LHCb at EPFL: computing report

Stefano Villa

Laboratoire de Physique des Hautes Energies, EPFL

CHIPP LHC Computing and Analysis workshop

Manno, August 25-26, 2005

Contents

- Present and recent activities in computing and software in Lausanne.
- The LHCb computing model for Tier-2 centres.
- Estimate of resources needed by EPFL at Tier-2 and Tier-3.

DISCLAIMER: all numbers are preliminary and based on extrapolations: allow for **big uncertainties**...

LHCb software and computing at EPFL

- **Level 1 and High Level trigger**:

  - Development, implementation and maintenance of L1 trigger.

  - Integration of L1 software in online monitoring system (PVSS).

  - Some exclusive HLT selections.

- **Simulation, reconstruction and analysis software**:

  - Description of Inner Tracker geometry and material in the simulation.

  - Algorithm for evaluating efficiencies and purities of trigger and selection.

  - Flavour tagging algorithms and monitoring.

  - Improvement of online tracking (VELO-TT matching).

  - Jet reconstruction.

LHCb software and computing at EPFL

- **Physics analyses**

  - $B_s \to \eta_c \phi$, $B_s \to J/\psi \, \eta$, $B_s \to J/\psi \, \eta'(\eta\pi\pi)$ selections and sensitivity studies (extraction of ϕ_s).

  - $B_s \to D_s \pi$ selection and first look at performance at high lumi. (measurement of Δm_s).

  - Photon polarization in $\Lambda_b \to \Lambda \gamma$

  - Light Higgs: selection and sensitivity studies.

- **Distributed data analysis**

  - Limited experience in our group: one user tried GANGA to submit jobs to CERN LSF batch system.

User's analysis: LHCb's model

- LHCb's computing model: Tier-2 centres are primarily devoted to MC production (LHCb TDR 11)

  MC production, LHCb requirements for Tier-2:

  (stays roughly constant in time)

	All Tier-2s	Manno (~10%)
CPU (MSI2k)	7.65	0.765
Storage (TB)	23	2.3

User's analysis: LHCb's model

- User's analysis at Tier-2 is possible if, **on top of what is needed for MC production**, there is:

  - At least ~0.2 PB of disk storage (1 copy of latest DST).

  - Some minimal CPU (see later our needs).

  - Enough network, to get DSTs every time they are produced: ~50 MB/s during 1 month of stripping, 4 times per year.

- At Tier-2, user analysis consists of batch jobs: interactive analysis (Ntuples...) must be done elsewhere (Tier-3 or local).

Estimate of needed resources: people

- Today's composition of the group (LHCb + other projects)

  - Faculty: 3 profs, 1 prof. NFS, 1 MER, 1 senior researcher

  - 5 postdocs

  - 15 PhD students

- Assuming the same size of the group and phasing out of other activities by 2008, ~ 15 FTE physicists doing analysis ("active users").

Estimate of needed resources: assumptions

- Assume all MC production done at Tier-2, centrally coordinated (**not included in the following**).

- EPFL has/will have a few Linux clusters; not yet clear how much grid integration we can expect. It might work as (be part of) Tier-3.

- Assume that a big part ($\sim 50\%$) of analysis jobs will be sent to Tier-2. The rest will run at Tier-1 centres or CERN.

Estimate of needed resources: numbers

- TDR estimates of computing needs per active user:

  - CPU: 5.6 kSI2k

  - Storage: 1.4 TB

- Correct for efficiencies: 60% "chaotic" CPU usage and 70% of disk usage

  - CPU: 9.3 kSI2k

  - Storage: 2.0 TB

- Assume "slow" startup: only 10% of analysis done at Tier-2 in 2008 (the rest at Tier-1).

Estimate of needed resources at Tier2: evolution

- Ramp up to full (50% of total) use by 2009.

- Assume no change in manpower; CPU usage and disk storage should grow linearly with data:

Resource requirements at Tier-2 for EPFL:

	2008	2009	2010
CPU (kSI2k)	14	140	210
Storage (TB)	3	30	45

Estimate of needed resources at Tier3

- Assume less need of CPU for "local" analysis than for "batch" analysis (1/2). Time evolution should also be slower (1/2).

- Storage should be smaller (1/2), most files can stay at Tier-2.

Resource requirements at Tier-3 for EPFL:

	2008	2009	2010
CPU (kSI2k)	70	105	140
Storage (TB)	15	30	45

Final considerations

- It is very difficult to formulate reliable estimates on how people will run their analysis in 3-5 years.

- At the end users will actually use whatever works better/faster, regardless of what we "plan" today.

- But... **we have to plan!**