### **Usecases of LCG-2 at GridKa**

Johan Blouw

Max Planck Institute for Nuclear Physics



Usecases of LCG-2 at GridKa – p. 1

### Overview

- Introduction
- LHC
- LHCb
- Distributed Computing
- GridKa
- DC03
- DC04
- Conclusions



### Introduction

B-physics in the 21-st century:

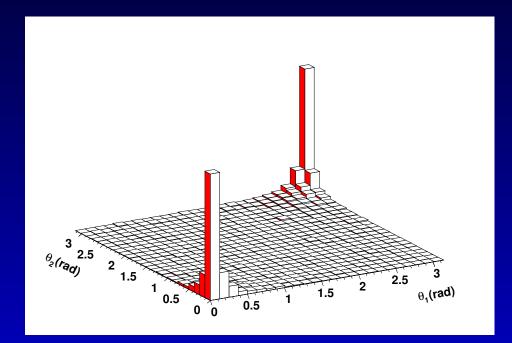
- Use hadron machines (CERN, DESY, Fermilab...)
- high cross section and high interaction rate  $\rightarrow$  lot of data
- Measure cp-violation in B<sub>s</sub> meson system
- Search for rare decays
- over-constrain CKM-triangle
- Search for physics beyond the Standard Model





### Introduction

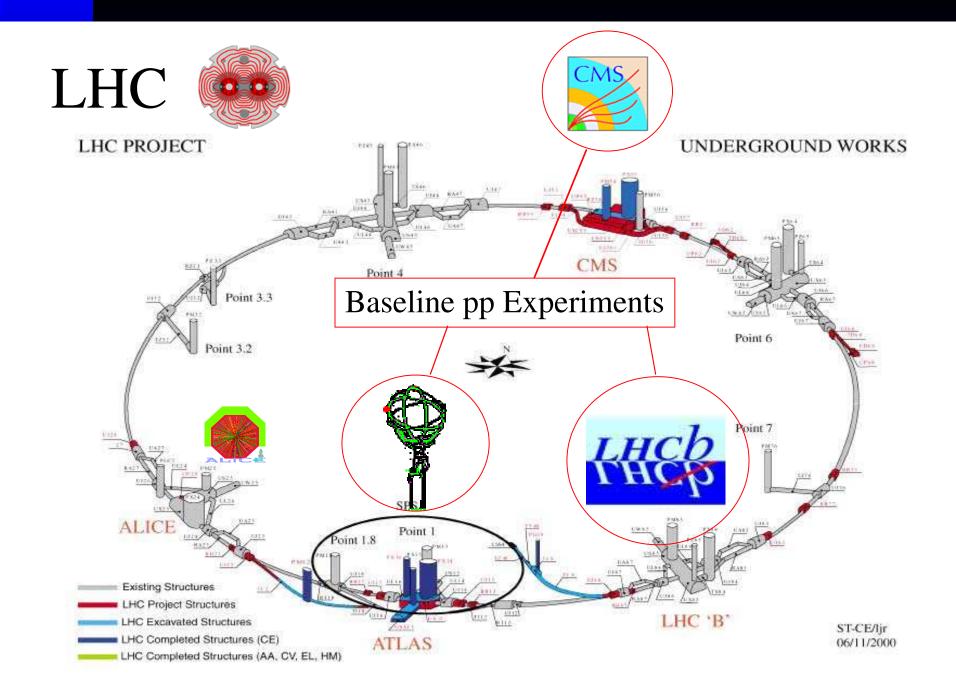
### • $b\bar{b}$ mesons produced at forward/backward angles



• large boost  $\rightarrow$  may resolve  $B_s - \bar{B}_s$ oscillations...

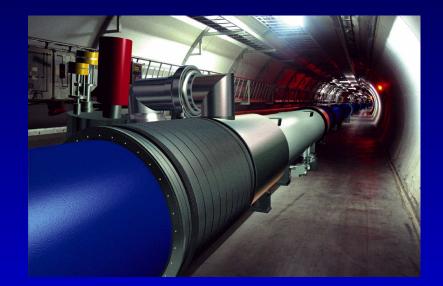


 simple forward spectrometer sufficient (HERAB, BTEV, LHCb)
 Usecases of LCG-2 at GridKa - p. 3



### LHC at CERN

- $p\bar{p}$  collisions at  $\sqrt{s} = 14 \text{ TeV}$
- 40 MHz bunch crossing rate
- $\mathcal{L} \sim 10^{32} \,\mathrm{cm}^{-2} \mathrm{s}^{-1}$



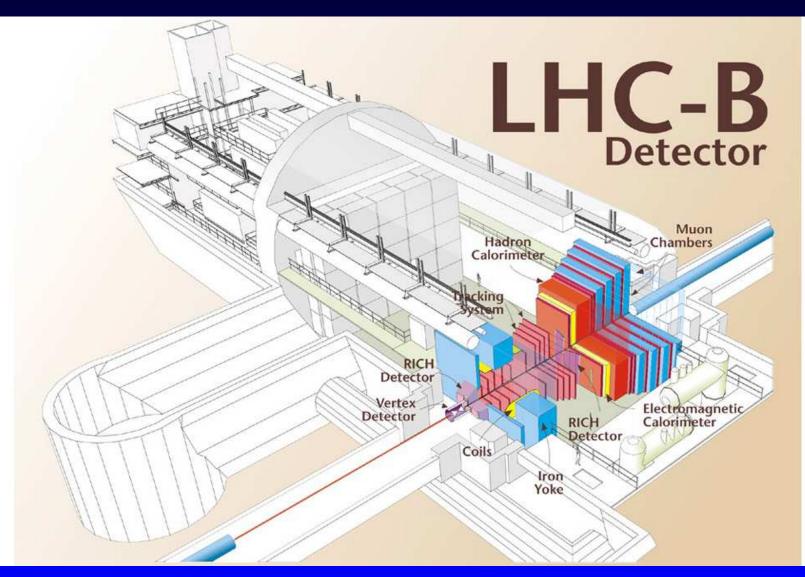


### LHC at CERN

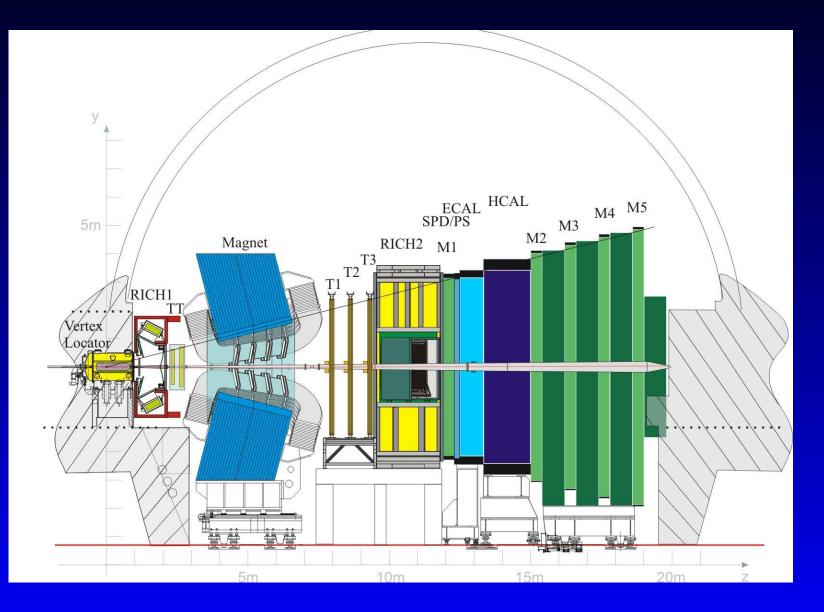
- large  $b\overline{b}$  cross section:  $\sigma \sim 500 \ \mu b$
- signal/noise ratio:  $\sigma_{b\bar{b}}/\sigma_{\rm inelastic} \sim 5 \times 10^{-3}$
- $10^{12}$  bb pairs per year



### LHCb detector located in DELPHI pit:









• Study physics performance

depends on

- trigger performance (efficiency, event rejection...)
- tracking performance (efficiency, ghosts, momentum resolution)
- material budget
- $\Rightarrow$  use detector and event simulation

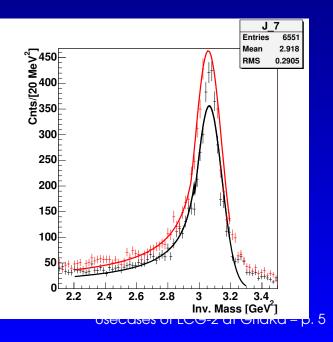


predicted rates for 1 Snowmass year:

- $10^{12} b\bar{b}$  events per year
- at 10 MHz of visible pp crossings
- Level 0: 40 MHz input, reduce to 1 MHz
- Level 1: reduce 1 MHz to 40 KHz
- HLT: reduce to 200 Hz
- Data volume:  $200 \text{ Hz} \times 1 \text{ Snowmass year} = 2 \times 10^9 \text{ events}$
- at 50 KByte/event  $\rightarrow \mathcal{O}(10^4)$  TByte/year



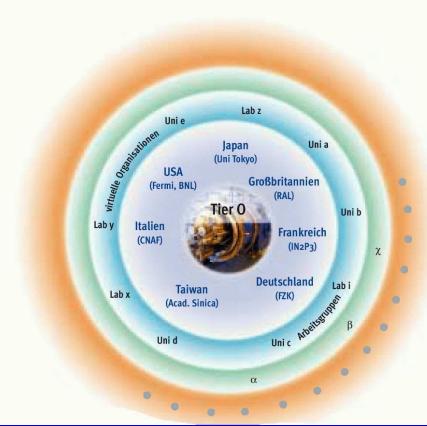
- enormous amount of simulated data needed
- detector simulation:  $\sim \text{ minute/event/1GHz}$
- trigger studies
- background studies
- mass reconstruction efficiency studies
- cp-reach, rare decays, etc. etc...
- $\Rightarrow$  too much data for CERN alone





## **Computing Challenge**

- use distributed computing resources (cpu, storage)
- need high bandwidth connections between computing centers
- $\Rightarrow$  birth of GRID...





### **Computing Challenge** GRID: nice idea, but no easy meat...

### Installation Automation, Operation, Control Configuration + monitoring Fault tolerance Infrastructure Electricity, Cooling, Space **Batch** system (LSF, CPU server) Storage system (AFS, CASTOR, disk server) Network Benchmarks, R&D, Architecture GRID services !? Prototype, Testbeds Purchase, Hardware selection, **Resource** planning Coupling of components through hardware and software



Distributed Infrastructure with Remote Agent Control... Different from DataGRID philosophy:

- jobs not time critical
- many more computing jobs than computing resources
- ! use "PULL" concept instead of "PUSH".
- run job-agents at each participating site
- job-agents request jobs from central job-server

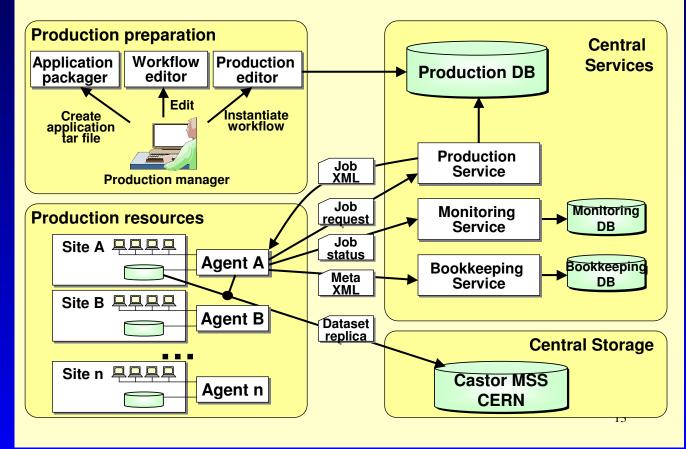


- Production task definition
- Software installation on production sites
- Production steering
- Job scheduling
- Job monitoring
- data transfer
- bookkeeping



• use simple system of job submission

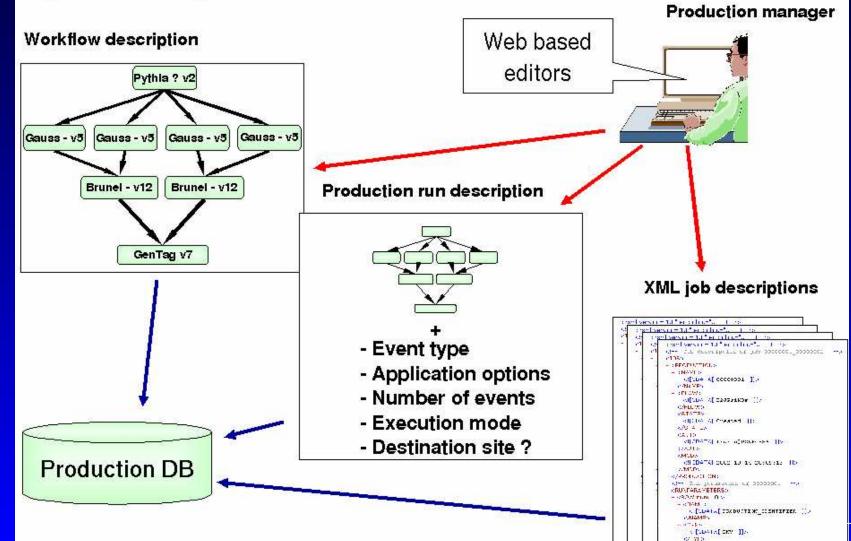
### **DIRAC** architecture reminder





### Job description in 'sandbox':

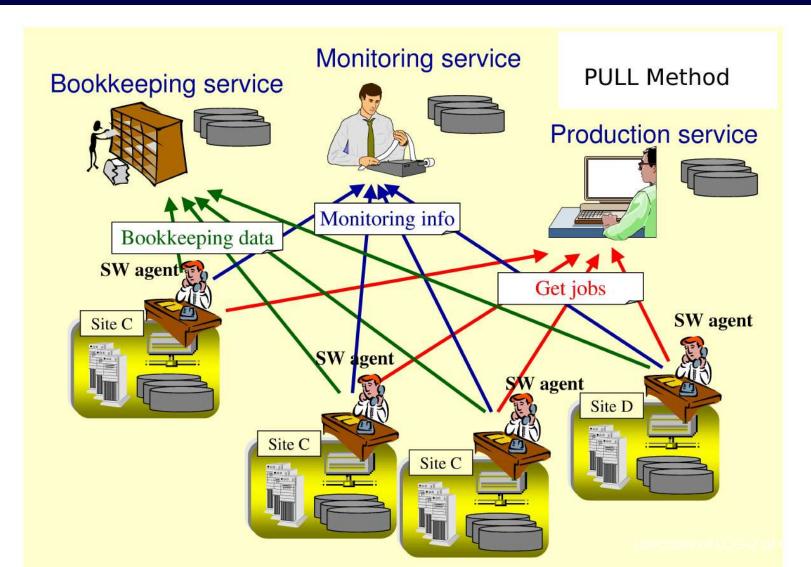
### Job description



p. 7



### • use the "PULL" approach:





## **LHCb DC03 Goals**

Goals:

- Produce 30 million minimum bias events
- 10 million BB events
- 38 different signal event channels (50K each)
- use data for "re-optimized TDR"
- physics reach studies



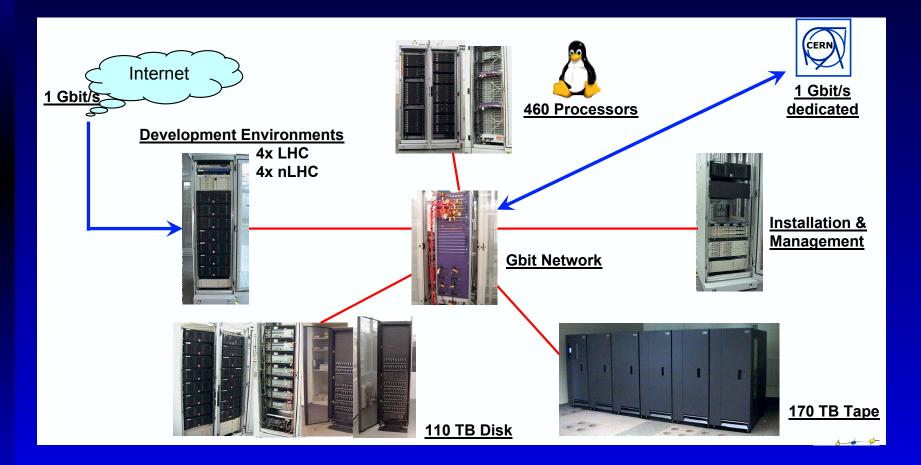
## LHCb DC03 Goals

Participating sites: CERN Imperial ScotGrid Bologna Lyon RAL Karlsruhe Barcelona Rio Cambridge CESGA **Bristol** Oxford **VU** Amsterdam

dam ...

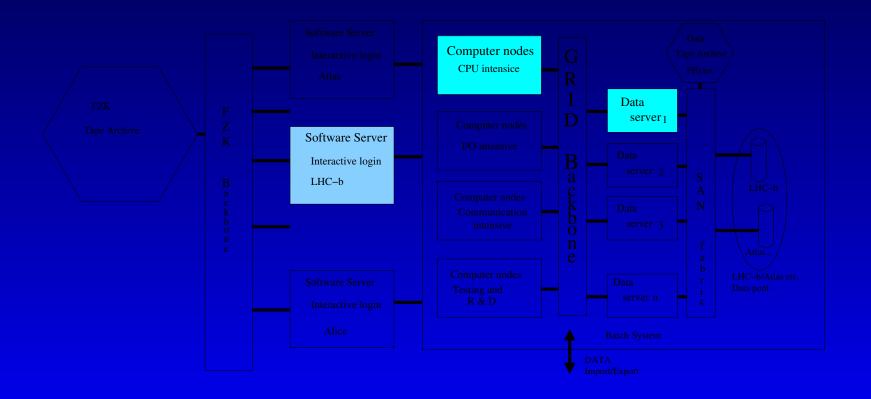


Usecases of LCG-2 at GridKa – p. 8



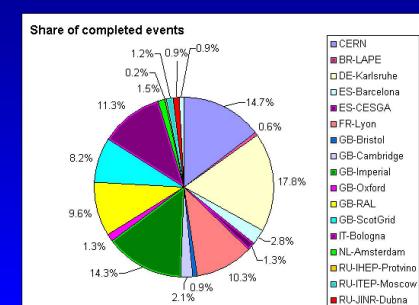


- But software did not use hardware optimally
- $\rightarrow$  run < 50 jobs because of limited disk i/o
  - major debugging necessary to solve bottleneck





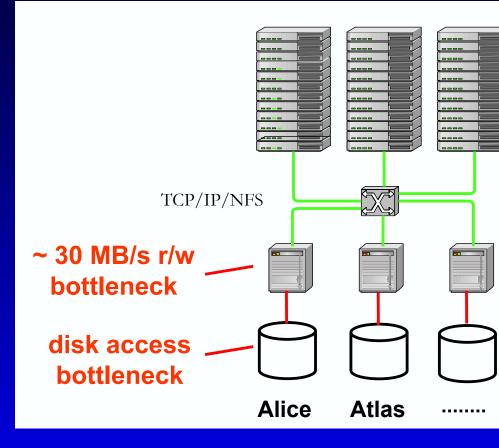
- Each WN accessed FS simultaneously for read/write
- data through-put to FS limited by FS
- $\Rightarrow$  write data to local disk first
  - ! copy data to FS at end of job
  - ! mainly hacking of job-submission software...



Usecases of LCG-2 at GridKa – p. 9

But:

- excellent cooperation with GridKa:
- → use three different FS (1
   big slow, 2 small fast
   servers)
  - while WNs write to fast FS
  - move data from 2nd fast FS to slow FS
  - alternate writing to fast FS1 and FS2





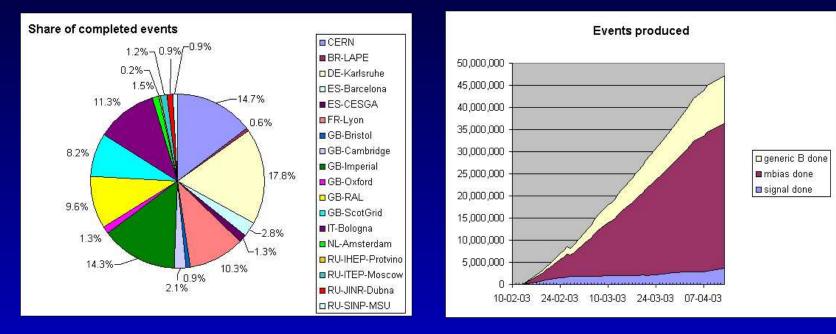
### **DC03 Results**

### Usage Example: LHCb Data Challenge'03 - Results

- A total of 47 Million events have been produced in two months
- 18 centres participate and 80% of CPU outside CERN
- 36 600 jobs have been run and each job:
  - producing between 250 and 500 events
  - using from 32 to 56 hours on 1GHz PC
- It would have taken more than 170 years on a single PC



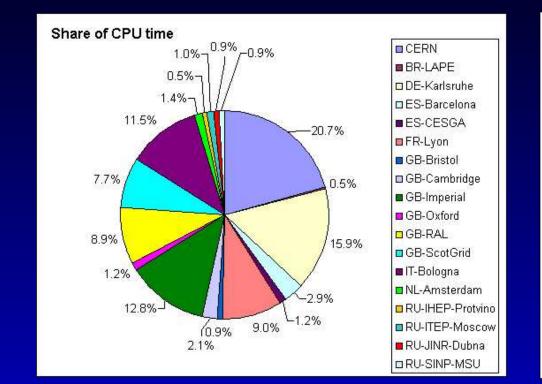
# DC03 Results GridKa contributed ~ 18% of produced events

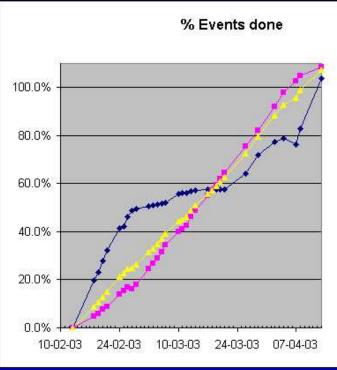


### Very successful!



### **DC03 Results**





🔶 % tot

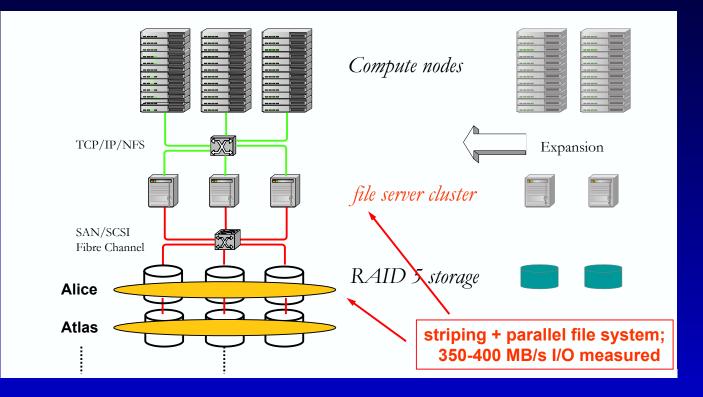
\_\_\_% tot

% tot



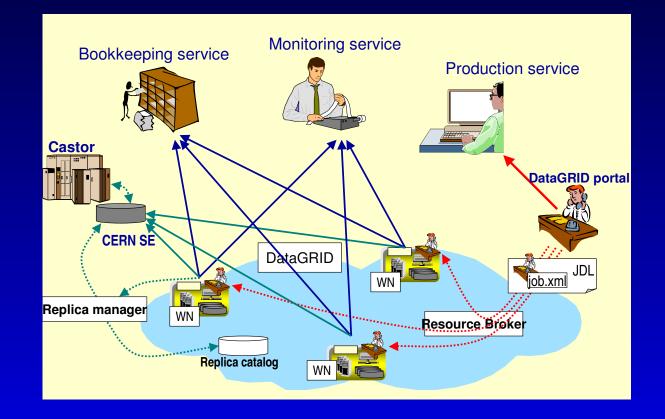
Usecases of LCG-2 at GridKa – p. 10

### GridKa improved internal data through-put:





# LHCb modified DIRAC to include job submission through LCG

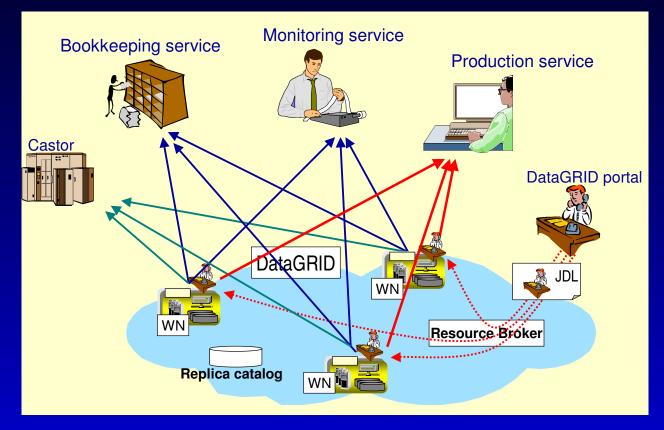




- accomodate LHCb specific software with
- "sandbox" technology:
  - job description
  - launch script
- use Grid software for transferring data to CERN
- but: need sandboxes for log files...
- DataGrid not very stable
  - Resource Broker unavailable
  - Proxy expires before job is over
  - large variation in available CPU power
  - some WN have no outbound IP connection

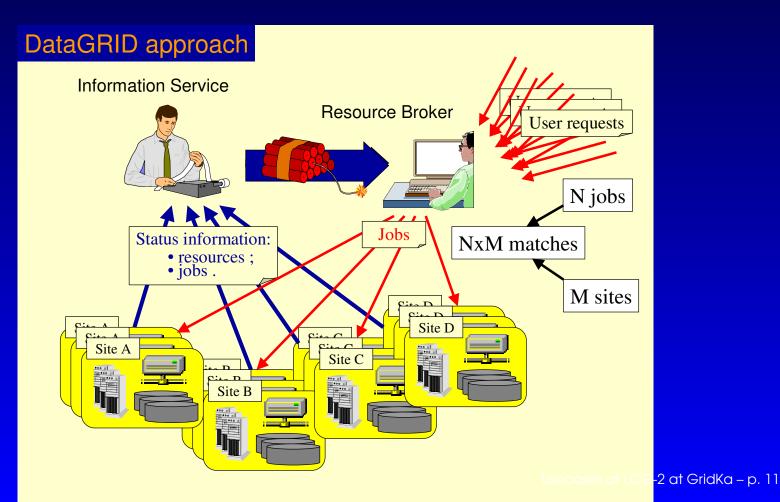


### Circumvent GridPortal...





DIRAC $\iff$ DataGridPULL $\iff$ PUSHdecentralized job $\iff$ centralized job



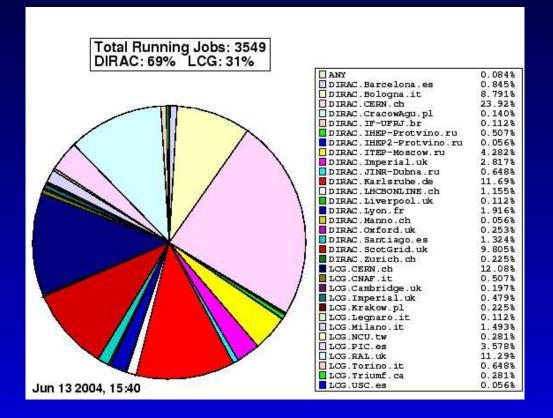


- Job submission software more complicated...
- site configuration much easier
- DIRAC also on DataGrid middleware
- $\rightarrow$  non-trivial exercise...



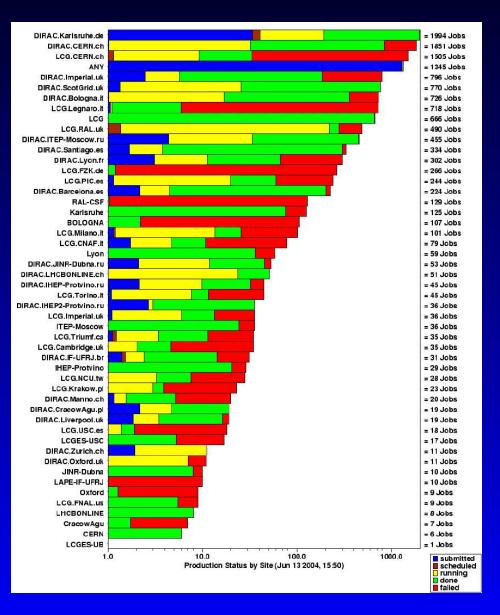


# By using LCG, many more sites may contribute to data challenge





- many updates of DIRAC software
- hiccups in data transfer to CERN
- $\rightarrow \sim 1300 \text{ files}$ waiting to be transferred from FZK to CERN...
- → transfer can not keep up with data production!





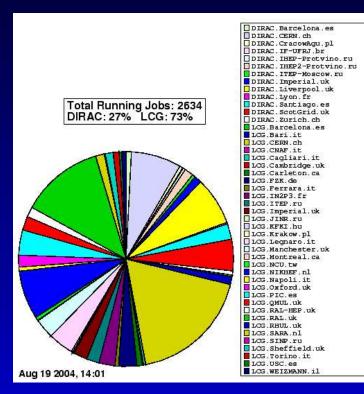
LCG related problems:

- Resource Broker bottle-neck
- Determine Ranking (which site gets the LCG jobs?)
- Faulty WM sucks up all LCG jobs
- Software installation
- Site configuration (not always according to LCG-2 specs)
- Proxy expiration
- 10% of jobs aborted at 'job-finilization' stage



### Hardware problems:

- RB server machine at CERN suffer from high-load problems
- log-file server sshd problems
- Bookkeeping dbase problems
- HW problems at local sites





0.797%

7 555%

0.493%

0 075%

0 379%

1.328%

0.493%

1.100%

7.213%

0.227%

2.353%

4.745%

0.607%

0.227%

1.214%

18.22%

0.037%

0.303%

0.189%

0 7975

2 733%

0.531%

2.429%

1.898%

2.088%

0.189%

0.037%

0.227%

3.796%

2.961%

0.075%

0.531%

7.3278

0.607%

1.784%

3.796%

2 088%

1.518% 12.41%

0.037%

1.290%

0.151%

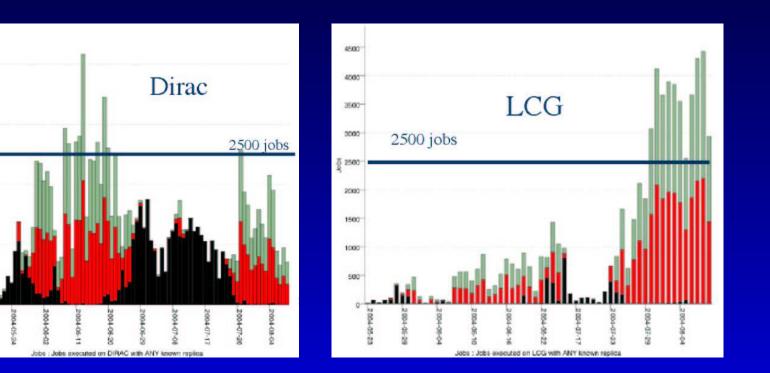
1.176%

0.911%

0.379%

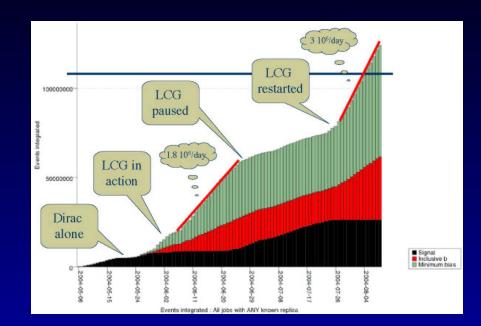
0.645%

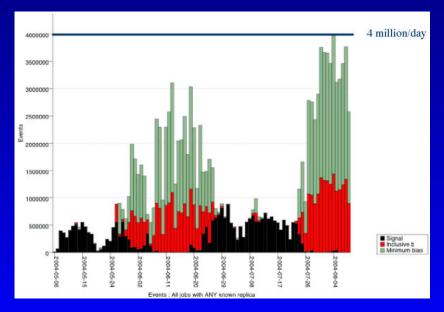
## DC04 Experiences... DIRAC & LCG contributions: DIRAC



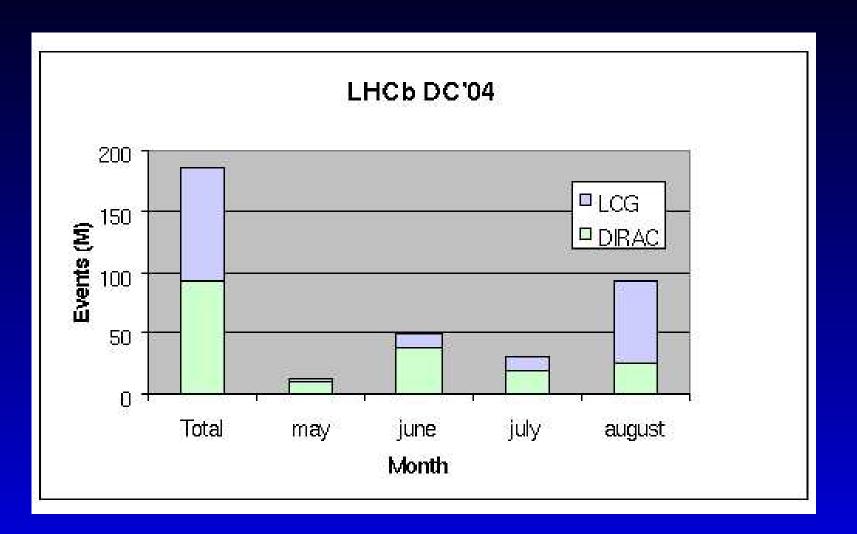


LCG



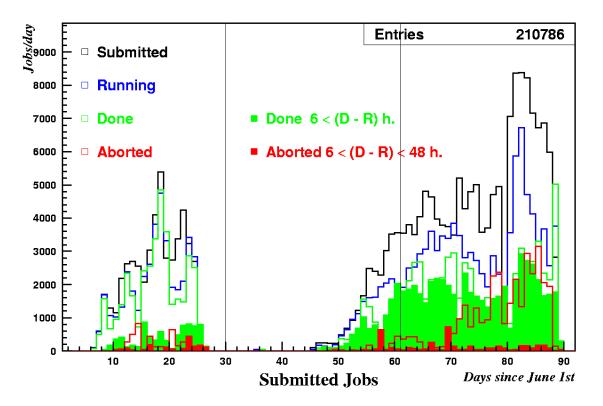








### LHCb DC'04 (LCG)



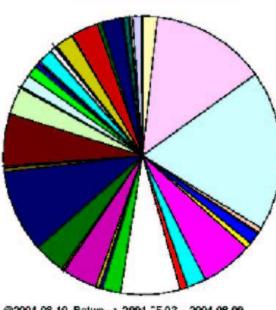


Max-Rlanck-Institut für Kernphysik

## **DC04 Summary**

### 283 CPU.Years

#### CPU Time: 2482042 h



@2004-08-10 Between 2004 05-03 - 2004-08-09

ANY	0.0019
CracowAcu	0.0031
DIRAC Barcelona es	1.723
DIRAC.Bologna.it DIRAC.CERN.ch	13.348
DIRAC.CERN.ch	18.651
DIRAC.CracowAgu.pl DIRAC.IF-UFRJ.br DIRAC.IHEP-Protvino.ru DIRAC.IHEP2-Protvino.ru DIRAC.IHEP2-Protvino.ru DIRAC.IHEP-Moscow.ru DIRAC.Inperial.uk DIRAC.JINR-Dubna.ru DIRAC.JINR-Dubna.ru	0.6049
DIRAC. IF-UFRJ.br	0.1315
DIRAC. IHEP-Protvine. ru	1.3621
DIRAC. IHEP2-Protvine. ru	0.702
DIRAC. ITEP-MOSCOW. FU	6.0019
DIRAC. Imperial.uk	2.220
DIRAC . JINR-Dubna . ru	0.8631
DIRAC.Karlsruhe.de DIRAC.LHCBONLINE.ch DIRAC.Liverpool.uk	6.9021
DIRAC . LHCBONLINE . ch	2.1278
DIRAC. Liverpool.uk	0.3691
DIRAC. Lpool.uk	0.4421
DIRAC. Lyon. fr	4.0371
DIRAC . Manno . ch	0.0771
DIRAC . Onford. uk	0.251
DIRAC.Santiago.ea	3.6831
DIRAC. ScotGrid. uk	9.6001
DIRAC. Zurich. ch	0.451
DIRAC.Liverpool.uk DIRAC.Lpool.uk DIRAC.Manno.ch DIRAC.Manno.ch DIRAC.Santiago.es DIRAC.Santiago.es DIRAC.ScotGrid.uk DIRAC.Zurich.ch LOG.BHAM-HEP.uk LOG.BARCelona.es LOG.CERN.ch	0.2231
LOG.Barcelona.es	0.0221
LOG.CERN.ch	6.051
	3.195
LOG.Cambridge.uk	0.1411
LOG.Carleton.ca	0.006
LOG.FNAL.US LOG.FZK.de	0.0561
LOG.FZK.de	1.4938
LOG. IN2P3. fr	0.0094
LOG.IN2P3.fr LOG.Imperial.uk LOG.KFRI.hu	1.290
LOG .KFKI .hu	0.4371
LOG.KFEKOW.D1	0.1061
LOG.Lancashire.uk	0.2241
LOG.Legnare.it	1.6721
LOG.Lancashire.uk LOG.Legnaro.it LOG.Nanchester.uk	0.0521
LOG.Milano.it	0.614
LOG.Montreal.ca	0.014
LOG.Milano.it LOG.Montreal.ca LOG.NCU.tw	0.250
LOG.NIKHEF.nl LOG.Napoli.it LOG.Oxford.uk LOG.PIC.es LOG.RAL-HEP.uk LOG.RAL.uk	2.0991
LOG.Napoli.it	0.0331
LOG.Oxford.uk	0.0478
LOG.PIC.es	3.1371
LOG.RAL-HEP.uk	0.4799
LOG.RAL.uk	2.3871
LOG.Roma.it	0.0861
LOG.Roma.it LOG.Sheffield.uk LOG.Torino.it LOG.Toronto.ca LOG.Triunf.ca	0.108
LOG. Torino.it	0.5921
LOG. Toronto.ca	0.165
LOG. Triumf. ca.	0.2871
LOG.UCL-CCC.uk	0.9671
LOG.USC.es	0.1775
Oxford	0.005

LCG Job Statistics

	Jobs	Jobs(k)	%Sub	%Remain
Submitted	210786	211	100.00%	
Cancelled	25669	26	12.18%	
Remaining	185117	185	87.82%	100.00%
Aborted (not Run)	37157	37	17.63%	20.07%
Running	147596	148	70.02%	79.73%
Aborted (Run)	34228	34	16.24%	18.49%
Done	113368	113	53.78%	61.24%
Retrieved	113331	113	53.77%	61.22%

	Jobs	Jobs(k)	%Sub	%Remain
Done	113368	113	53.78%	61.24%
Done (CPU > 6h)	69296	69	32.88%	37.43%
Aborted (CPU > 6h)	9971	10	4.73%	5.39%

Output Sandbox Analysis: 69K Successful jobs

LHCb Accounting: 81K Successful jobs



### Conclusions

- DC03 was very successful
- DIRAC works! ("PULL" vs. "PUSH")
- DC04: many hiccups/problems solved
- include running on LCG (non-trivial)

To do:

- skimming of data
- incorporate distributed analyses
- employ more gridtools

