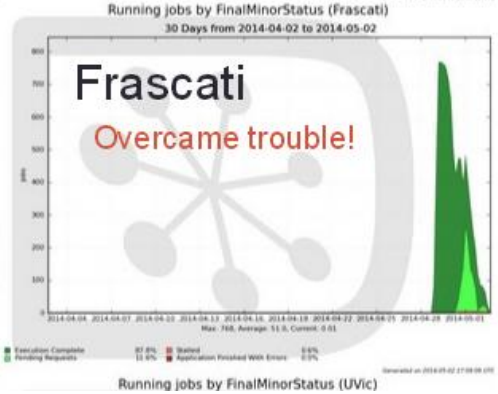












Running jobs by FinalMinorStatus (ULAKBIM)

Running jobs by FinalMinorStatus (BINP)

Running jobs by FinalMinorStatus (UVic)

SITE STATUS DURING THE MC Campaign 3.0

Belle II