LHCb Computing Experience with First Data

Marco Adinolfi - University of Bristol
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.
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.
Data Flow

- 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.
Data Flow - Stripping

- 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.

Computing Model foresees in 1 year:
- 2 reconstruction
  - quasi real time
  - during LHC shutdown
- 4 stripping

Very first data already reprocessed 7 times.
Data Flow - Analysis

- 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.
2010 Data

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.

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
Reconstruction

- 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 about 3 days.
- Now that nominal conditions have been reached such frequent reprocessing are no longer possible.
• Currently processing 40% of the jobs at CERN corresponding to 43% of the CPU.
• Roughly corresponds to the Computing Model.
Data Transfer

• RAW Data is replicated to one of the Tier-1
• Albeit some initial problem, data is now successfully transferred on regular basis.
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.