LHCb Computing Experience with First Data

Marco Adinolfi - University of Bristol

Friday, 23 July 2010

The LHCb Experiment
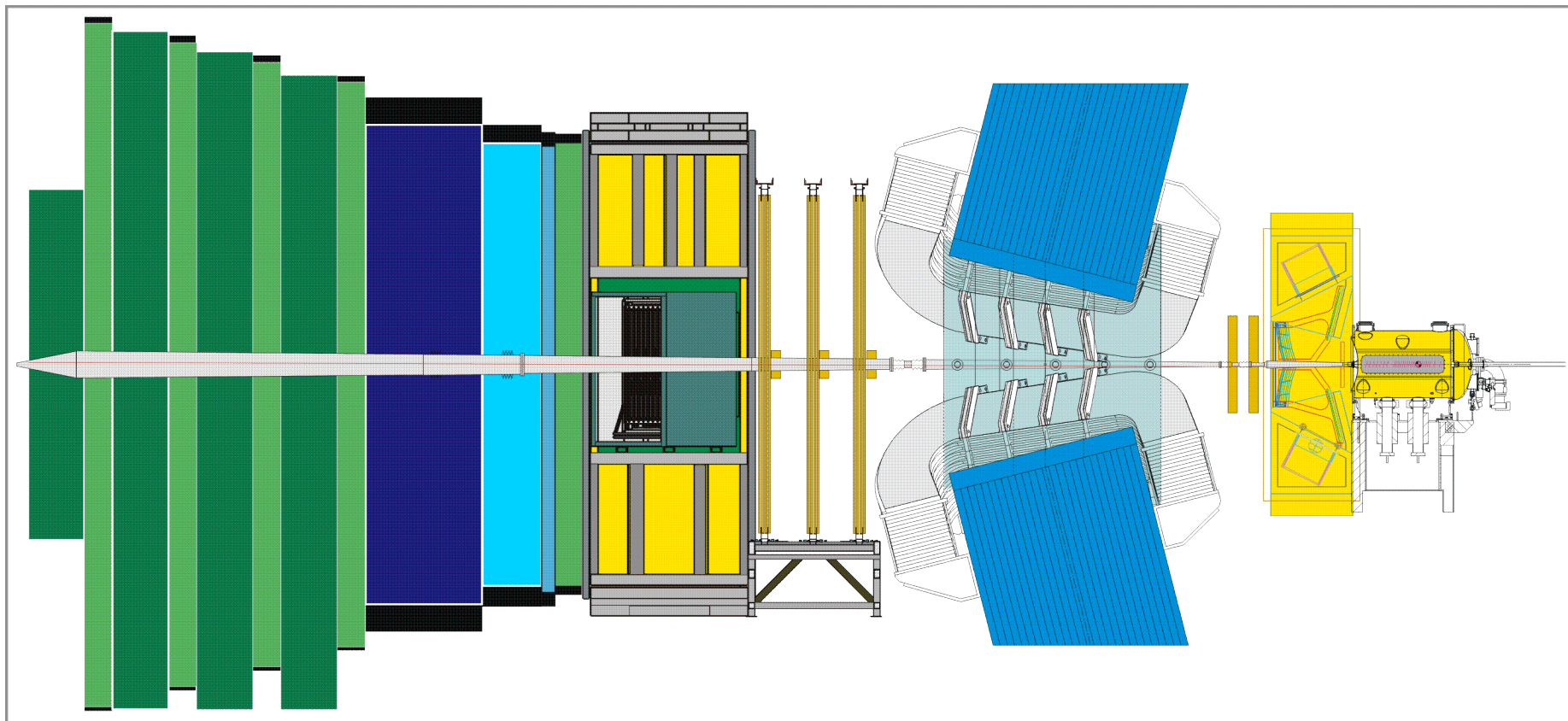
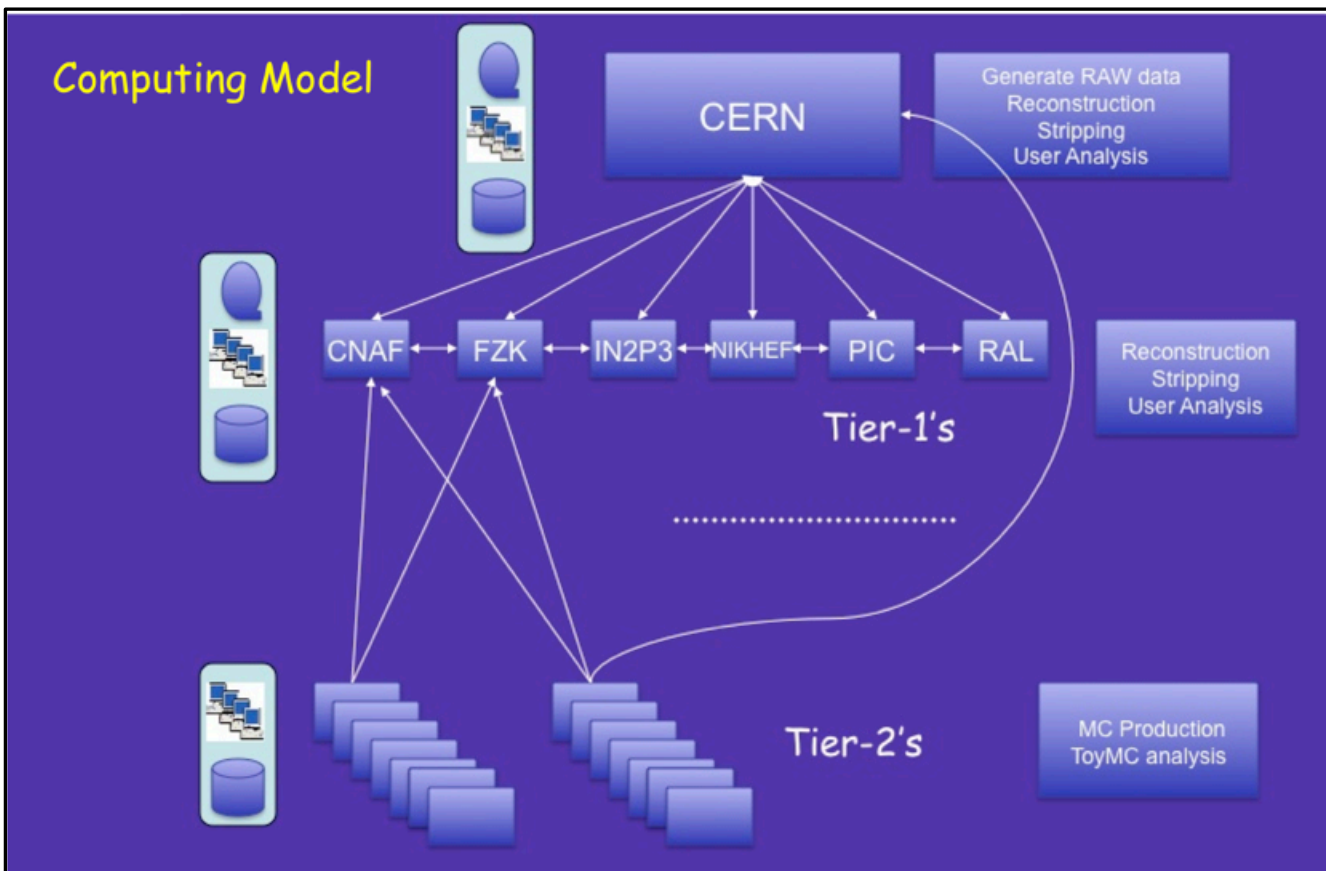Muon   H/ECAL   RICH-2   Tracker   Magnet   RICH-1   VELO

- LHC experiment dedicated to the search of rare B decays and New Physics.
- As of July LHCb is running at nominal speed with a trigger rate of 2 KHz, as in computing model.

M. Adinolfi - University of Bristol   2   ICHEP 2010 - Paris - 24 July 2010

Friday, 23 July 2010

The LHCb Computing Model

- User analysis supported at CERN and at the 6 Tier-1 centres.

- Tier-2 centres used for Monte Carlo production.

- Plan to use the LHCb Online farm for reprocessing.
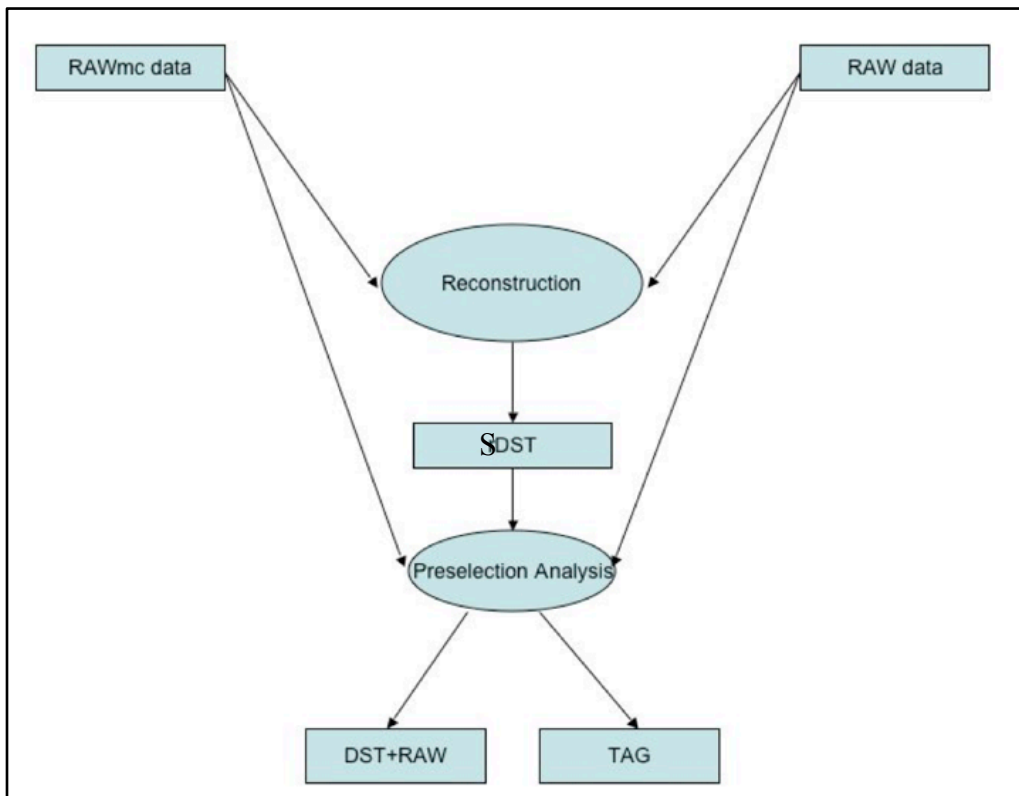
- RAW Data is reconstructed
  ‣ Calorimeter energy clusters
  ‣ Particle ID
  ‣ Tracks...

- At reconstruction only enough information is stored to allow a physics pre-selection to run at a later stage: stripping DST (SDST)
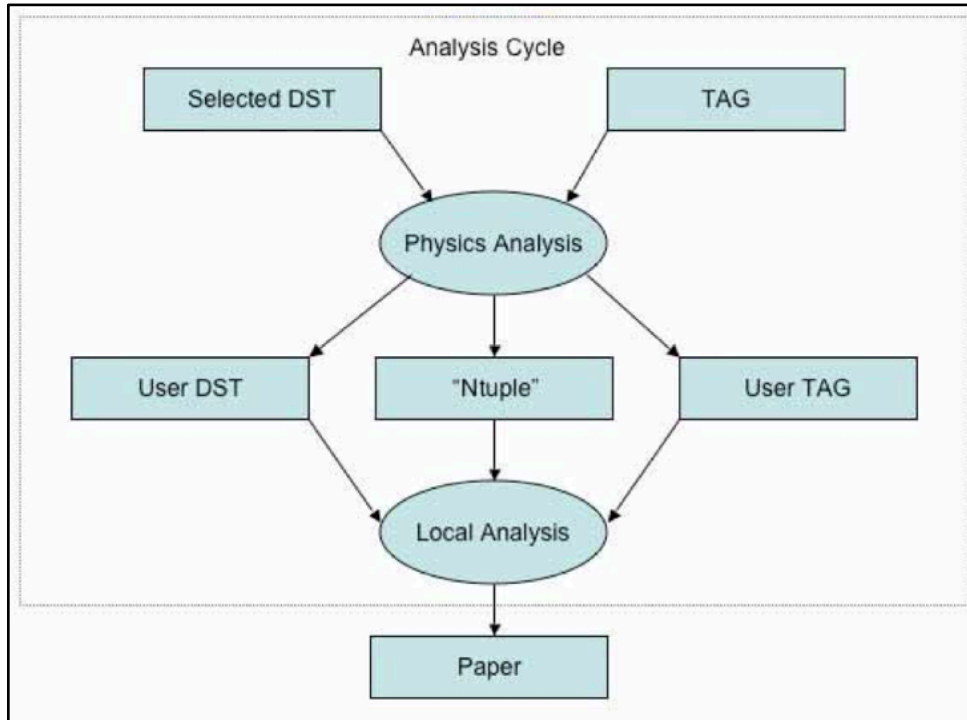
- SDST stored separately from RAW data.

Friday, 23 July 2010

Data Flow - Stripping
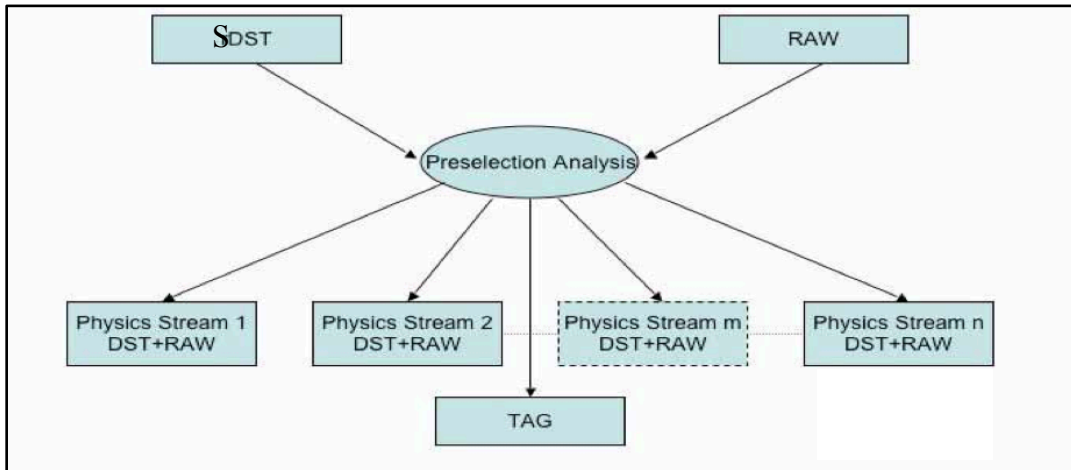
Computing Model foresees in 1 year:
▸ 2 reconstruction
  ● quasi real time
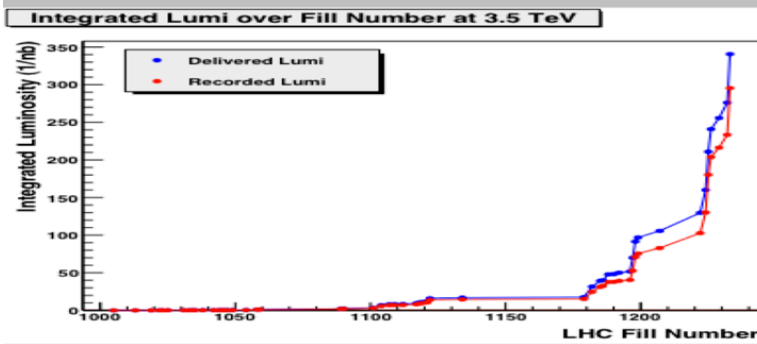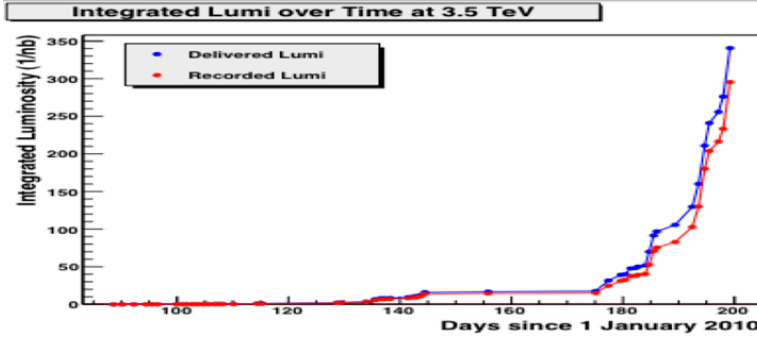 ● during LHC shutdown
▸ 4 stripping

Very first data already reprocessed 7 times.

- Data reduction factor 10.
- SDST analysed during production.
- Event streams saved for further analysis.
  ▸ Currently 11 streams.
- Selection algorithms developed by physics working groups.
  ▸ Currently over 250 algorithms.
- Stream output includes the event RAW data.
- Event Tag Collection, containing metadata, created to allow quick access to data.
- Data only accepted after a thorough quality check based on histograms produced during reconstruction.

Friday, 23 July 2010

Data Flow - Analysis

Analysis Cycle

- Selected DST
- TAG
- Physics Analysis
- User DST
- "Ntuple"
- User TAG
- Local Analysis
- Paper

- User physics analysis performed on the stripped data.

- Output of the stripping is self contained, i.e. no need to navigate through files.

- Analysis generates semi-private data: ntuple and/or personal DST.

- Semi-private data can be Grid accessible to enable remote collaboration.

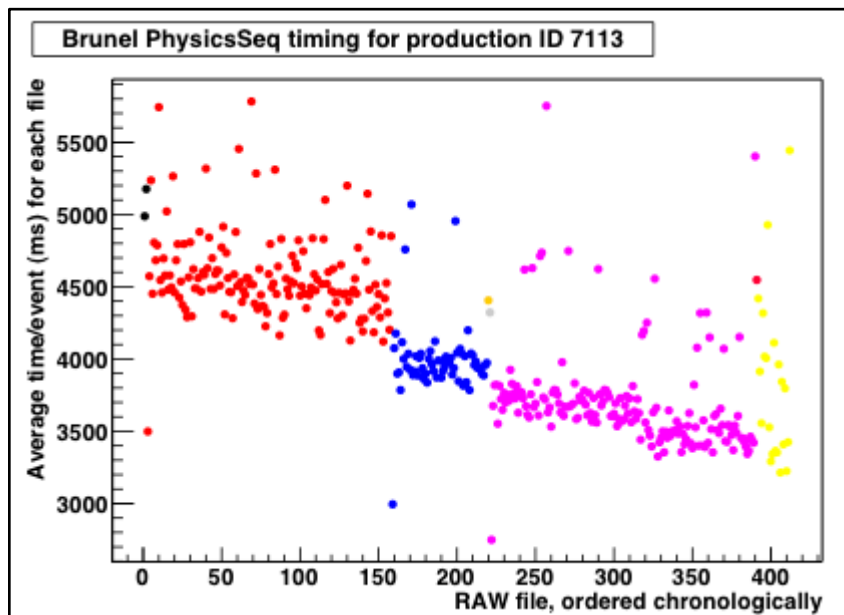- Ganga, a Grid User Interface, developed and fully functional.

Friday, 23 July 2010

2010 Data

Integrated Lumi over Time at 3.5 TeV

Integrated Lumi over Fill Number at 3.5 TeV

RAW Data event size ~ 60 kB   CM ~ 35 kB

LHCb now running with higher number of Primary Vertices than design.
1PV 50.6 kB - 2PV 66.4 kb - 30PV 80.5 kB

SDST   event size ~ 45 kB
DST(b)   event size ~ 170 kB
DST(MB)   event size ~ 42 kB

- As of 19 July LHCb has collected ~295 nb-1 of ~339 nb-1 delivered.
  ‣ 33000 files
  ‣ 900M events
  ‣ 42 TB

- About 2% rejected by Data Quality checks.

- 2 copies of RAW data: 1 at CERN and 1 at a Tier-1 centre.

- All files of the same run stored at the same Tier-1 centre.

- Stream DST of a run merged once the run has been reconstructed and stripped.

Friday, 23 July 2010

Reconstruction

Brunel PhysicsSeq timing for production ID 7113

- Average reconstruction time per event currently 3500 ms.

- Reconstruction time dependant on pile-up.

- LHCb running at an higher pile-up rate than the design.

- RAW data file size adapted to maximise job efficiency.

Data set processing
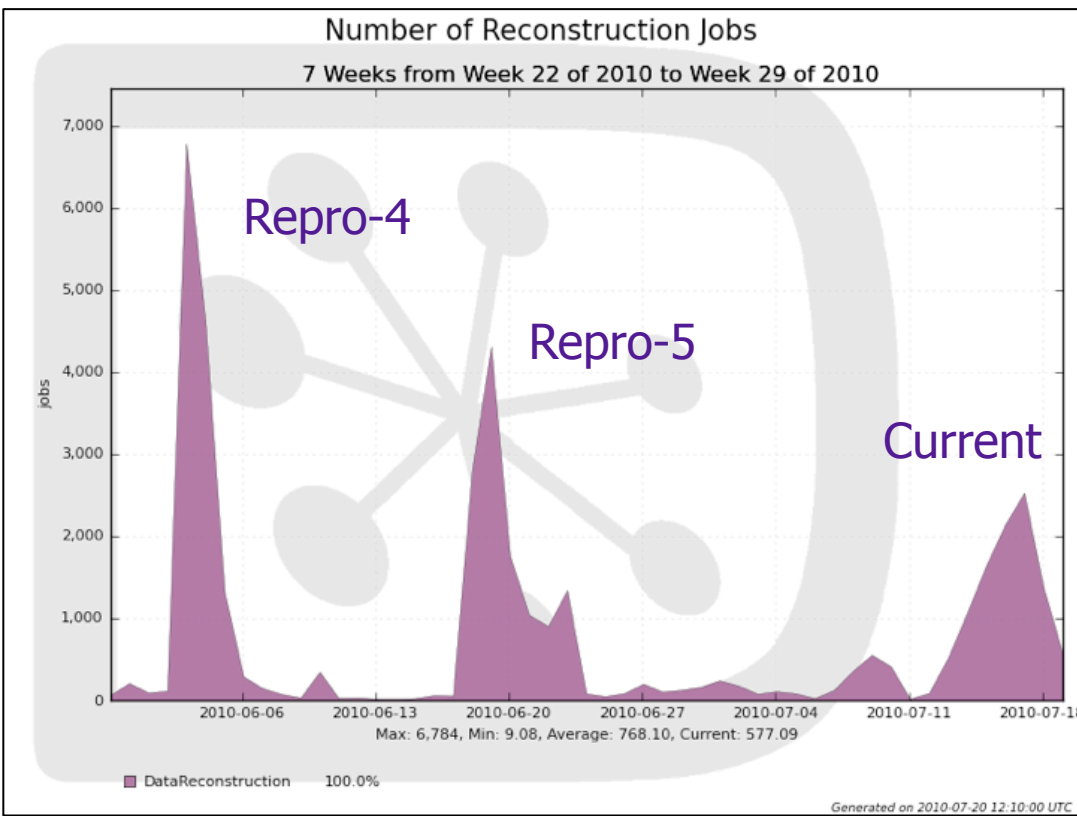
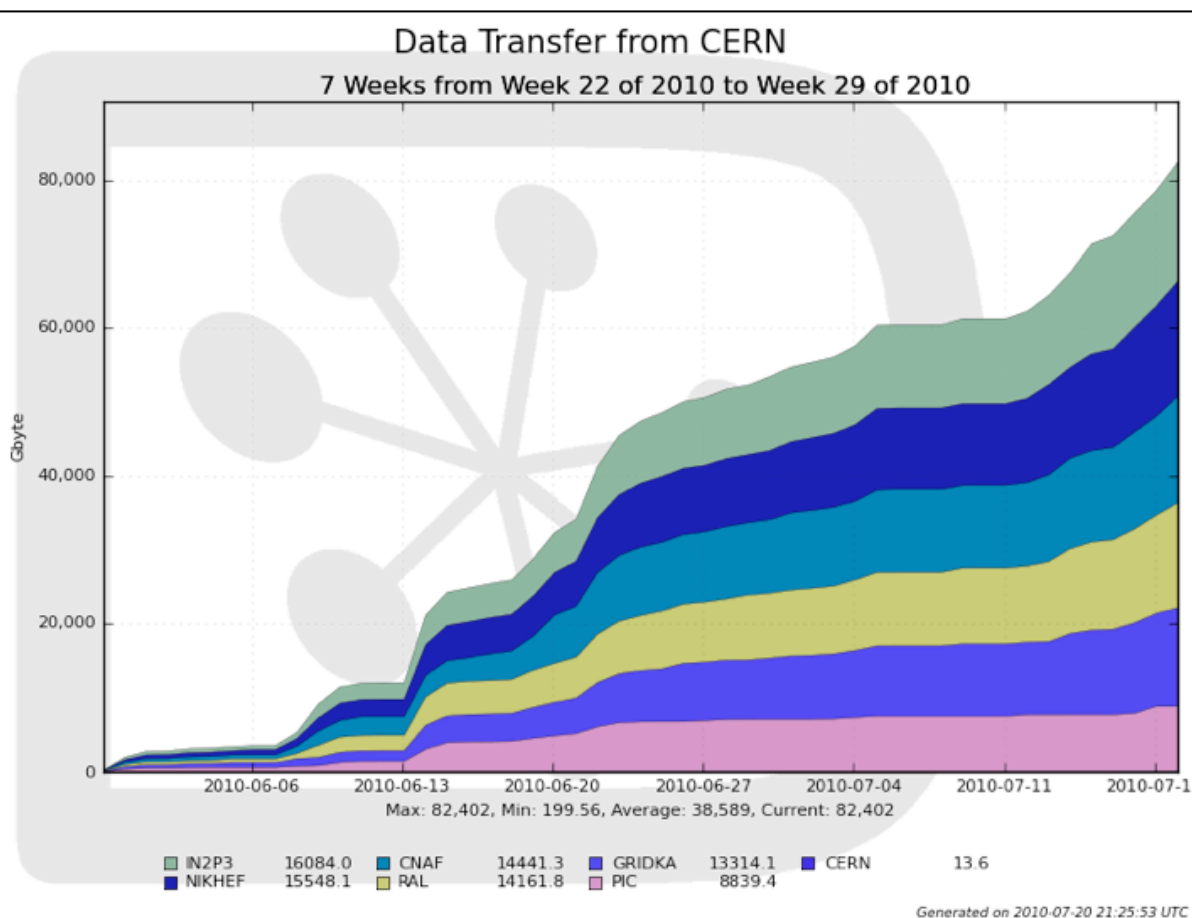- Data collected up to early June (~14nb-1) processed several times as new alignment and improved reconstruction are made available.

- 90% of the datasets is reprocessed in abut 3 days.

- Now that nominal conditions have been reached such frequent reprocessing are no longer possible.

Friday, 23 July 2010
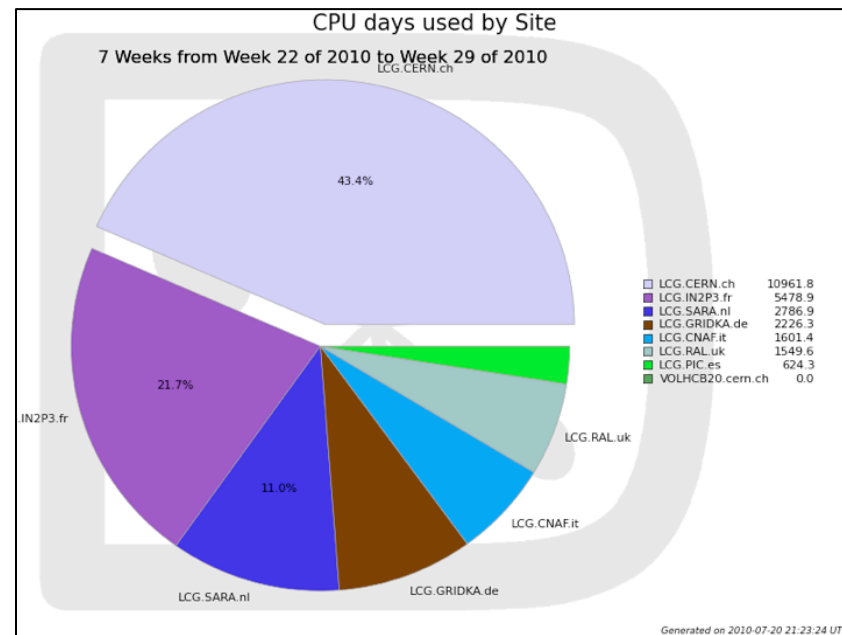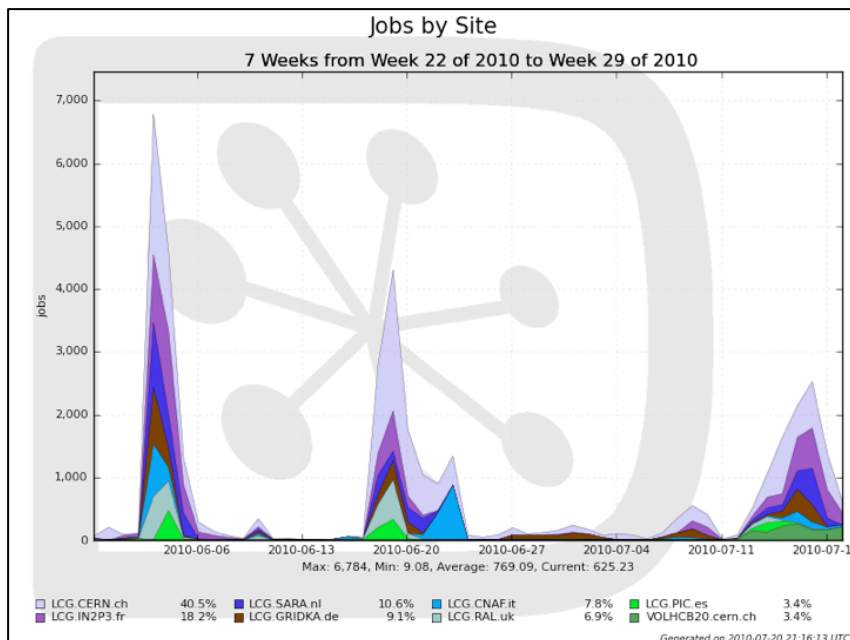
- Currently processing 40% of the jobs at CERN corresponding to 43% of the CPU.
- Roughly corresponds to the Computing Model.

Friday, 23 July 2010

Data Transfer

Data Transfer from CERN
7 Weeks from Week 22 of 2010 to Week 29 of 2010

Max: 82,402, Min: 199.56, Average: 38,589, Current: 82,402

☐ IN2P3	16084.0	☐ CNAF	14441.3	☐ GRIDKA	13314.1	☐ CERN	13.6
☐ NIKHEF	15548.1	☐ RAL	14161.8	☐ PIC	8839.4		

Generated on 2010-07-20 21:25:53 UTC

- RAW Data is replicated to one of the Tier-1

- Albeit some initial problem, data is now successfully transferred on regular basis.

Friday, 23 July 2010
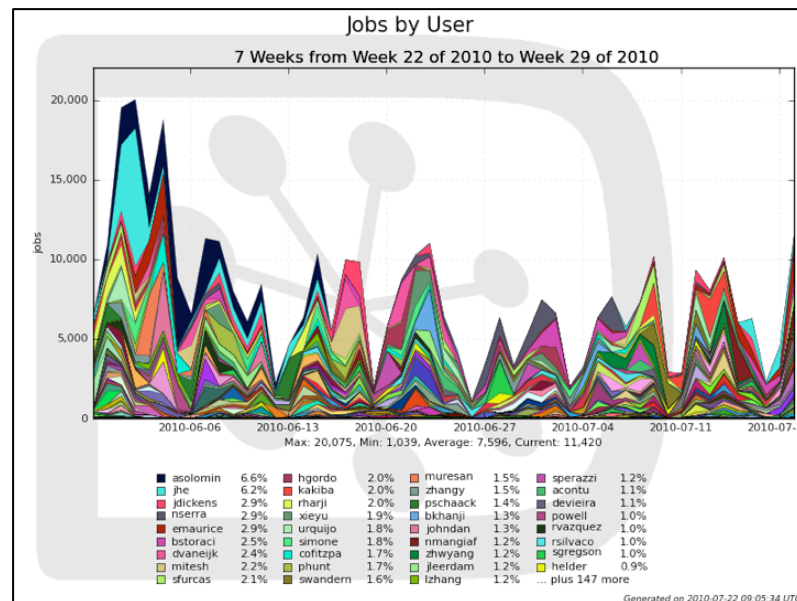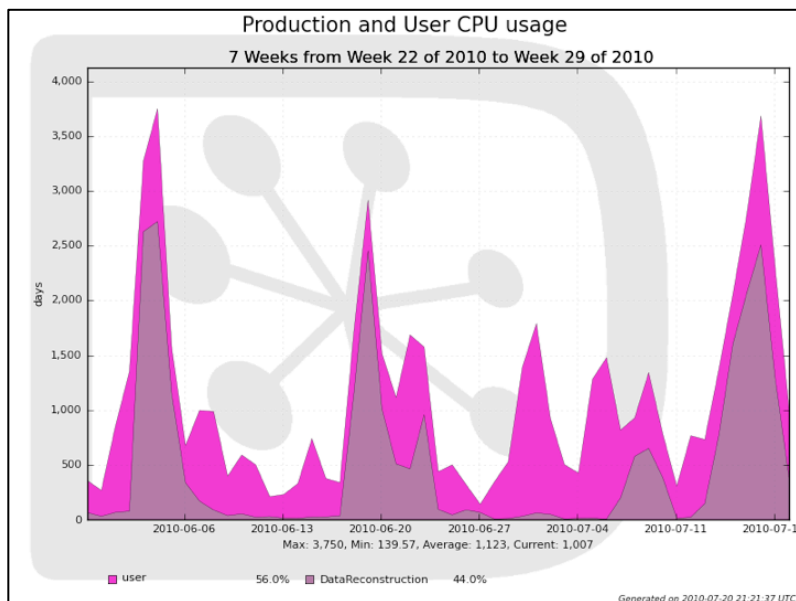
User Analysis Jobs

- CPU at Tier-1 centres roughly distributed 60/40 between user and reconstruction jobs.

- Over 200 LHCb users have submitted analysis jobs over the Grid.

- As many as 30k jobs in a day.

Summary and Conclusions

- LHCb has developed robust, efficient and flexible Computing Model and software framework to process and analyse its data.

- The Model flexibility has proved invaluable when modification have been required to adapt to real data.

- The RAW data is rapidly and successfully transferred from CERN to the Tier-1.

- Fully qualified new data is reconstructed, stripped merged and made available for analysis to the users within a few days.

- Intense analysis program fully ongoing using the Grid resources.

Friday, 23 July 2010