Tom Fifield
fifieldt@unimelb.edu.au

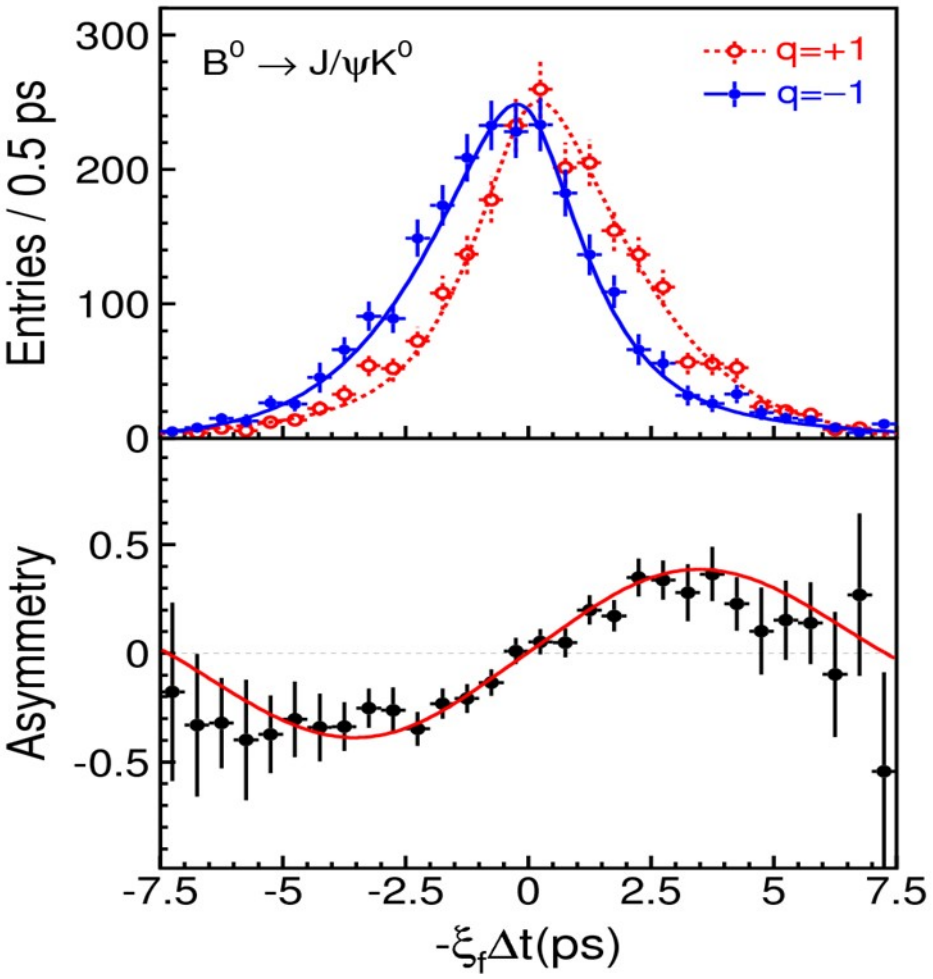
Cloud computing in the Belle II Experiment

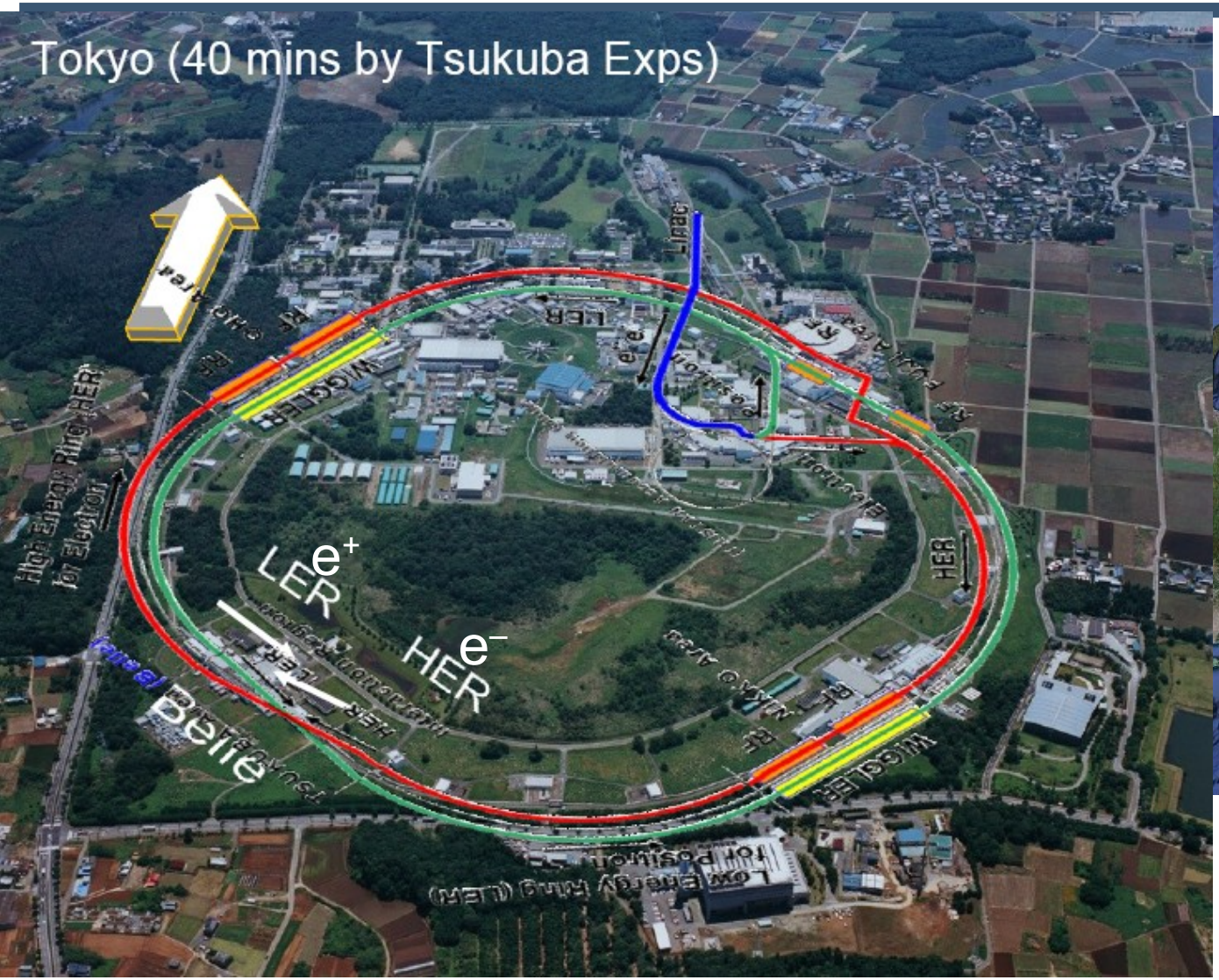
Outline

- A little bit of particle physics
- The grid
- Our use of cloud
- Tying it all together

Physics objectives of Belle and Belle II

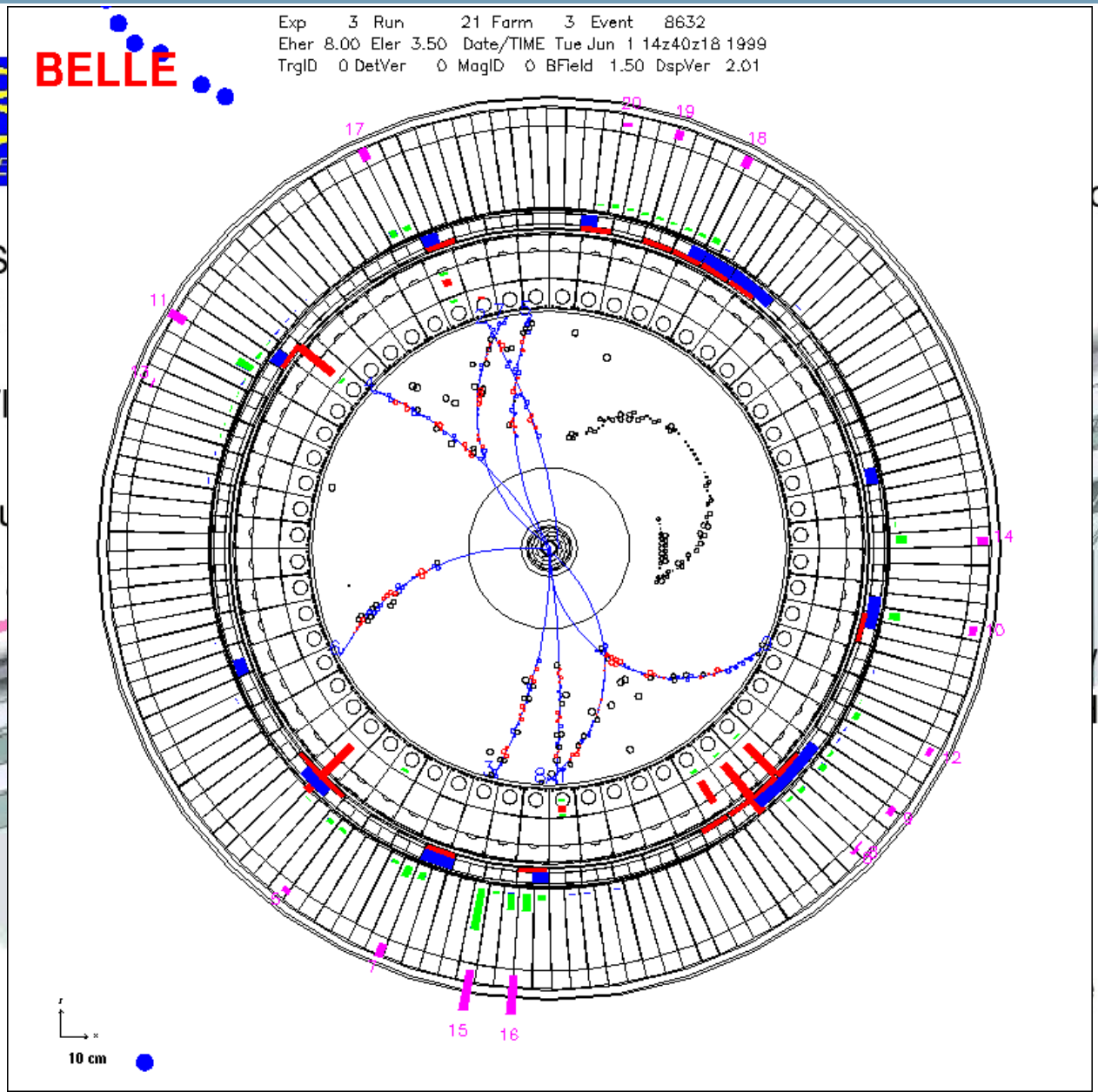
- ✔ Confirmation of KM mechanism of ~~CP~~ in the Standard Model
- ✗ CP violation in the SM by many orders of magnitude too small to generate observed baryon asymmetry in the universe
- ➢ Need sources of CP violation beyond the SM
  - ➢ Super B factory
- ➢ For more physics, see http://belle2.kek.jp/



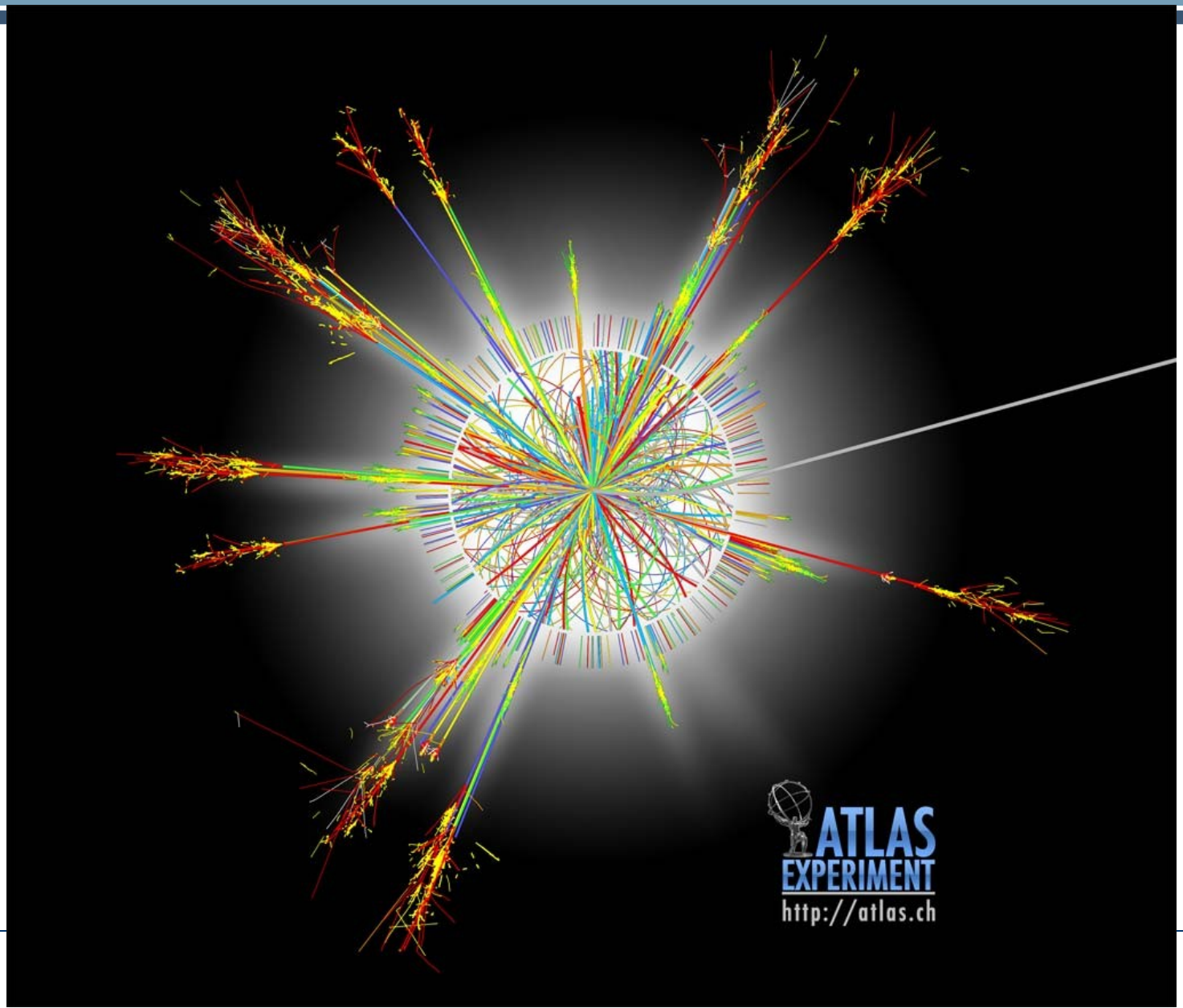

Tokyo (40 mins by Tsukuba Exps)
LER
e+
HER
e−
Belle


Tsukuba
Tokyo
Narita
Mt. Fuji

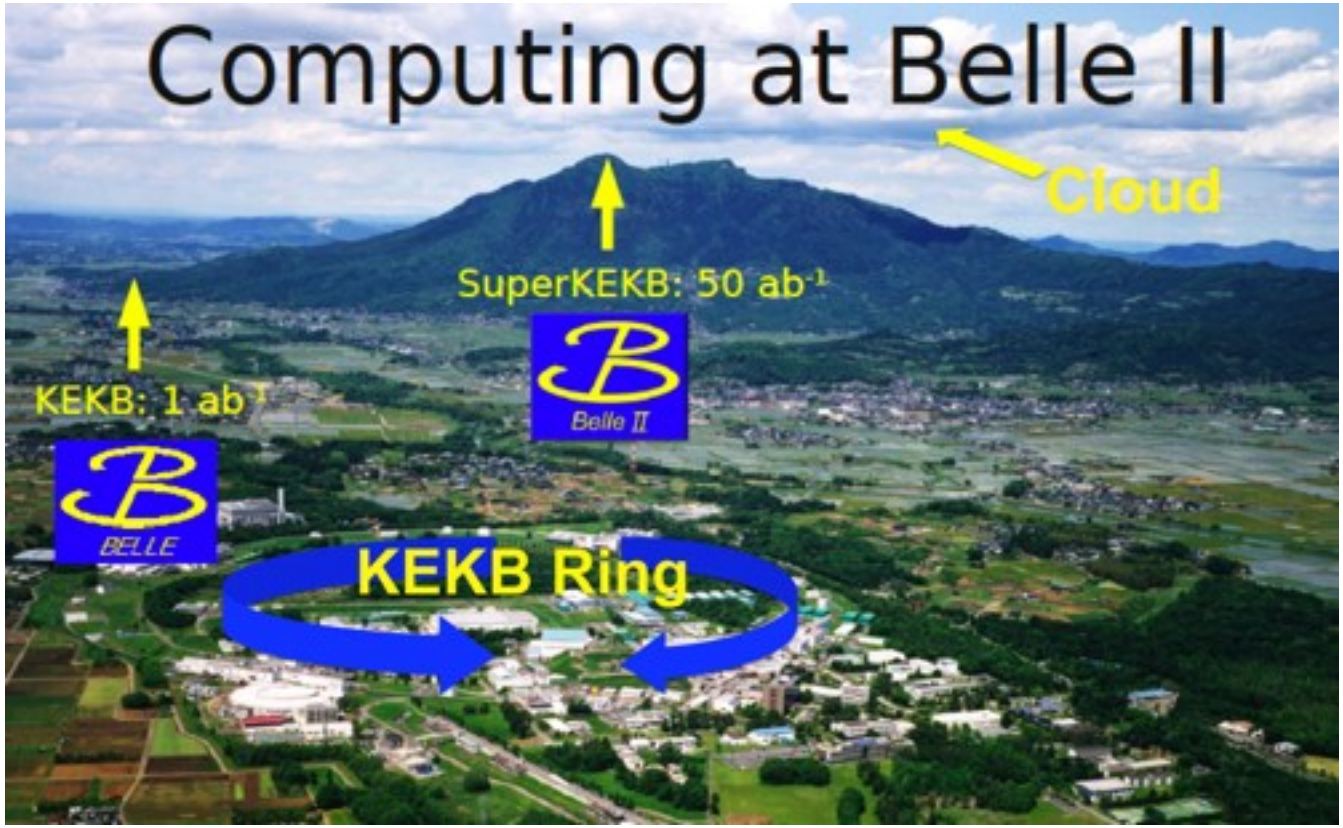
An “event”


BELLE
Exp 3 Run 21 Farm 3 Event 8632
Eher 8.00 Eler 3.50 Date/TIME Tue Jun 1 14z40z18 1999
TrgID 0 DetVer 0 MagID 0 BField 1.50 DspVer 2.01
10 cm

A slightly more complicated “event”

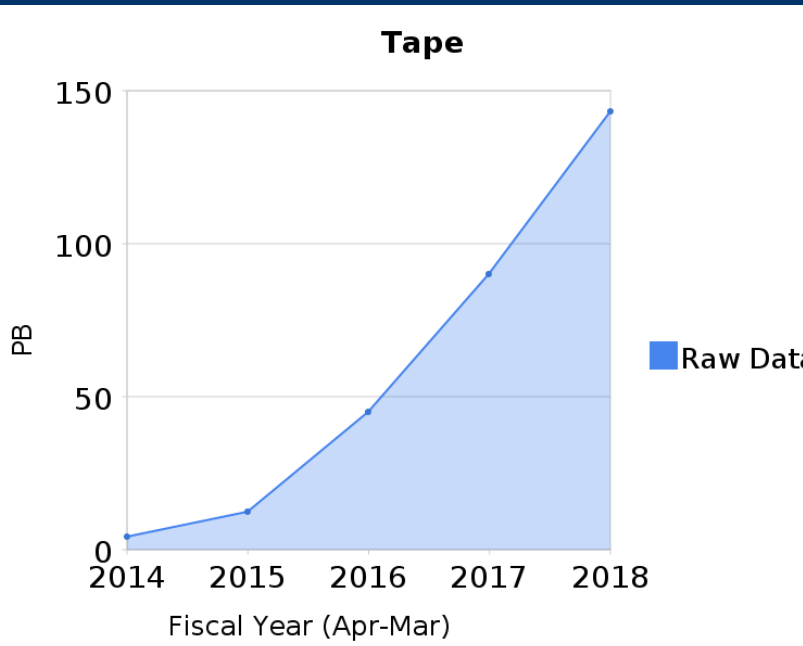


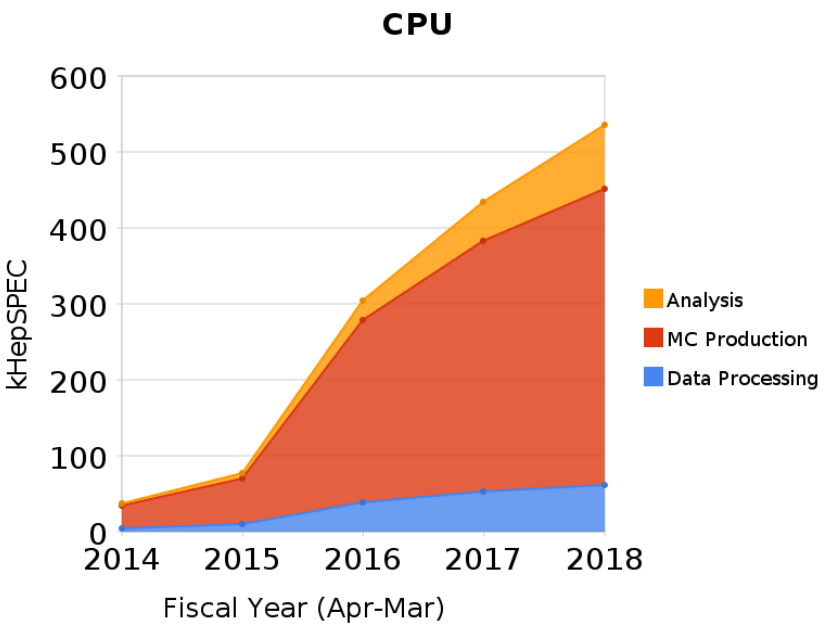
We base all our numbers on Mt Tsukuba

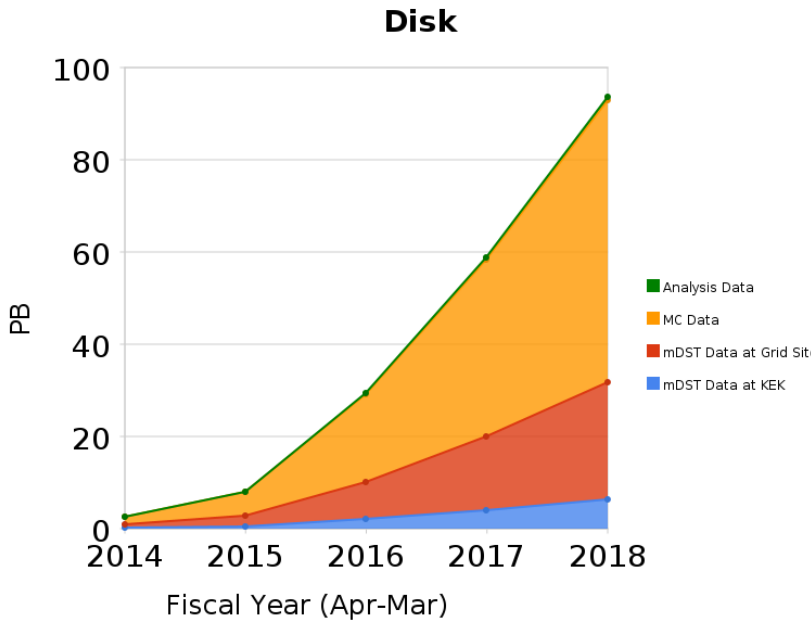


Hardware

- Preliminary estimates depend on many unknown parameters (accelerator performance, data reduction, performance of simulation/reconstruction, analysis requirements, ...)







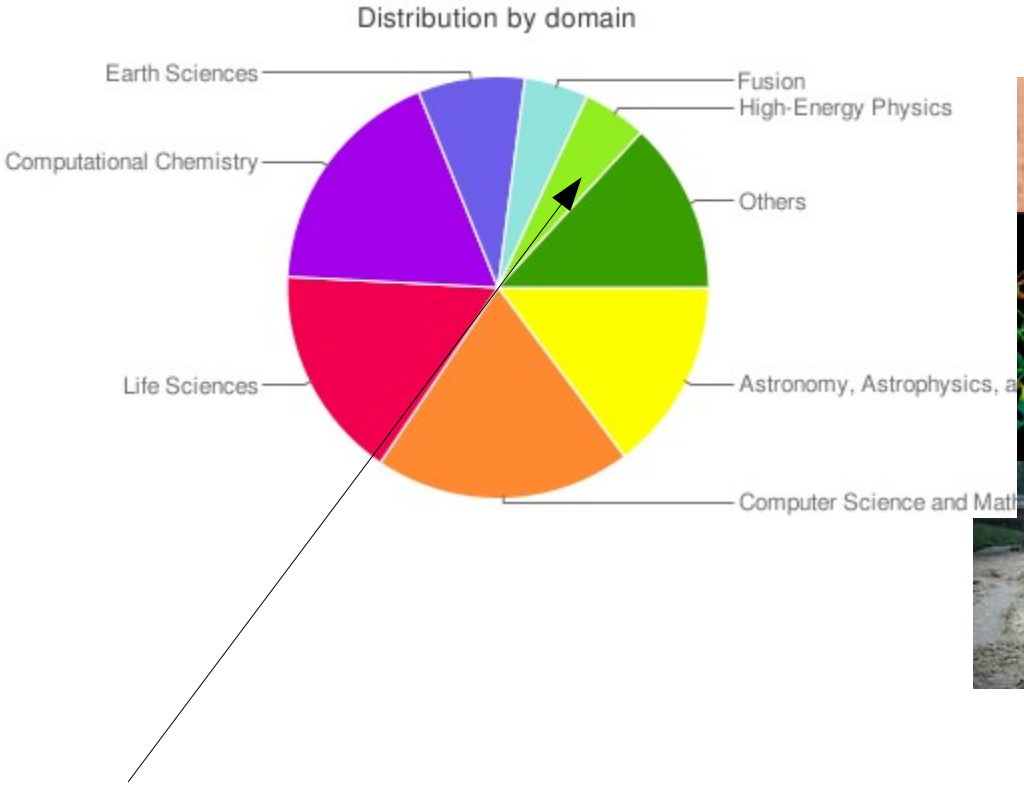
The Grid

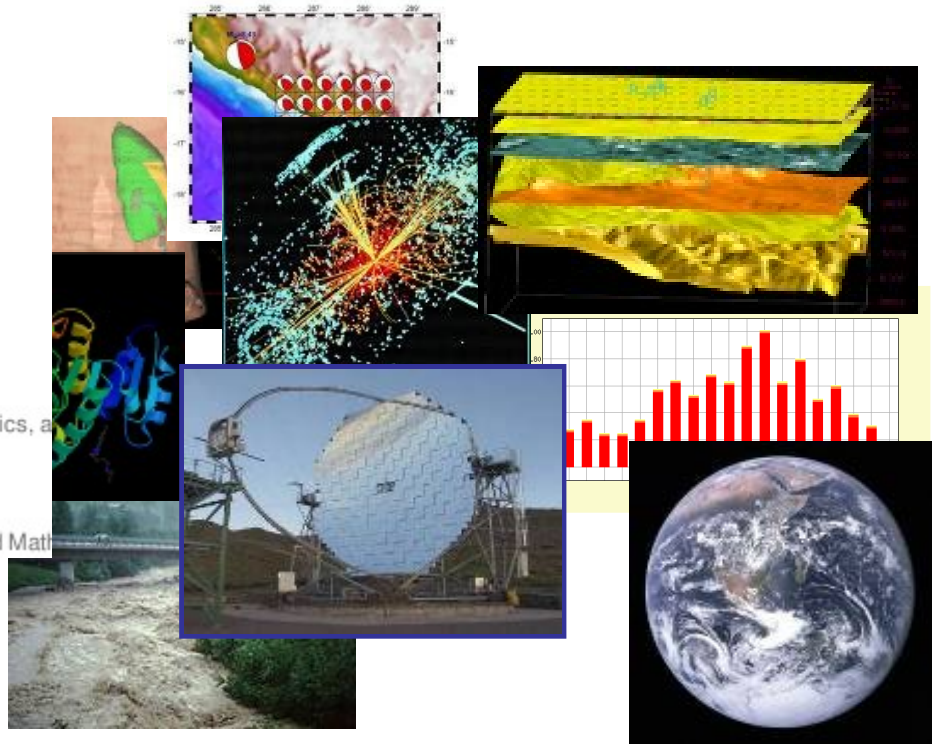
- Grid computing aims to take sparse resources and collect them in a coherent system available worldwide
- Integrate – Abstract – Manage
- Interactions with the underneath layers (batch systems, storage) must be ***transparent*** to the user
- Therefore, there is the need for a Middleware

- It's complex - I run a 3-day training course on this :)

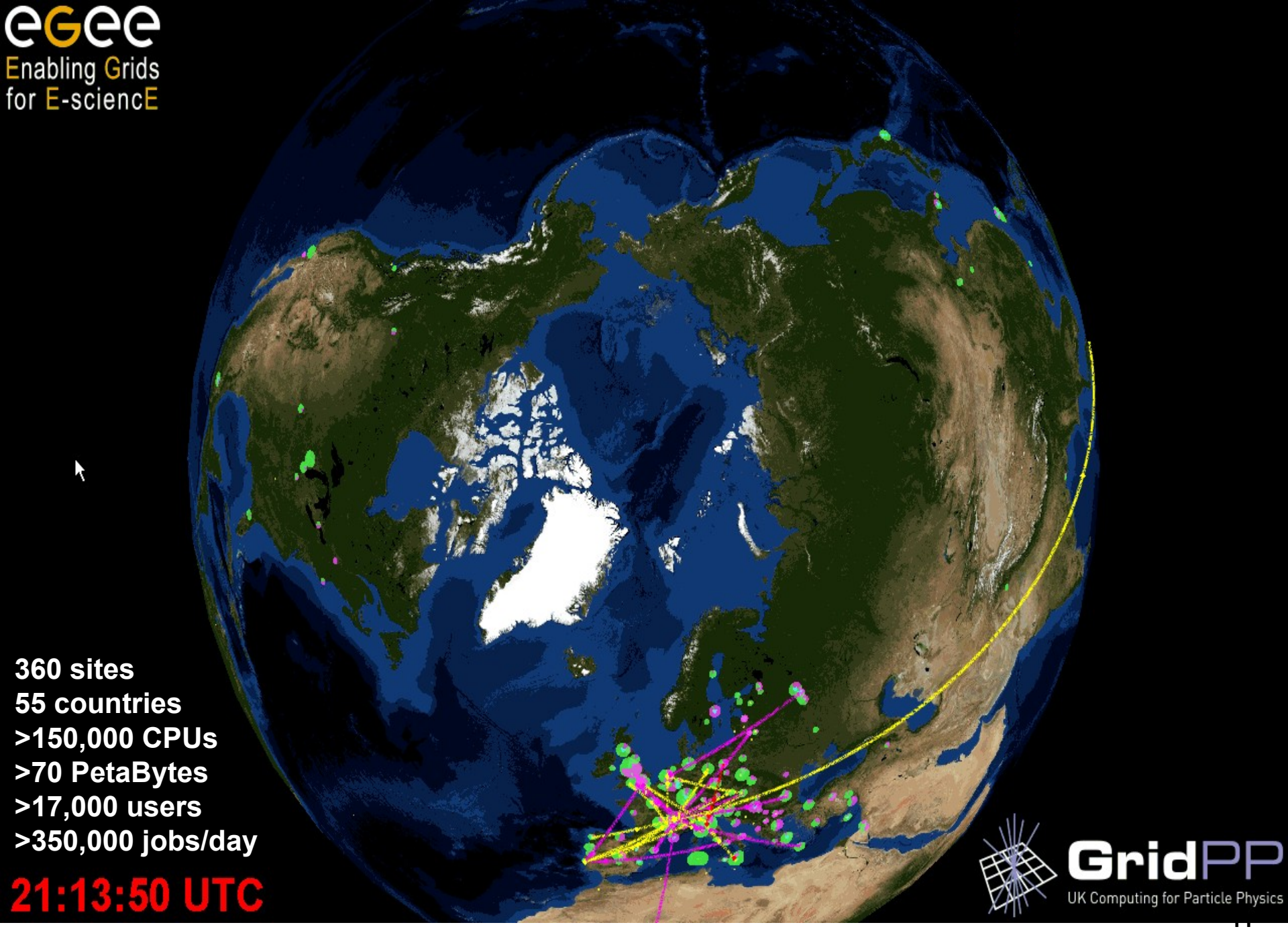
Applications

>270 Virtual Organisations from several scientific domains

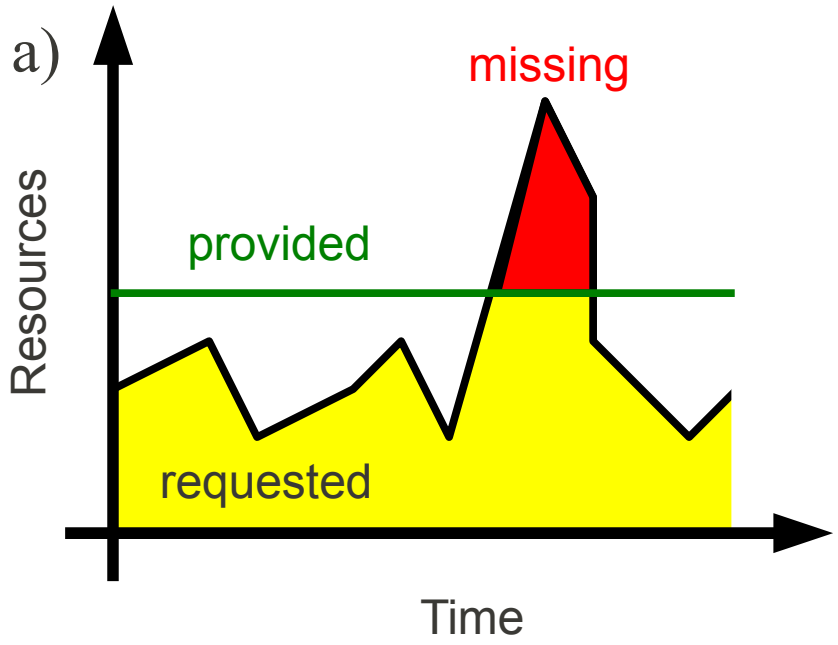


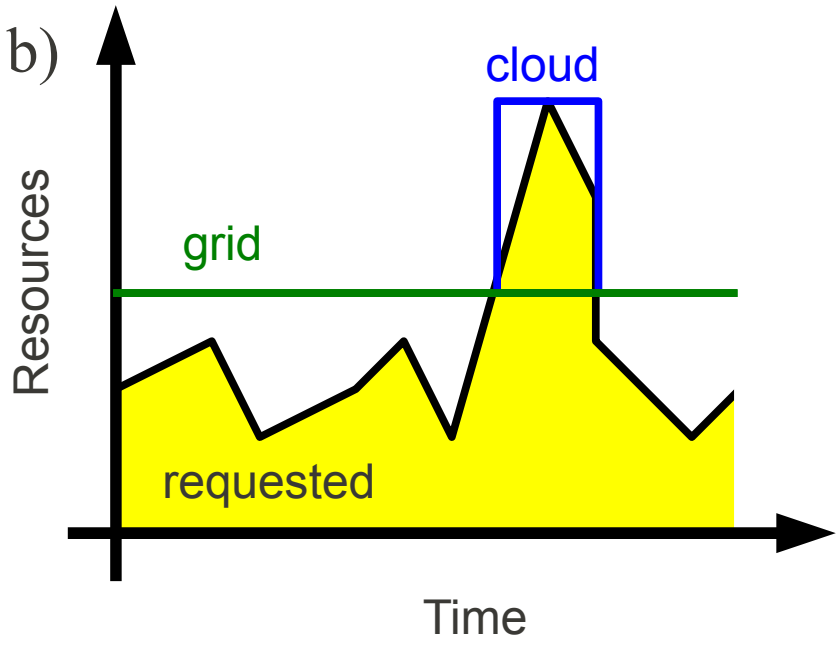


15PB+ new data/year to process


eGee
Enabling Grids
for E-sciencE
360 sites
55 countries
>150,000 CPUs
>70 PetaBytes
>17,000 users
>350,000 jobs/day
21:13:50 UTC
GridPP
UK Computing for Particle Physics

Belle II … to the clouds





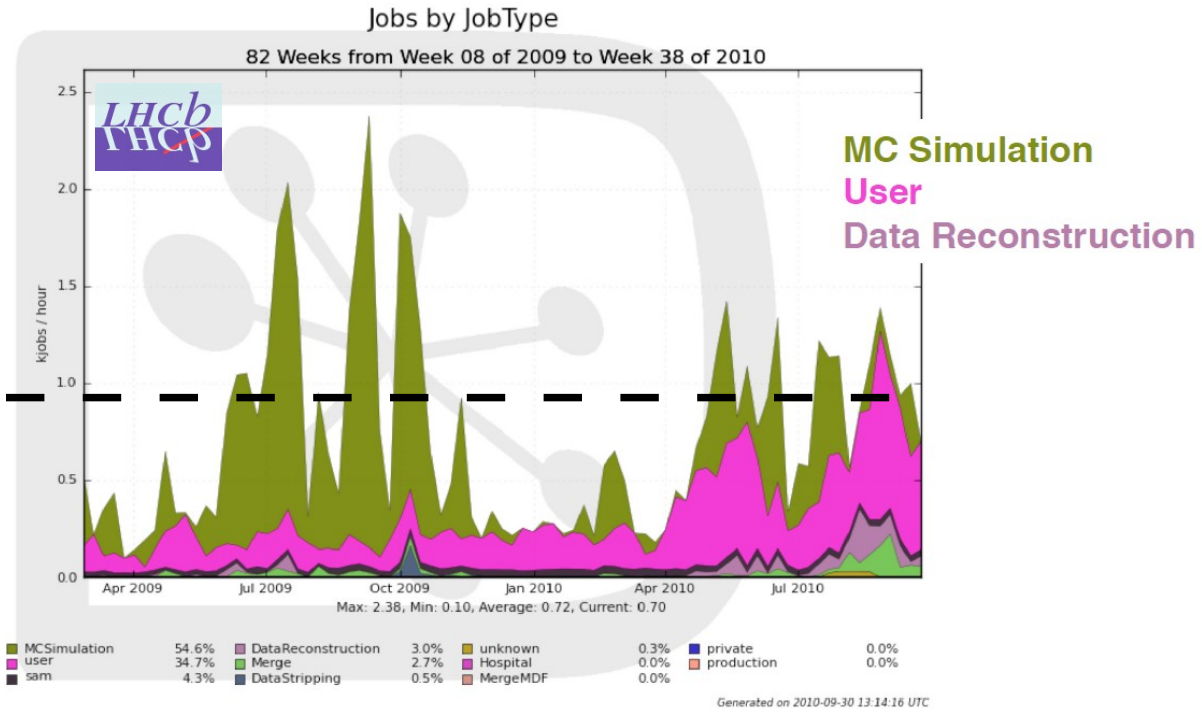
Why Cloud?

- Reduce TCO of Belle II computing...

↑data ↓$

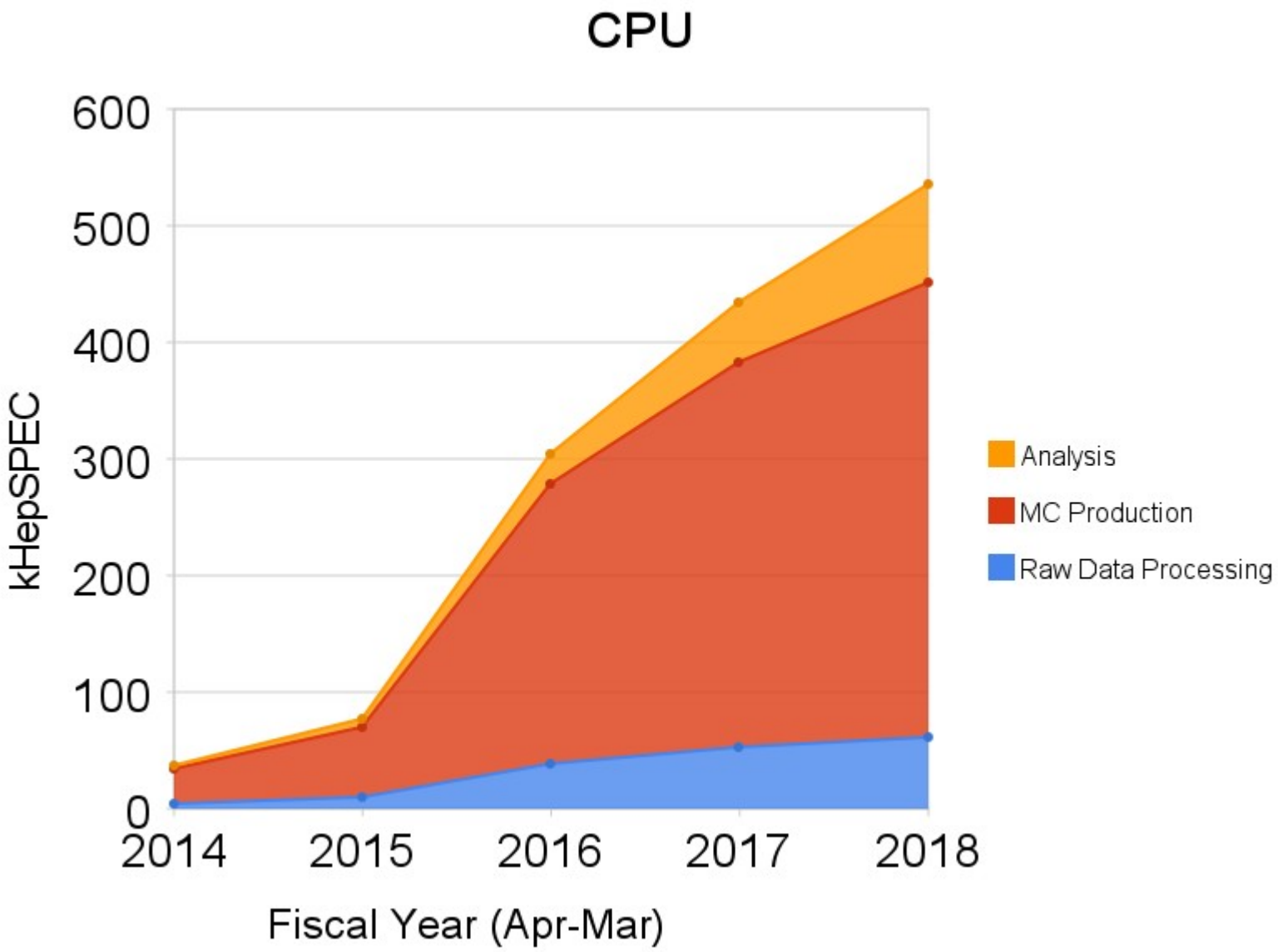
cloud?

grid



- Computing dominated by Monte Carlo production

Monte Carlo


CPU
600
500
400
300
200
100
0
kHepSPEC
2014
2015
2016
2017
2018
Fiscal Year (Apr-Mar)
Analysis
MC Production
Raw Data Processing

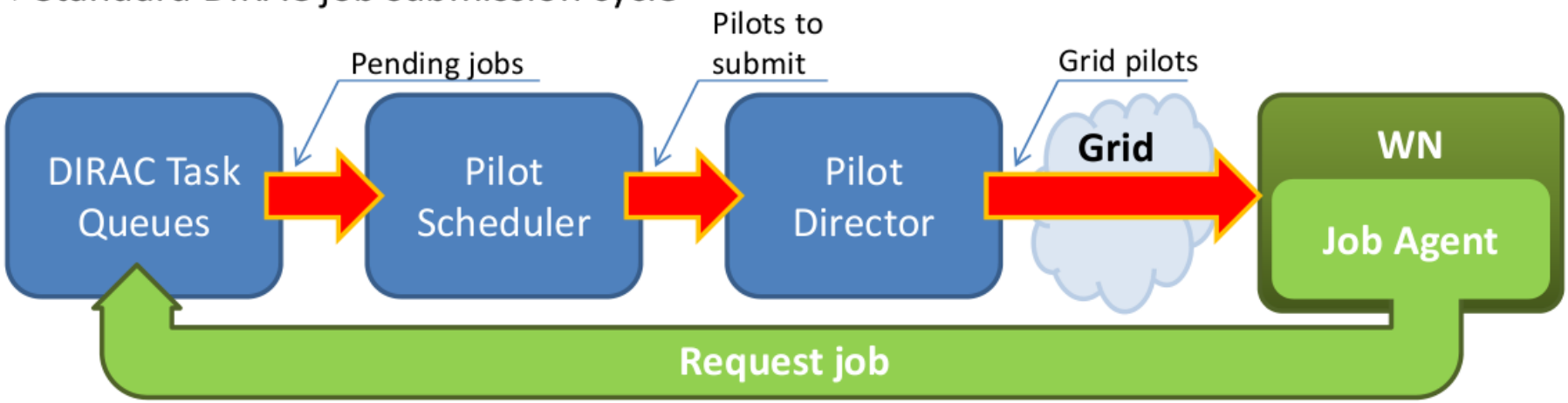
Cloud Computing / Infrastructure as a Service

- Suppose that you need several thousand computers for, say, the next hour
- Got a credit card ?
- 8-core machine, 7GB RAM, 1.7TB disk ~USD0.25/hr
  - http://aws.amazon.com/ec2/instance-types/
- Data out: UDS0.15 per GB (first 10TB)
  - first 1GB free, inbound free
  - http://aws.amazon.com/ec2/pricing/
- 99.95% uptime
  - http://aws.amazon.com/ec2-sla/
- Prepare a VM image, or use an existing one
- Click a button, you've got root access.

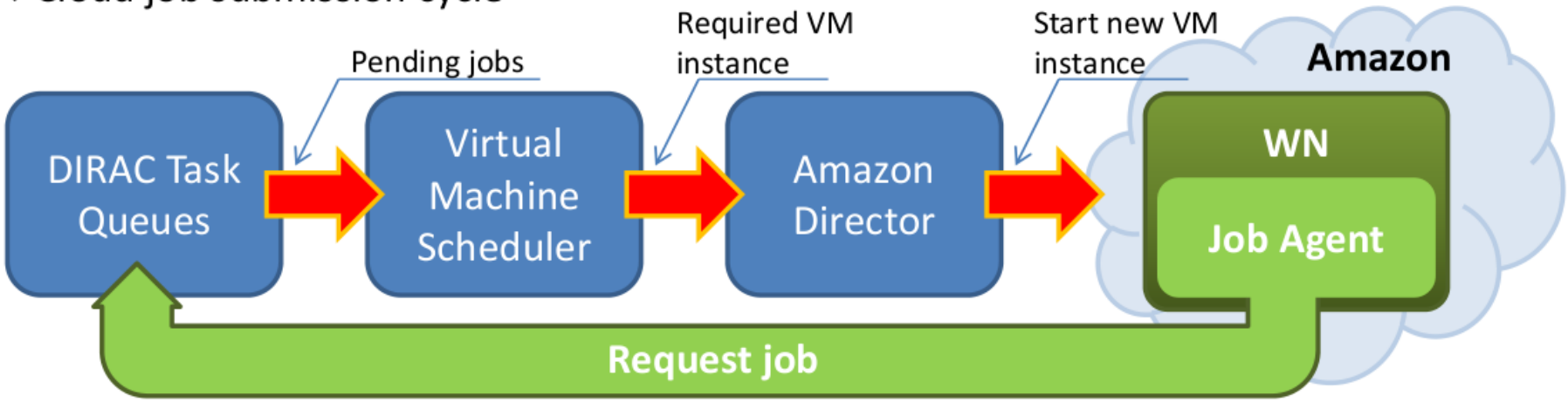
- DIRAC is a framework for distributed computing developed by the LHCb collaboration, that we use for Belle II
- DIRAC is written in python, as a number of **collaborating systems**, each providing the framework with a subset of the required functionality
- DIRAC Systems provide functionality using **Servers** and **Agents** that operate in a coordinated manner
- Virtual Organisation-Centric
  - tries to fill the gap between the resources and the community
- Code is here: http://code.google.com/p/dirac-grid/

Job Submission Concept

❖Standard DIRAC job submission cycle



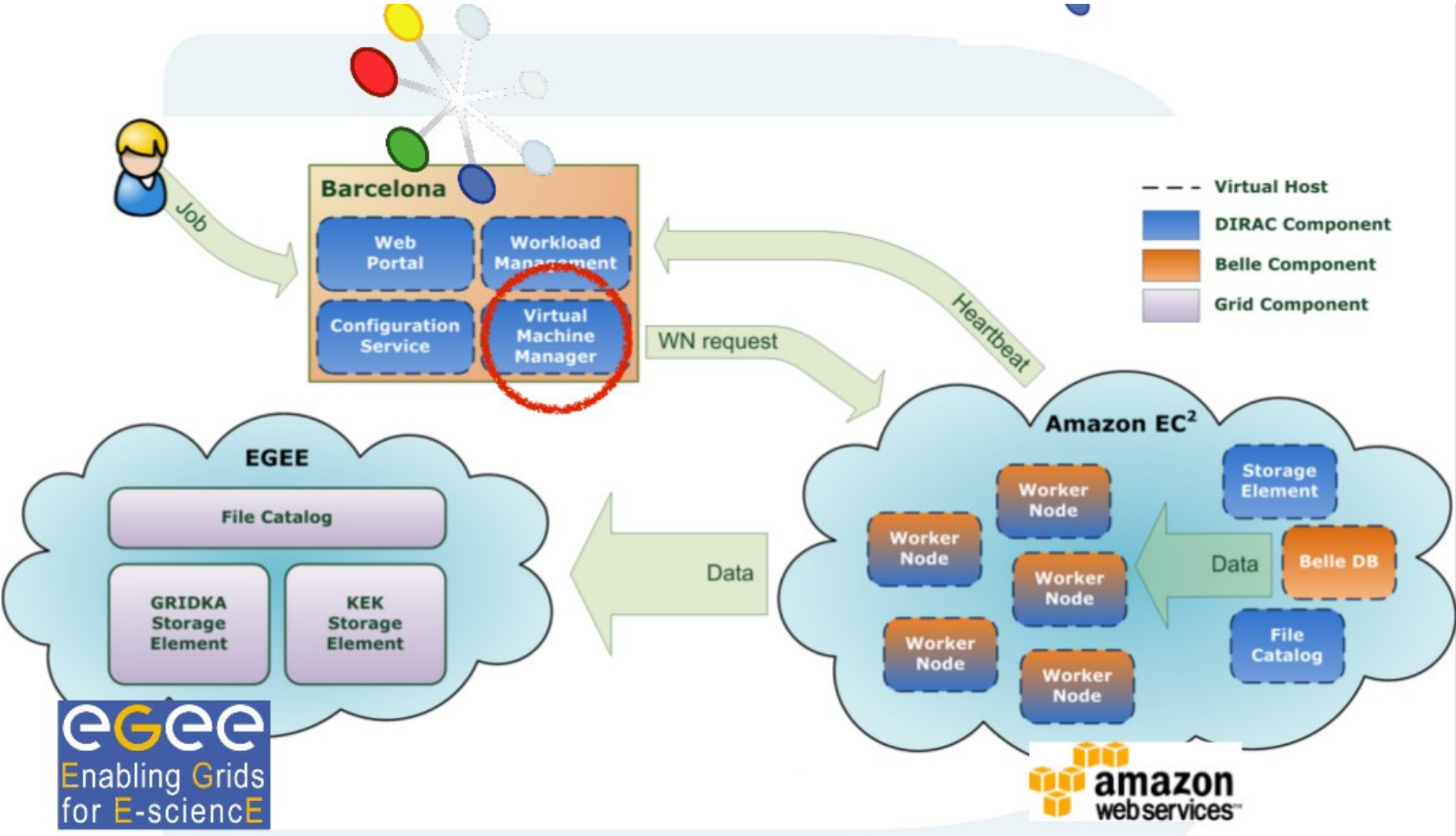
❖Cloud job submission cycle



DIRAC Cloud components

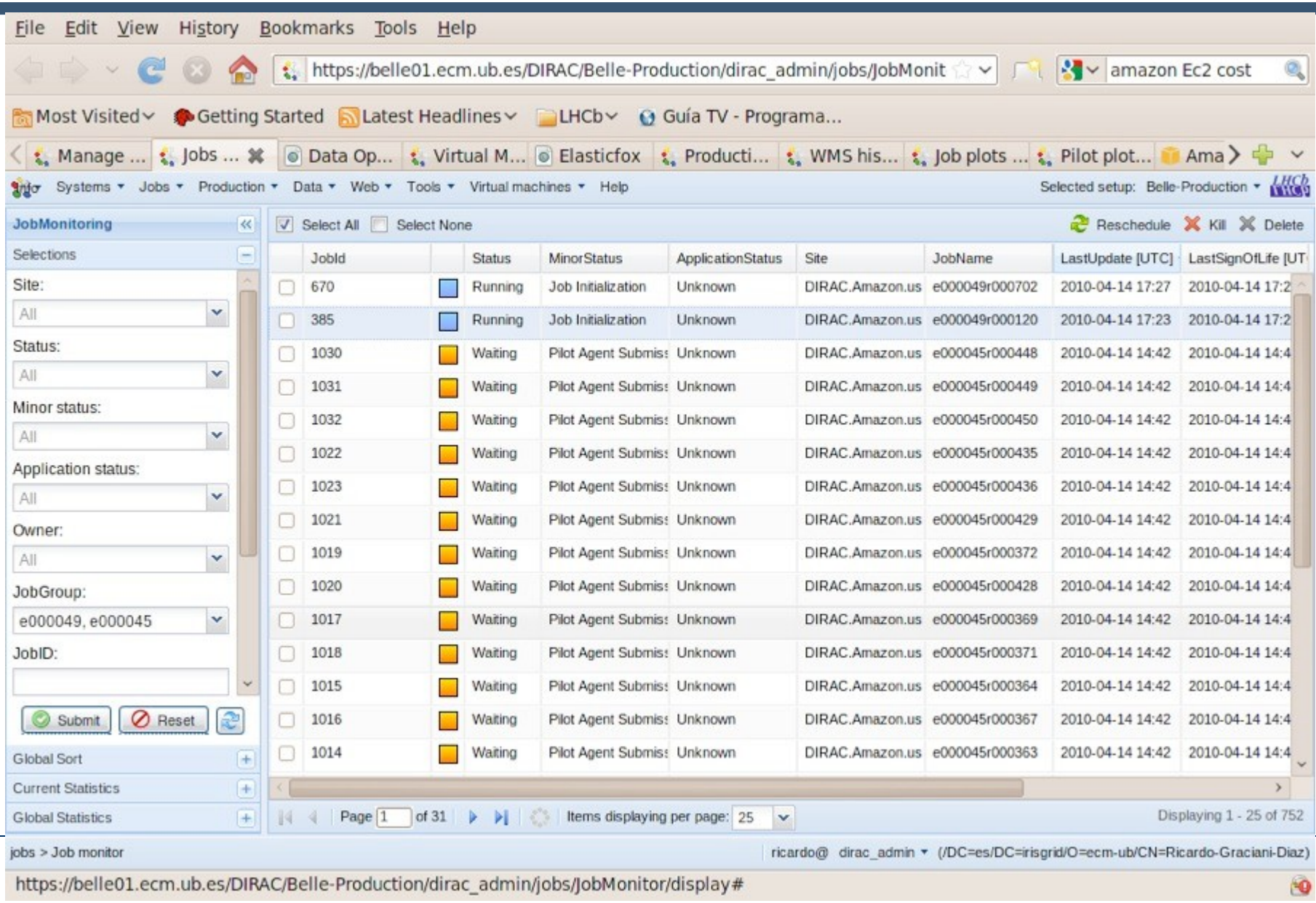
- Aim: Minimal dependence on cloud API
- Aim: Keep the proven **scalability** of DIRAC
- Replace pilot submission with virtual machine instantiation
- VirtualMachine Scheduler
  - Monitor DIRAC TaskQueues and request new VM from resource provider as appropriate
- VirtualMachine Monitor
  - On-VM module that reports activity and halts VM if no longer needed
- VirtualMachine Manager
  - Collects information about requested, running and halted VMs, and provides usage monitoring

The Solution



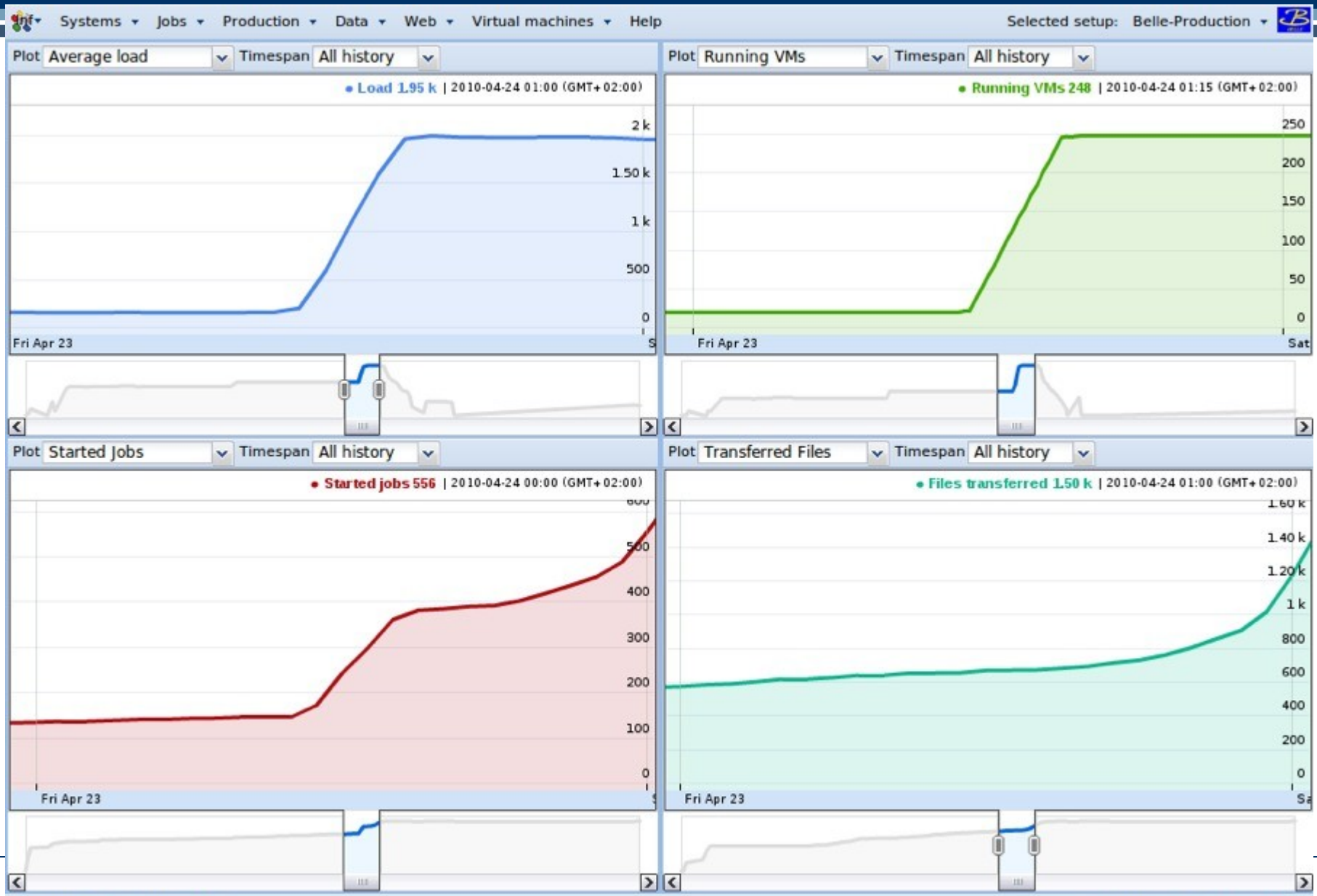
THE UNIVERSITY OF MELBOURNE

Submitting the first jobs to the cloud...

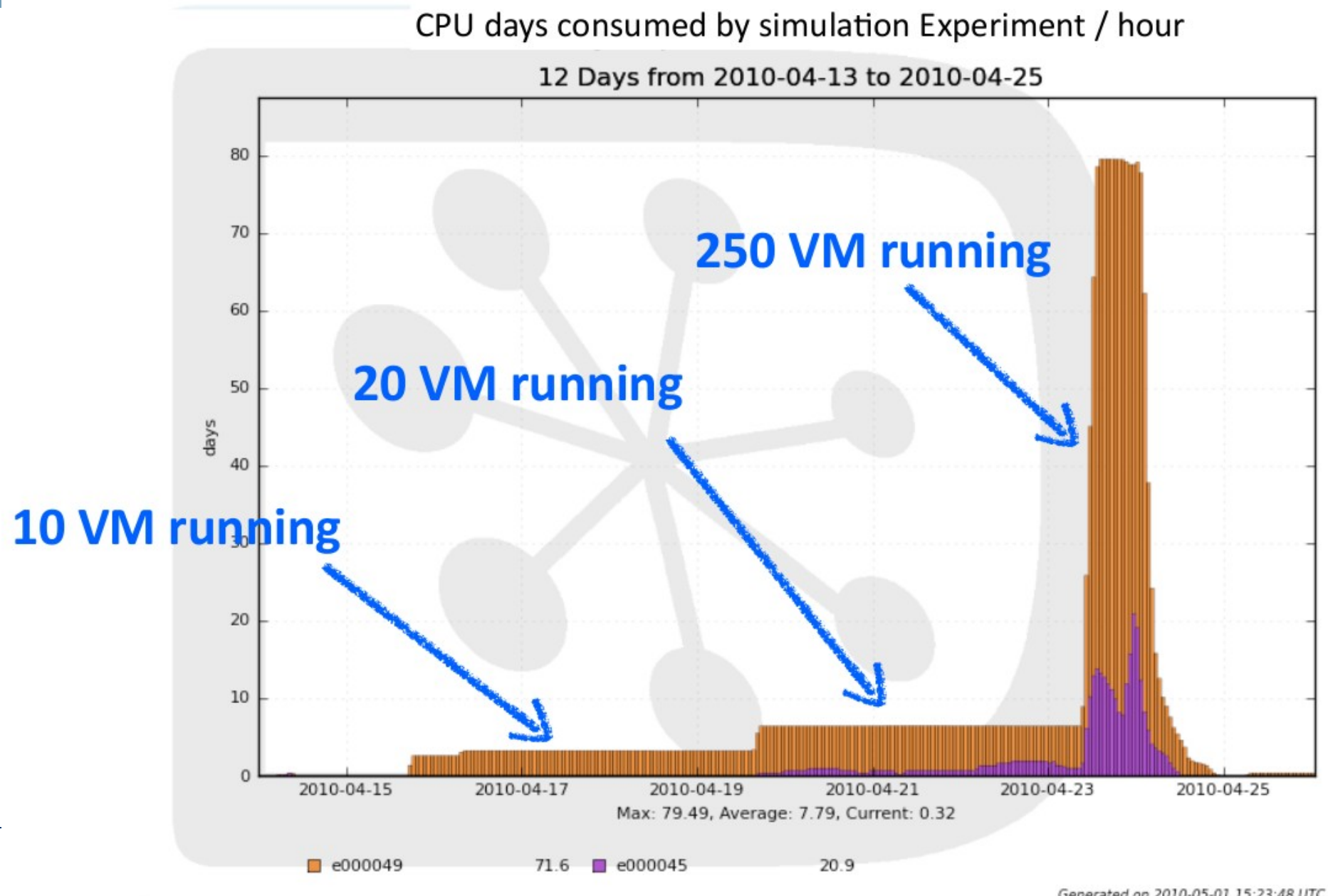


THE UNIVERSITY OF MELBOURNE

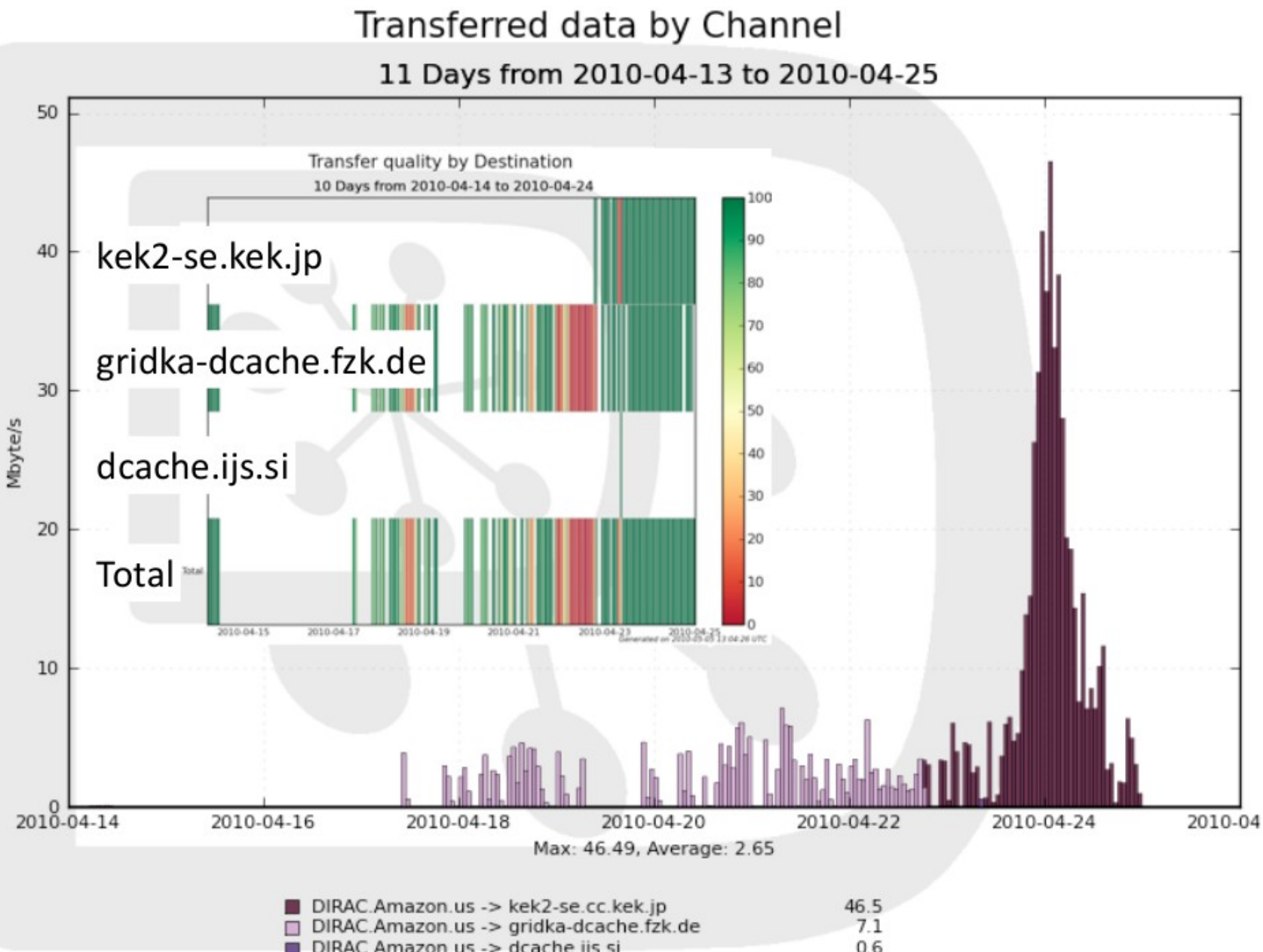
Monitoring the ramp-up



Phase One Testing

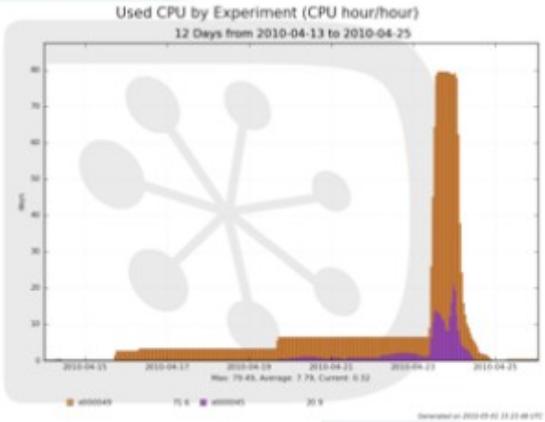


...and the data back to the grid

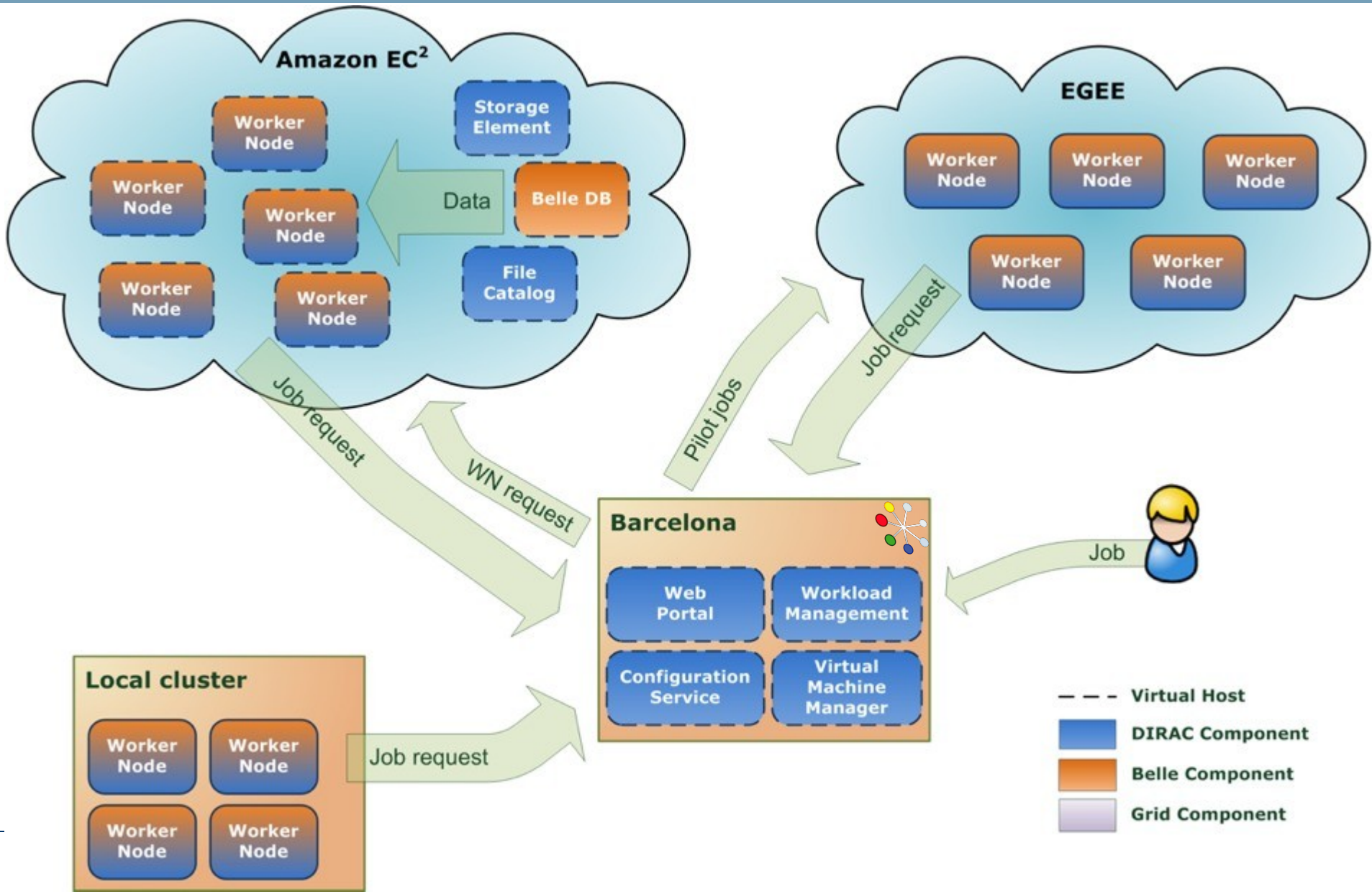


Results (I)

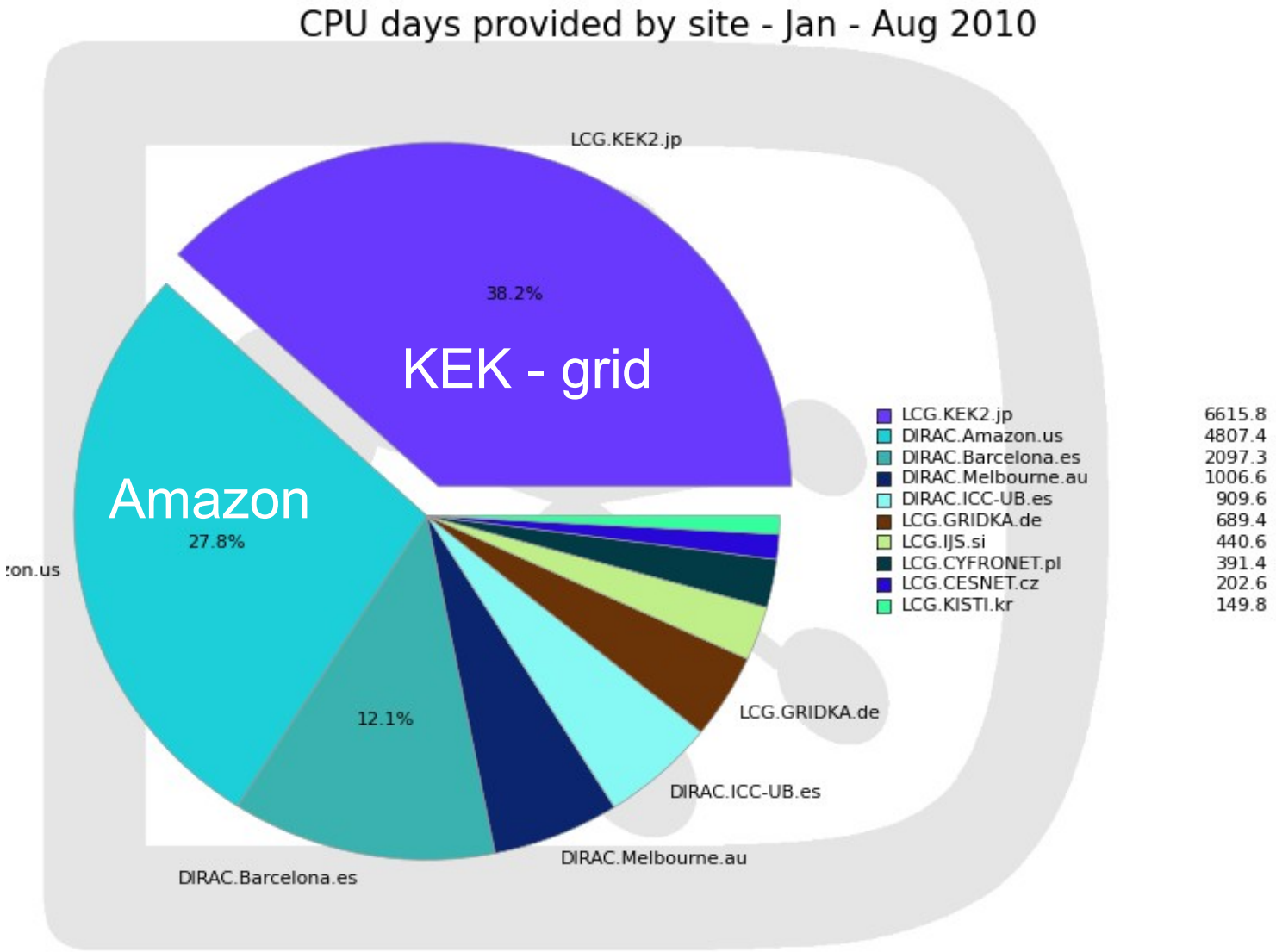
- Phase I (cloud test):
  - ☑ production ready:
    - 5% of Belle production in 10 days
    - 120 M evt (~2.7 TB)
    - 2250 CPU days used
  - ☑ proven stability and scalability:
    - 2000 CPUs peak achieved in < 4 hours
    - > 90 % efficiency in CPU usage
  - first cost estimation:
    - 0.46 USD/10k evt
  - input data pre-uploaded to Amazon SE VM.
  - few bug fixes



Interoperating – Clouds, grids, local clusters

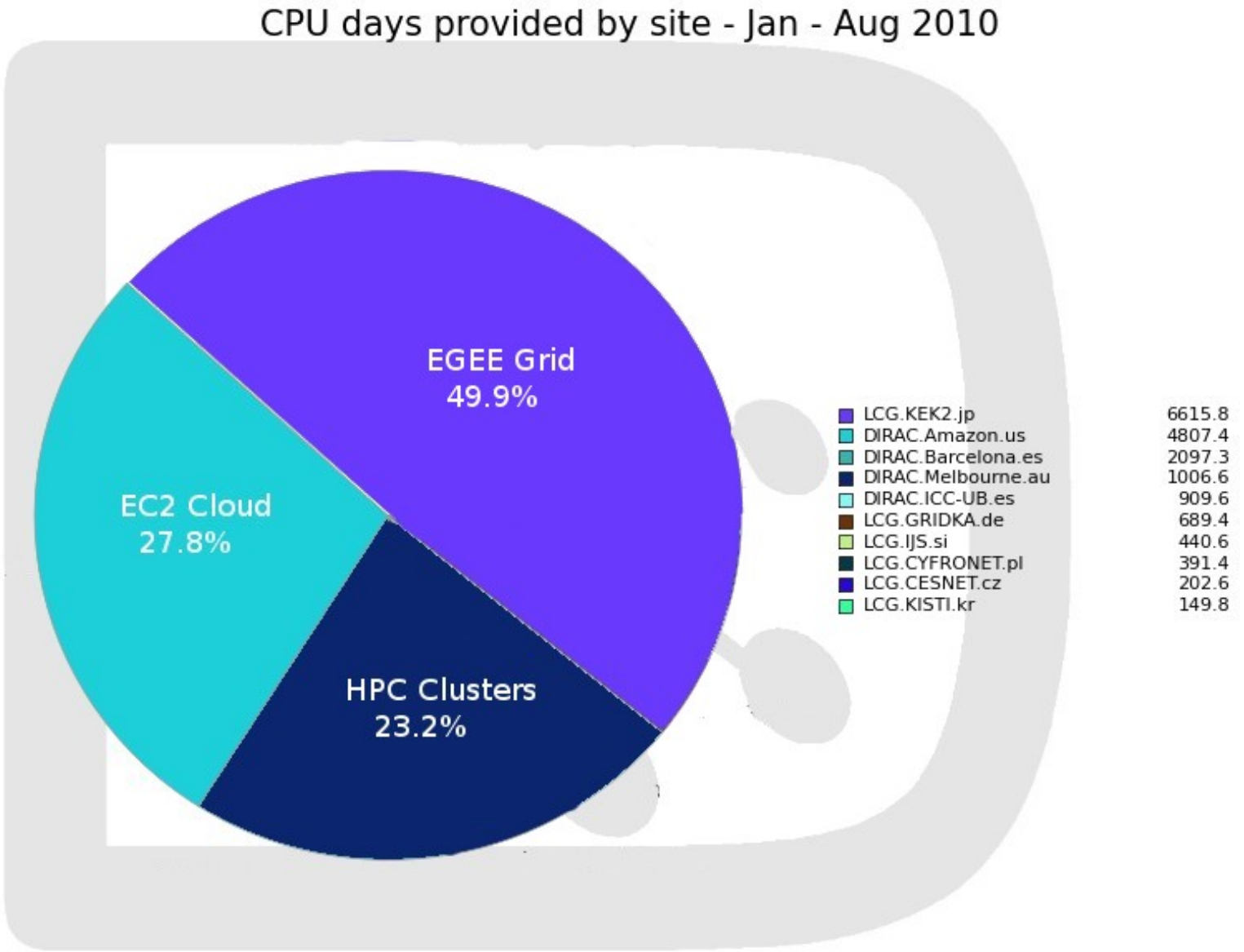


Since then...

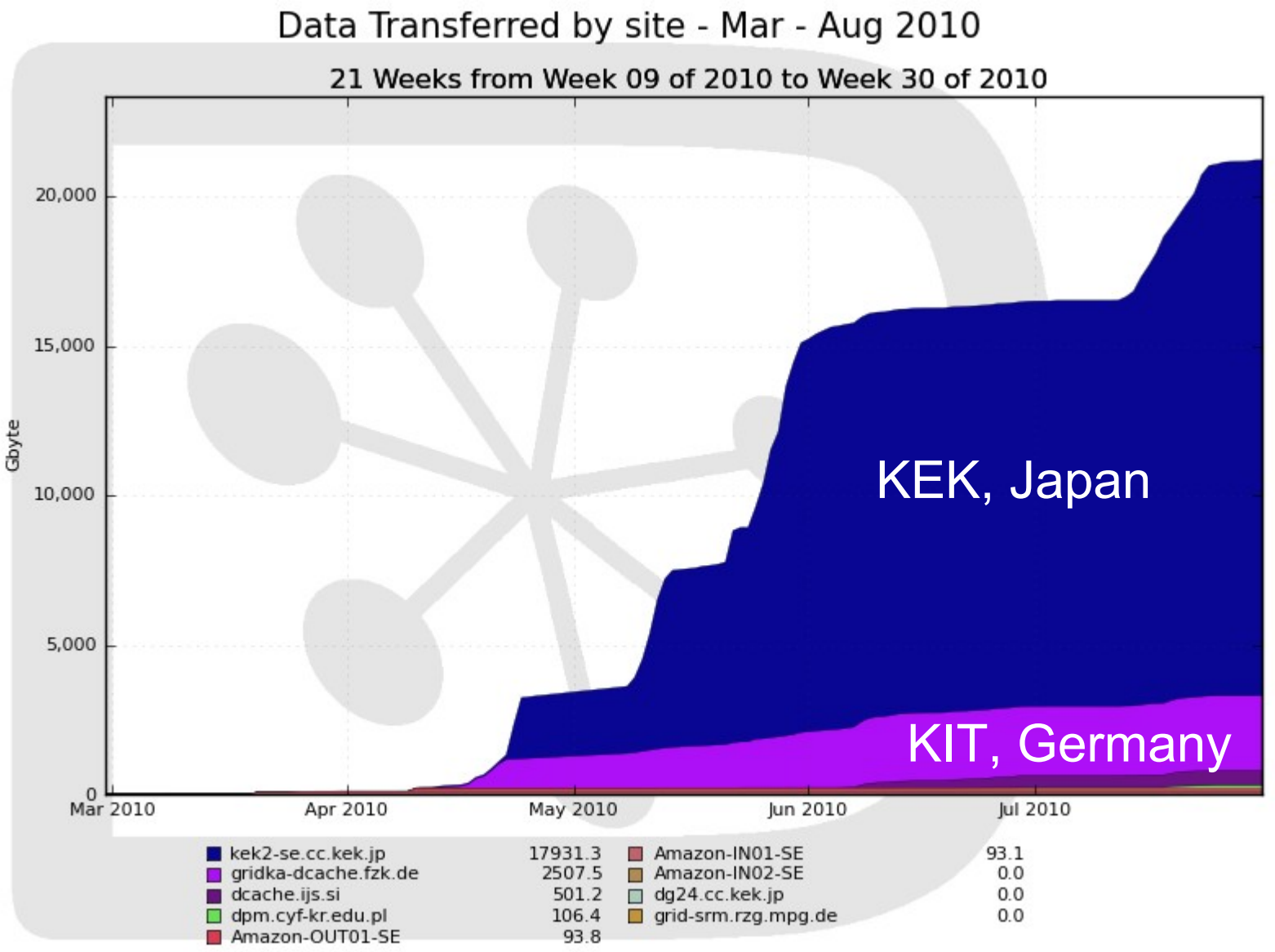


Generated on 2010-08-05 05:52:04 UTC

Total CPU distribution



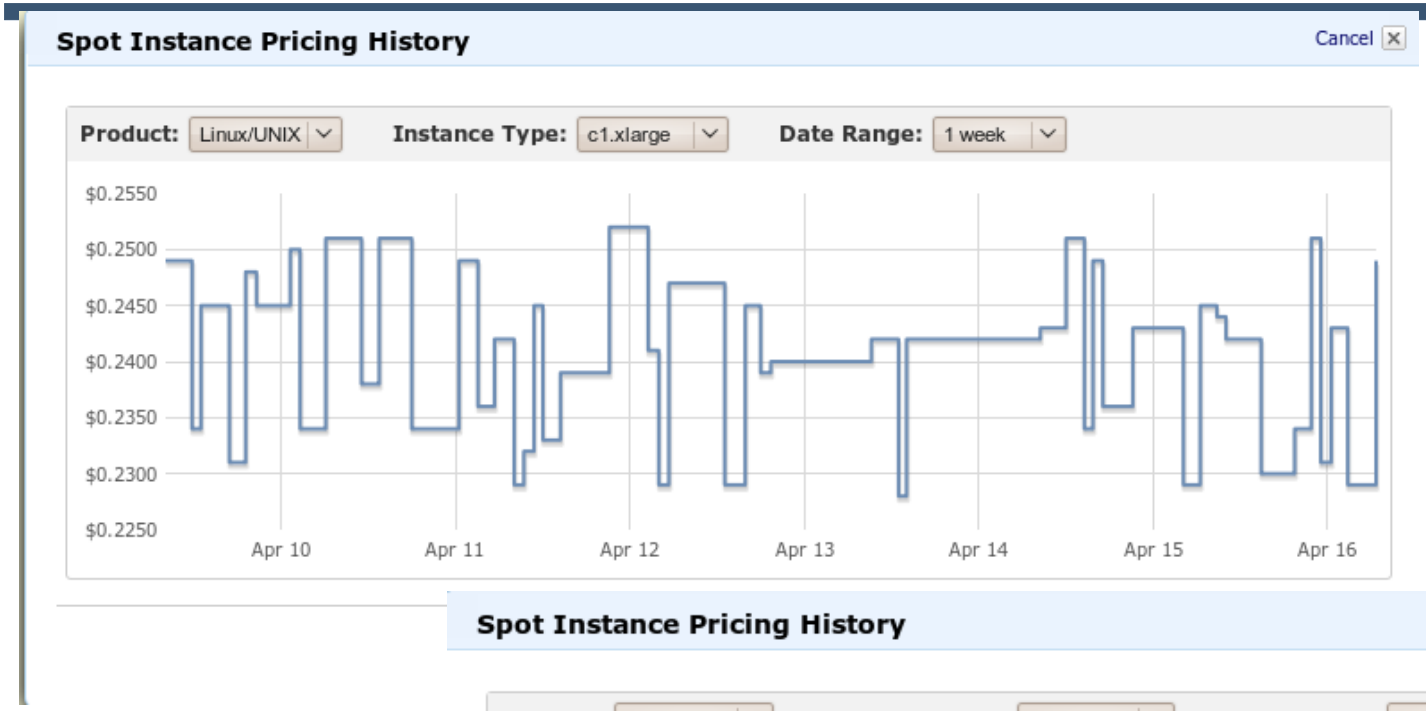
Since then...

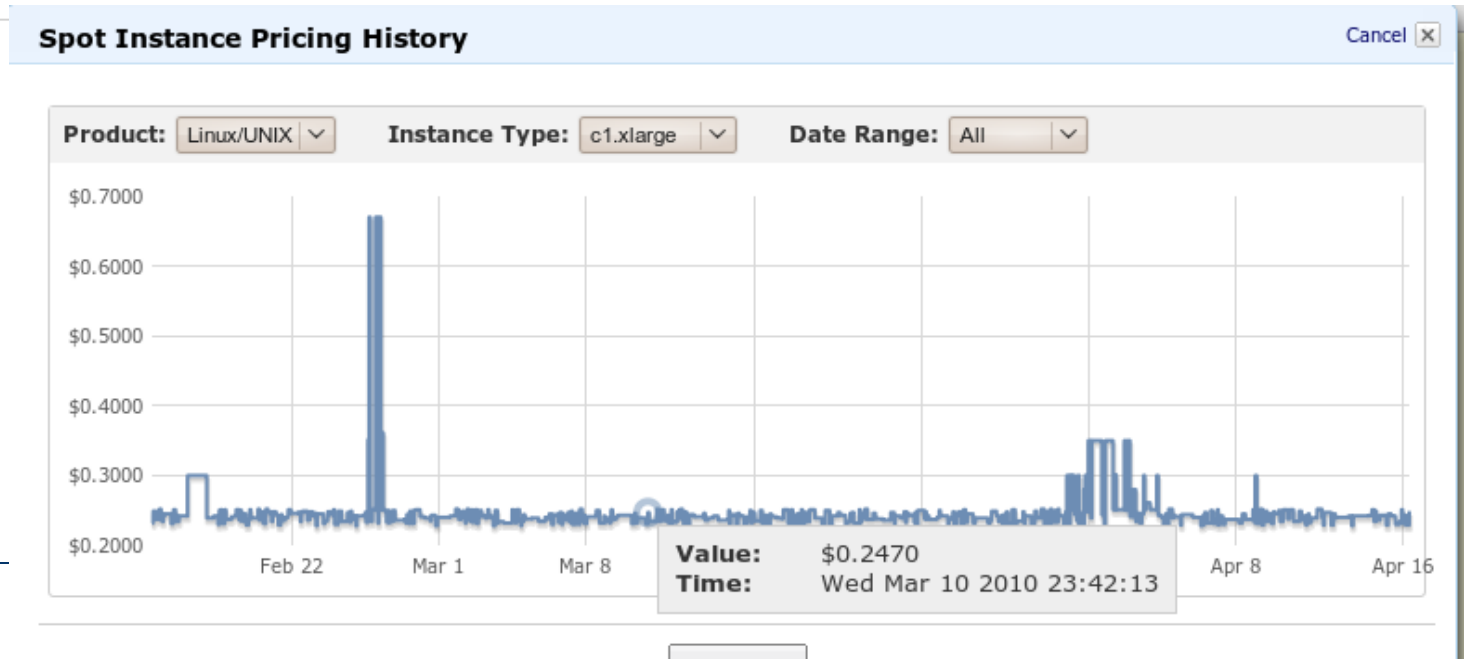

Data Transferred by site - Mar - Aug 2010
21 Weeks from Week 09 of 2010 to Week 30 of 2010
Gbyte
20,000
15,000
10,000
5,000
0
Mar 2010
Apr 2010
May 2010
Jun 2010
Jul 2010
KEK, Japan
KIT, Germany
kek2-se.cc.kek.jp 17931.3
gridka-dcache.fzk.de 2507.5
dcache.ijs.si 501.2
dpm.cyf-kr.edu.pl 106.4
Amazon-OUT01-SE 93.8
Amazon-IN01-SE 93.1
Amazon-IN02-SE 0.0
dg24.cc.kek.jp 0.0
grid-srm.rzg.mpg.de 0.0
Generated on 2010-08-18 02:51:31 UTC

What worked

- CPU efficiency was >95%
- Cloud very stable: no job failures on cloud
- Cloud can support long jobs and multi-core jobs (grid has issues)
- DIRAC, running on a couple of 1 core 2GB RAM VMs in Barcelona scaled very well
- Input data worked equally well whether it was located on cloud or grid, from any of three paradigms
- Network – we could run our storage at the maximum rate
  - Other groups have tested international bandwidth to 500MB/s
  - Peering with academic/research networks could be useful

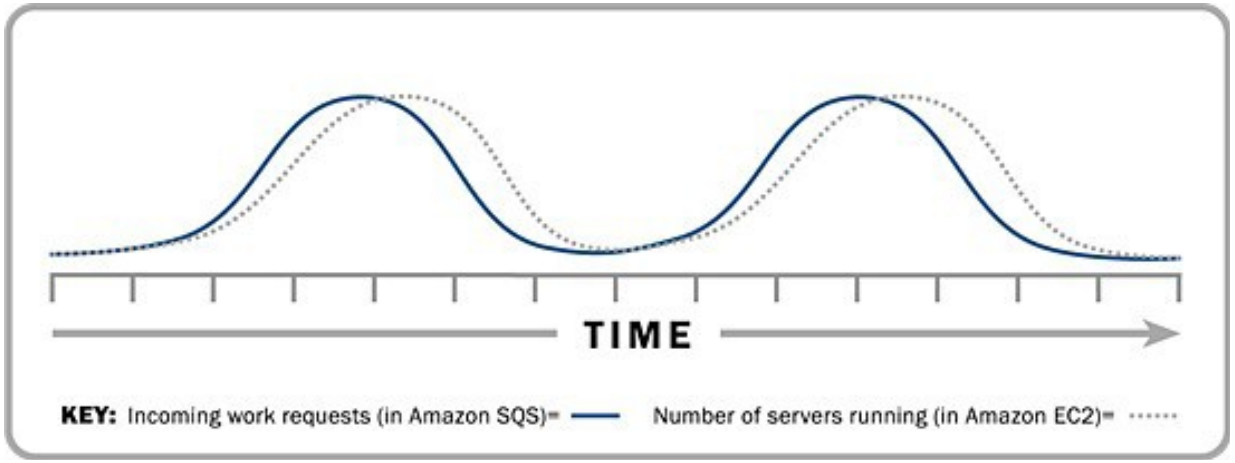
Cloud Computing – Spot Pricing





Tips for using the cloud

- Only keep VMs you need running



- Data inside the cloud is free
- Pull Scheduling
- To avoid vendor lock-in effect, treat cloud as truly elastic

Costings

- On the cloud, it costs us USD0.20 for 10,000 simulated collisions, including data in/out, overheads etc
  - We just buy capacity for 5 months of the year
- $4000 server, 5 events/sec, 3 years
  - USD0.08/10k simulated collisions
- Electricity, 1kW @ UDS0.08/kWh (KEK, Japan rate)
  - USD0.12/10k simulated collisions
- Physical infrastructure? Rack space? Cooling? Network?
  - https://spreadsheets.google.com/ccc?key=toEOU0bONc8D-z6xU0FRt-w
- SysAdmin time?
  - How much would it cost for a 2000 core cluster?
- Depreciation of computing output value over time?
  - $VWO = \sum_{t=0}^{3} Xe^{-\lambda . t} \simeq \int_{0}^{3} Xe^{-\lambda t} = 2.05\,X$

Cloud for you?

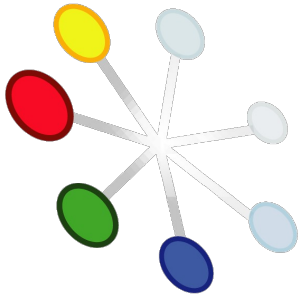
- If you're latency sensitive (eg MPI), regular offerings might not be appropriate
  - Try http://aws.amazon.com/ec2/hpc-applications/
  - A little expensive. Still works with DIRAC though!
- If you have a similar processing model (H**T**C), you can use this work
  - Haven't got access to a cluster?
    - You can test it today, no approval, no paperwork
  - Need a specific operating system, or package?
    - Have exactly what you want.
  - Short-term needs? Pre-conference rushes?
    - Just buy what you need

Further reading

- Computing at Belle II
  - http://www.kek.jp/intra-e/feature/2010/BelleIIComputing.html
- Our case study at Amazon
  - http://aws.amazon.com/solutions/case-studies/university-melbourne-barcelona/
- Musings on data transit
  - http://www.itnews.com.au/News/224403,researchers-rue-cost-of-public-cloud-data
- Background on DIRAC, vendor lock-in
  - http://www.itnews.com.au/News/229403,scientists-rein-in-the-commercial-cloud.as
- Article on the project with a photo that has lots of cables
  - http://www.theaustralian.com.au/australian-it/the-cloud-helps-with-lifes-curliest-qu
- Software desarrollado por científicos de la UB mejora la gestión de grandes procesos de cálculo mediante sistemas comerciales de computación
  - http://www.universia.es/portada/actualidad/noticia_actualidad.jsp?noticia=106910
- Above the Clouds: Managing Risk in the World of Cloud Computing
  - Kevin T. McDonald - IT Governance Ltd - February 23, 2010

Adria Casajus Ramo, Ricardo Graciani Diaz, Ana Carmona Agüero
Tom Fifield, Martin Sevior, the DIRAC team and
the Belle II computing group

Questions?
dirac.project@gmail.com



Cloud – NIST definition

On-demand self-service.

Broad network access.

Resource pooling.

Rapid elasticity

Measured Service.

Cloud Infrastructure as a Service (IaaS). The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls).