LHCb experience from recent productions / challenges

Joël Closier

CERN / LHCb

SC3 Detailed Planning Workshop

Outline

- DC04 phase 1
  - Monte Carlo production on all LCG sites and LHCb native sites
  - All DST transferred at Tier0 and Tier1
  - Summer 2004 (DC04 – v1) and December 2004 (DC04 – v2)
- DC04 phase 2
  - DST stripping on DC04 – v1 data on 3 Tier1s
  - Stripped DST transferred at Tier0 and Tier1
  - March 2005
- RTTC
  - Monte Carlo production on all LCG sites
  - All output after L0+L1 transferred at Tier0
- DC04 phase 2 next round
  - DST stripping on DC04 – v2 data
  - Stripped DST transferred at Tier0 and Tier1
  - June 2005

Results of the LHCb experiment Data Challenge 2004 (phase 1)

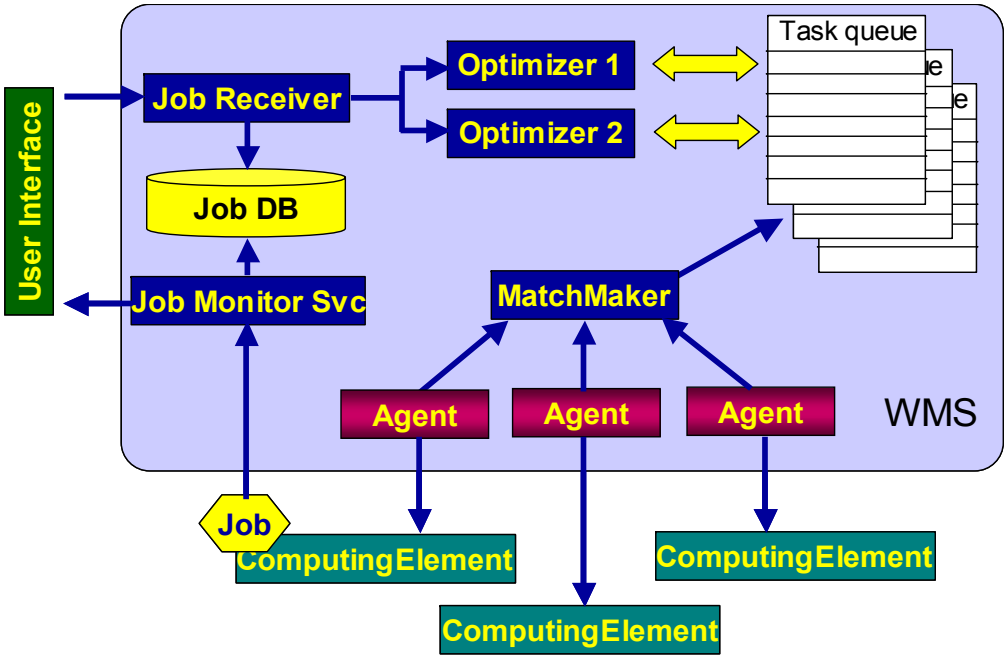
Summer 2004

LHCb DC'04 aims

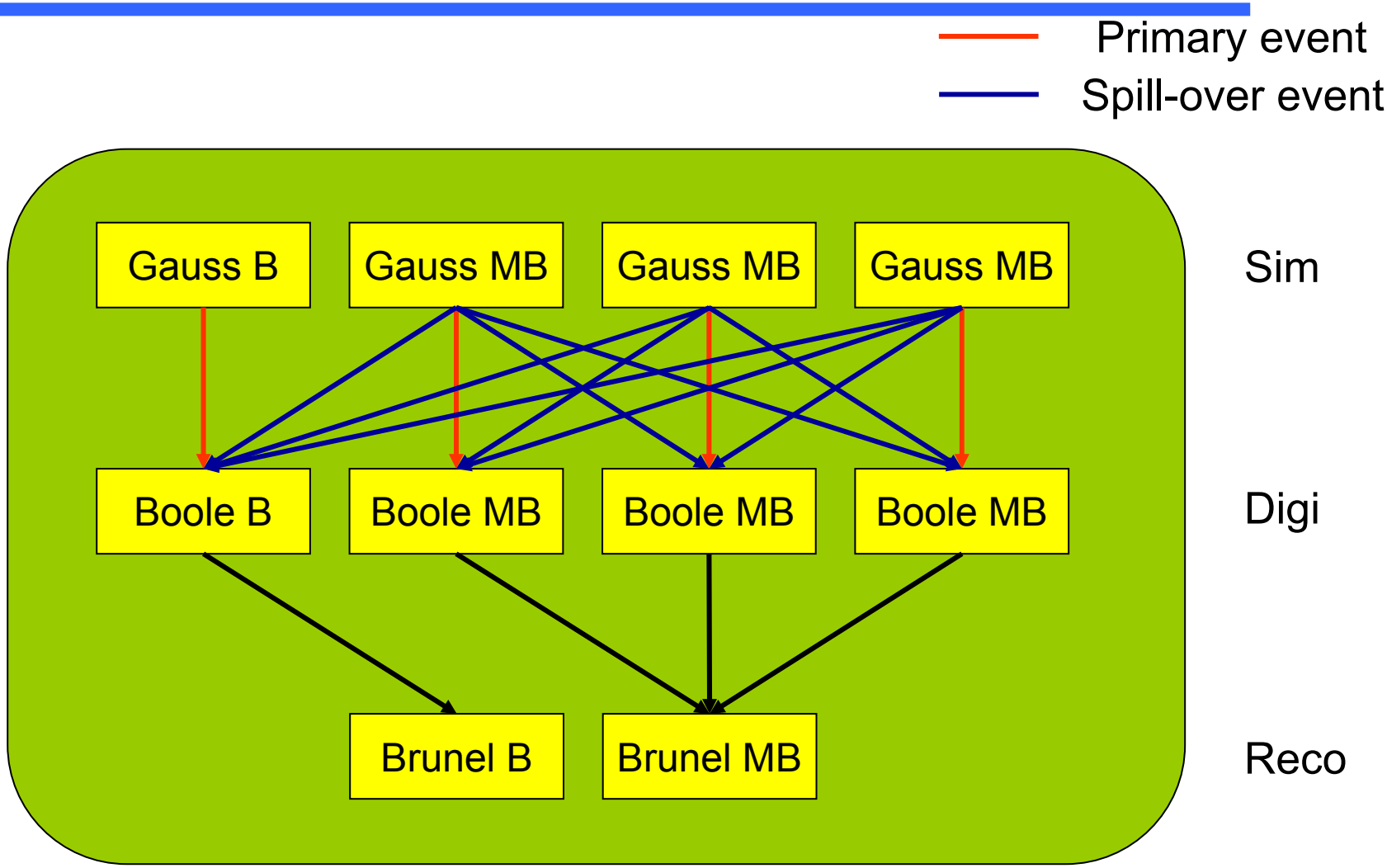
- Main goal :gather information to be used for writing the LHCb computing Technical Design Report
  - Robustness test of the LHCb software and production system
    - Using software as realistic as possible in terms of performance
  - Test of the LHCb distributed computing model
    - Including distributed analyses
    - Realistic test of analysis environment, need realistic analyses
  - Incorporation of the LCG application area software into the LHCb production environment
  - Use of LCG resources (at least 50% of the production capacity)
  - 3 phases
    - Production : MC simulation and reconstruction
    - Stripping : Event pre-selection
    - Analysis

DIRAC workload management

- Realizes ***PULL*** scheduling paradigm
- Agents are requesting jobs whenever the corresponding resource is free
- Using Condor ClassAd and Matchmaker for finding jobs suitable to the resource profile
- Agents are steering job execution on site
- Jobs are reporting their state and environment to central Job Monitoring service



Job workflow

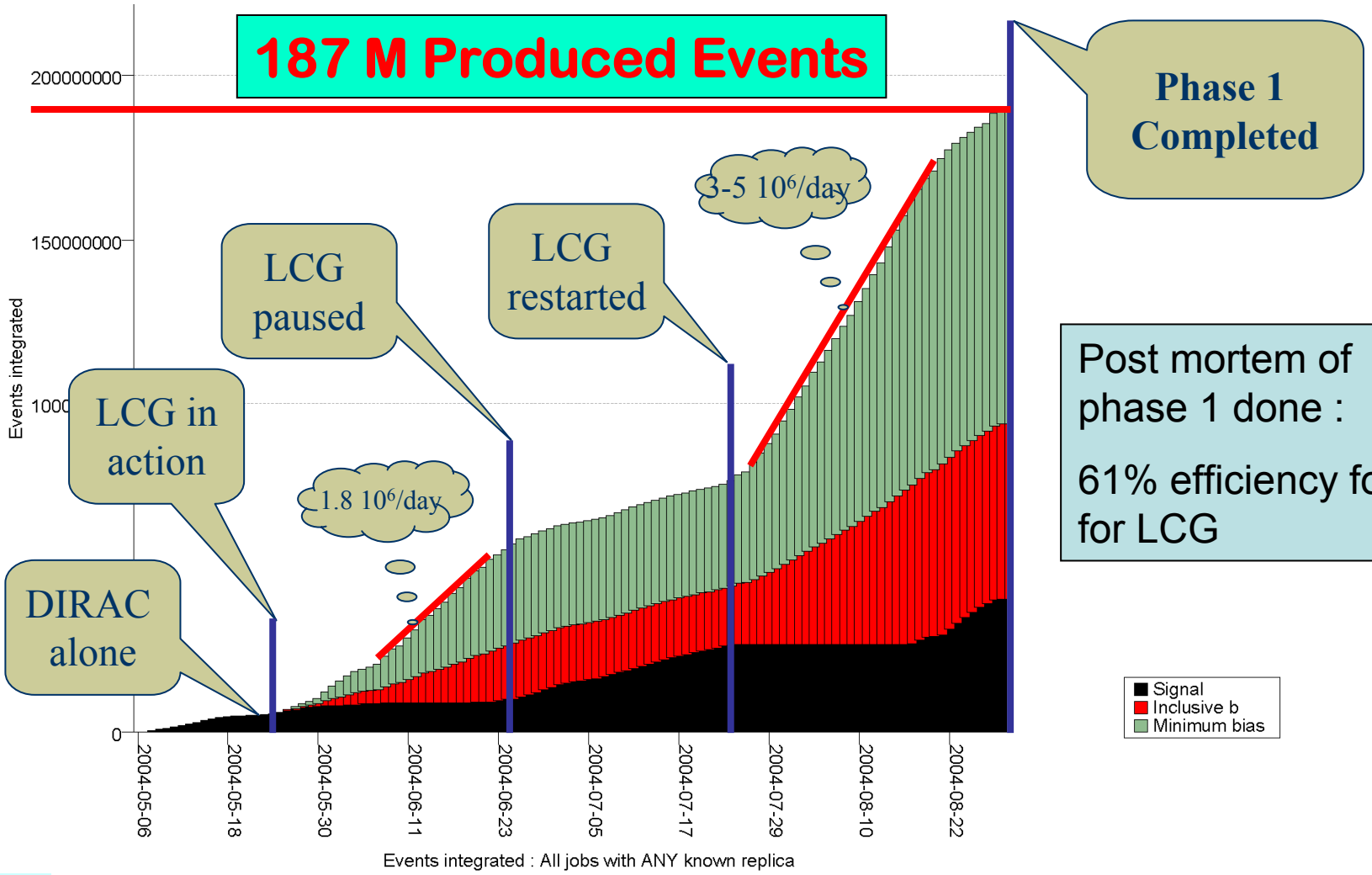


Data Storage

- All the output of the reconstructed phase (DST) were send to CERN (as Tier0)
- Intermediate files were not kept.
- DSTs were also stored in one of 5 centres (these centres map to our expected TIER1)
- TIER1
  - CNAF (Italy)
  - Karlsruhe (Germany)
  - Lyon (France)
  - PIC (Spain)
  - RAL (United Kingdom)
  - NIKHEF (Netherlands) (a candidate towards the end of the DC04 phase 1)

DC'04 performances

Phase 1 results



Post mortem of phase 1 done :

61% efficiency found for LCG

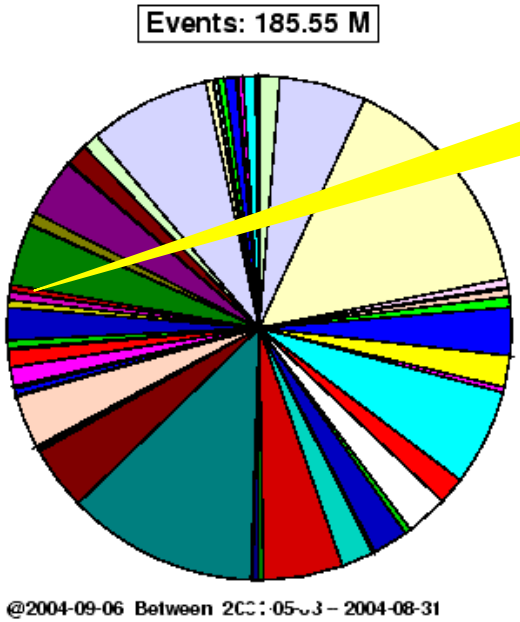
TIER storage

TIER 0	Nb of Events	Size (TB)
CERN	187 557 231	62

Tier 1	Nb of Events	Size (TB)
CNAF	37 129 350	12.6
RAL	19 462 850	6.5
PIC	16 505 010	5.4
Karlsruhe	12 486 300	4
Lyon	4 368 656	1.5

Sites involved

424 CPU years



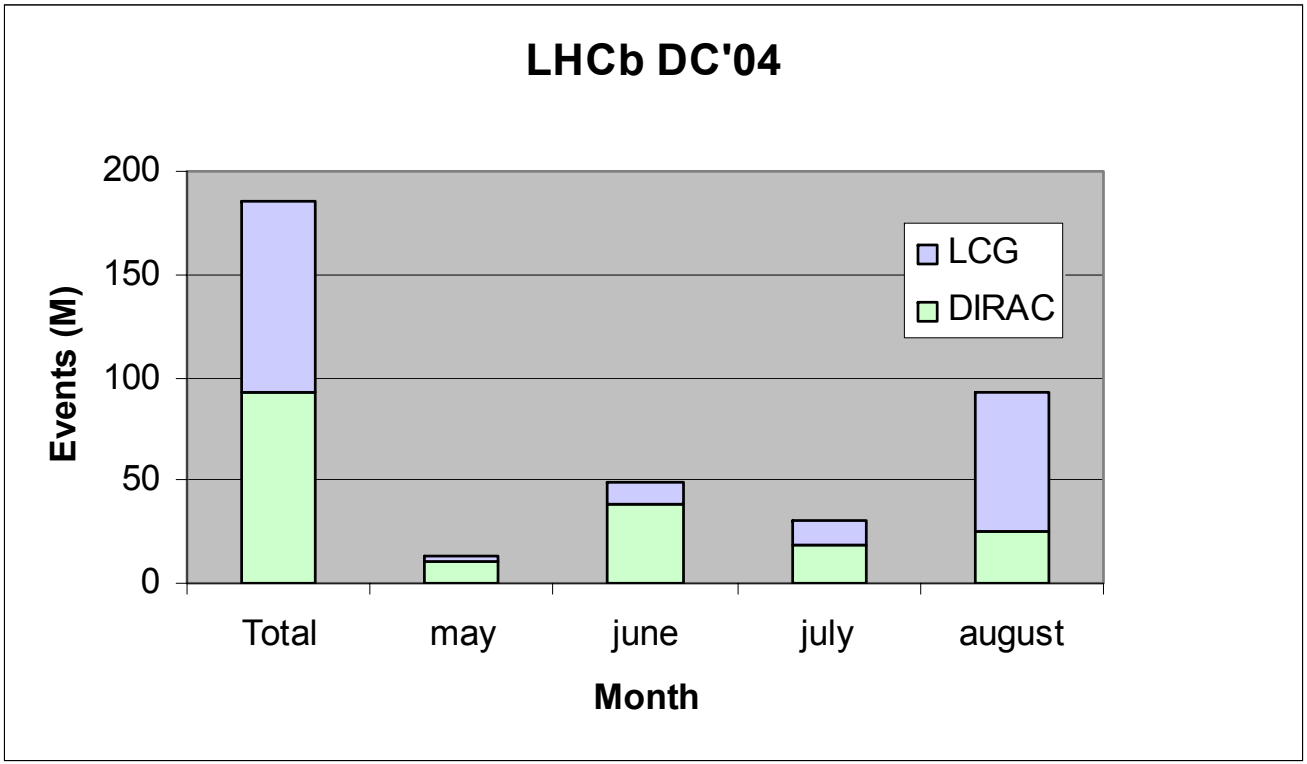
DIRAC.Barcelona.es	1.305%
DIRAC.Bologna.it	5.560%
DIRAC.CERN.ch	14.96%
DIRAC.CracowAgu.pl	0.532%
DIRAC.IF-UFRJ.br	0.124%
DIRAC.IHEP-Protvino.ru	0.504%
DIRAC.IHEP2-Protvino.ru	0.691%
DIRAC.ITEP-Moscow.ru	3.066%
DIRAC.Imperial.uk	2.017%
DIRAC.JINR-Dubna.ru	0.353%
DIRAC.Karlsruhe.de	6.181%
DIRAC.LHCBONLINE.ch	1.752%
DIRAC.Liverpool.uk	2.674%
DIRAC.Lpool.uk	0.405%
DIRAC.Lyon.fr	2.273%
DIRAC.Manno.ch	0.035%
DIRAC.Oxford.uk	0.137%
DIRAC.Santiago.es	2.053%
DIRAC.ScotGrid.uk	5.078%
DIRAC.Zurich.ch	0.355%
LCG.BHAM-HEP.uk	0.341%
LCG.Barcelona.es	0.106%
LCG.Bari.it	0.010%
LCG.CERN.ch	12.22%
LCG.CNAF.it	4.097%
LCG.Cagliari.it	0.049%
LCG.Cambridge.uk	0.128%
LCG.Carleton.ca	0.146%
LCG.Catania.it	0.031%
LCG.FNAL.us	0.017%
LCG.FZK.de	3.375%
LCG.Ferrara.it	
LCG.IN2P3.fr	
LCG.ITEP.ru	
LCG.Imperial.uk	.188%
LCG.JINR	0.021%
LCG.[illegible]	1.077%
[illegible].pl	0.127%
[illegible].Lancashire.uk	0.515%
LCG.Legnaro.it	2.076%
LCG.Manchester.uk	0.473%
LCG.Milano.it	0.527%
LCG.Montreal.ca	0.041%
LCG.NCU.tw	0.408%
LCG.NIKHEF.nl	3.963%
LCG.Napoli.it	0.062%
LCG.Oxford.uk	0.791%
LCG.PIC.es	3.716%
LCG.Padova.it	0.099%
LCG.QMUL.uk	1.417%
LCG.RAL-HEP.uk	1.042%
LCG.RAL.uk	7.726%
LCG.RHUL.uk	0.463%
LCG.Roma.it	0.052%
LCG.SARA.nl	0.246%
LCG.SINP.ru	0.034%
LCG.Sheffield.uk	0.420%
LCG.Torino.it	0.722%
LCG.Toronto.ca	0.143%
LCG.Triumf.ca	0.317%
LCG.UCL-CCC.uk	0.795%
LCG.USC.es	0.193%
LCG.WEIZMANN.il	0.034%

20 DIRAC Sites

Used resources from non-LHCb countries e.g. Hungary produced ~2M events

43 LCG Sites (8 also DIRAC sites)

DIRAC-LCG : events share



- 50% of events were produced using LCG
- 20 DIRAC sites + 43 LCG sites have been used
- End of phase 1, 75% produced by LCG

DC'04 lessons

Lessons learnt: DIRAC

- The concept of the light, customizable and simple to deploy agents proved to be very effective
- Easy update procedure - propagate bug fixes quickly of DIRAC tools
- Applications software installation triggered by a running job
- LHCb Strategy successful:
  - Submitting “empty” DIRAC Agents to LCG has proven to be very flexible allowing a success rate above LCG alone.
- Most of the central services were running on the same machine
  - Too many processes, high loads
    ⇒Improve Server Availability
- Improve Error Handling and Reporting.

Lessons learnt: LCG

- After the DC04, we produced a report for LCG team where we highlighted some problems:
  - Upload/retrieval output of job, particular failed jobs
  - Tools to deal with bulk operations
  - CE status collection
  - Add some intelligence on CE
  - Tools to navigate through LOG info
  - HowTo manuals

New production

- Due to a software problem, we had to redo part of the production of DC04
- With the same settings we produced in one month 100 M Events without any special babysitting over the 2 week Christmas period.
- efficiency looked the same as first round but no detailed analysis

Conclusions for DC04 phase 1

- The Production Target was achieved:
  - 186 M Events in 424 CPU years.
  - ~ 50% on LCG Resources (75-80% at the last weeks).
  - 100 M Events in the new production
- LHCb Strategy successful:
  - Submitting “empty” DIRAC Agents to LCG has proven to be very flexible allowing a success rate above LCG alone.
- Big room for improvements, both on DIRAC and LCG
  - DIRAC needs to improve in the reliability of the Servers:
    - big step already during DC.
  - LCG needs improvement on the single job efficiency:
    - ~40% aborted jobs, ~10% did the work but failed from LCG viewpoint.
  - In both cases extra protections against external failures (network, unexpected shutdowns…) must be built in.
- Success due to dedicated support from LCG team and DIRAC Site Managers

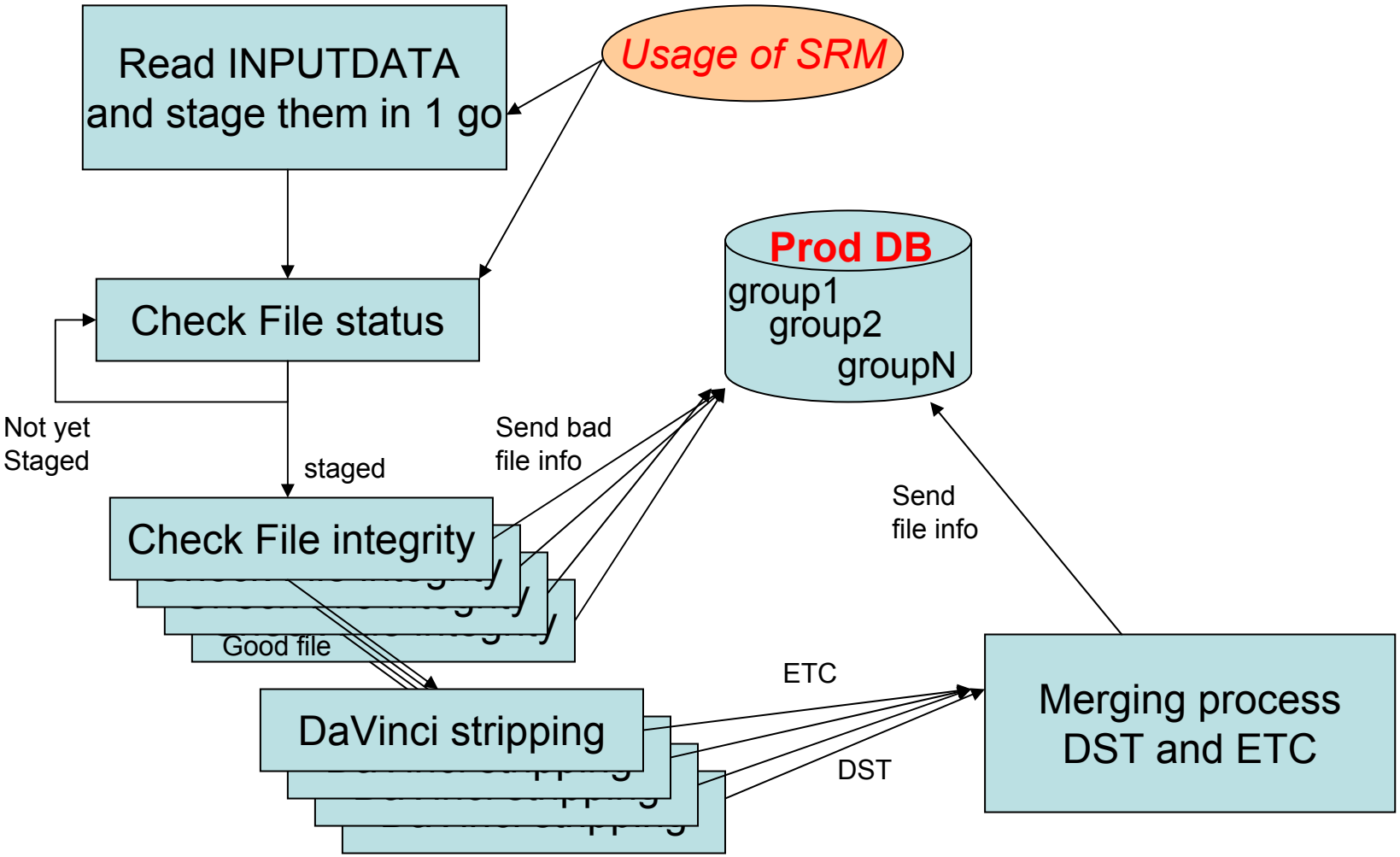
Results of the LHCb experiment Data Challenge 2004 (phase 2)

March 2005

Stripping

- Jobs with several INPUT files (between 40 and 80)
- Jobs sent to site where the data are placed
- Startup late due to LCG functionality not implemented (Storage Resource Management) and installed at some external sites for 1st time for LHCb
- 3 sites used CNAF, CERN and PIC based on CASTOR Mass Storage

Stripping job



Results of the stripping

- 20 M events processed which produced 460K stripped events.
- Not done with LCG.
  - Pressure from physics group
- Logic of these jobs tested and validate.

Production for the LHCb experiment Real Time Trigger Challenge

May 2005

RTTC production May/June 2005

- Goal: 150M minimum bias events to feed the online farm and test the software trigger chain
- Work-flow
  - Gauss v19r4 (3 steps)
  - Boole v8r4 (3 steps)
  - MergeRaw.RTTC (2 steps, Boole)
- RAW and DIGI are stored only at CERN
- 2+1 file catalogues in use
  - Book-keeping and Alien
  - LFC integrated in DIRAC
    - Running in production now in Lyon
    - Tests on other sites are ongoing
    - Not yet available on LCG

Available sites

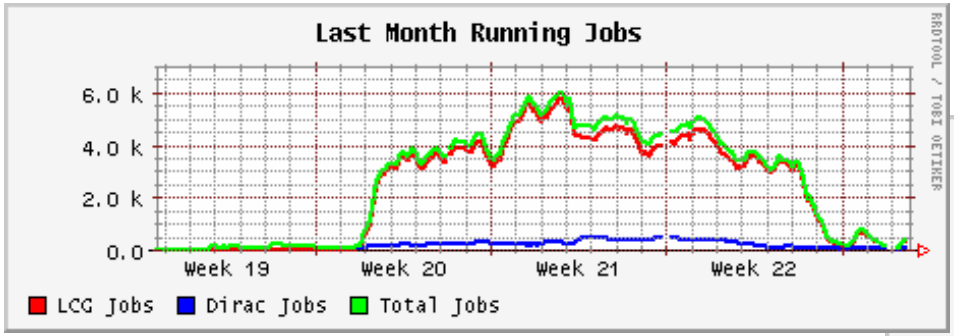
Sites in the production

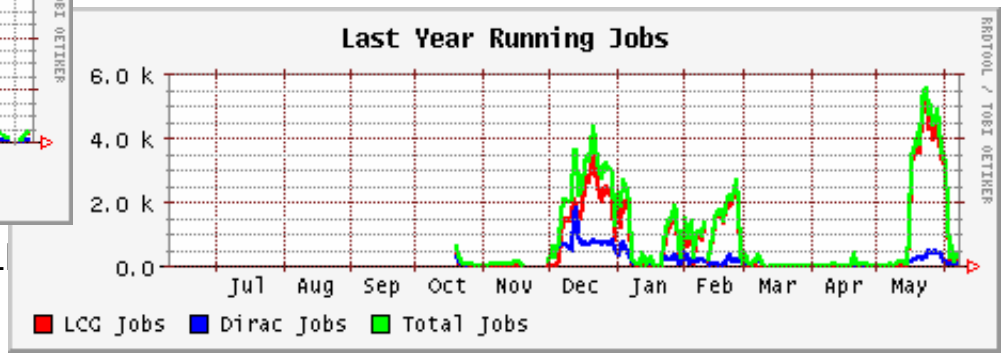
Countries	Dirac site	LCG site
Italy	1	13
UK	1	12
Spain	0	6
Grece	0	5
France	1	3
Russia	0	4
Swiss	2	2
Canada	0	3
Belgium	0	2
Netherland	0	2
Sweden	0	2
FKZ	0	1
Cyprius, Romany, Hungary, Brasil, USA	0	1+1+1 +1+1

- The RTTC production started since mid of May with a very fast startup
  - In one week almost all available sites in production mode
- More than 150M of events were produced in less than 3 weeks on 65 different sites
- 7 RBs used

Production Results

- 169M of events have been produced
- 3 production type
  - Prod. ID 846 (1500 events per job) → 10M
  - Prod. ID 847 (3000 events per job) → 84 M
  - Prod ID 848 (2550 events per job) → 75 M
  - The final output of events after L0 yes is 11 M now avalaible to the bookkeeping
- The production mean rate was of 10 M events per day with 4000 CPUs on avarage, with a peak of over 5000 CPUs





Shared data

Countries	Events produced
UK	60 M
Italy	42 M
Swiss	23 M
France	11 M
Netherland	10 M
Spain	8 M
Russia	3 M
Grece	2.5 M
Canada	2 M
Germany	0.3 M
Belgium	0.2
Sweden	0.2 M
Romany, Hungary, Brasil, USA	0.8 M

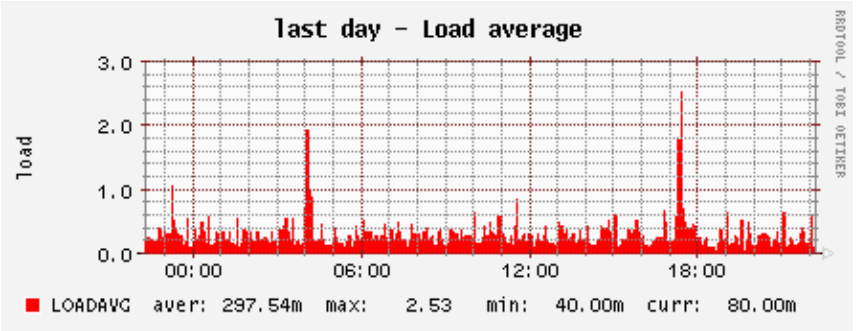
The data reported are preliminary (accuracy at 5%) as the accounting database is being populated now. In few days we'll have more precise numbers

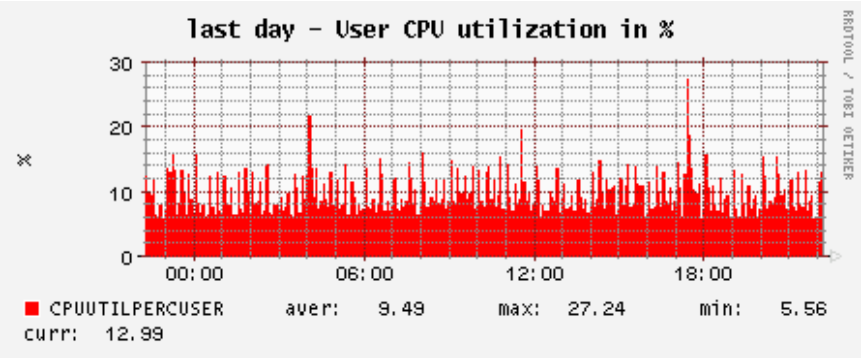
5% produced with plain DIRAC sites

95% produced with LCG sites

DIRAC performance

- Performance in the RTTC production
  - Over 5000 simultaneous jobs
    - Limited by the available resources
  - Far from the critical load on the DIRAC servers





Gridftp for file transfers

- Using gridftp tools for all the file transfers
  - Data
  - Log files
- All the DIRAC sites were obliged to use gridftp
  - Needs certificates for running local agents
  - Distribution of lcg_utils included with DIRAC
    - Platform independent distribution
    - Needs tools for automatic renewal of CA credentials (CRL's)

Problems met

- 30-40% of jobs aborted on the LCG, most of them at the start up time due to mis-configuration problems.
  - Problem seen also when installing LHCb software in the VO_LHCB_SW_DIR
  - Small or big site are affected in the same way.
- Without a pre-installed software a small fraction of LCG sites had the problem to install the software due to a combination of factor (http server overload, network occupancy, timeouts, etc)

Conclusions for the RTTC

- The RTTC production last just 20 days
- The startup was very fast
  - In few days almost all available sites were in production and the system was able to run with 4000 CPUs maintaining this rate for 3 weeks, with a peak of over 5000 CPUs. Sensible improvement with respect to DC04 data challenge.
- 168 M events produced (11 M events as final output after L0)

Results of the LHCb experiment Data Challenge 2004 (phase 2 next round)

June 2005

Using SRM

- Stripping activity necessitate the usage of advanced features in the storage handling
  - Staging files before the job starts
  - Checking file availability before the start of the application
- Special Module introduced in DIRAC to deal with SRM based storage
  - Command lines based on the GFAL libraries
  - Still needed direct Castor staging commands to ensure the deletion of the files
- Need more work to incorporate SRM based storage into the DIRAC framework

Stripping Status

- SRM is functional on the three sites CNAF, CERN, PIC
- The stripping test jobs have been submitted and run successfully at CERN, PIC and CNAF
- Some wrong entries on the file catalogue found
  - This is problematic to SRM CASTOR as these files will cause the whole SRM request to fail BUT without notification but files are still staged!
- Output will be sent to the three Tier1s.
- There are not any major problems to start the stripping of the 100M events
  - Some pending problems to be solved, but it should start now!

Summary

- Monte Carlo production can be run without any major problem
- Pilot agent approach protects us from inherent problems on Grid
  - generally once job starts it runs to completion
- LCG currently providing ~95% of our resource requirements for production
- stripping of data (using SRM) about to be launched