Computing Report CMS-ETHZ

Urs Langenegger
(ETH Zürich)

2005/08/26

CHIPP Computing Workshop

- Introduction
- ETHZ
  - Analyses/requirements
- User perspective
  - no pain, no gain

Many thanks to Gian Luca!

CMS Computing Model

- CMS computing TDR: CERN-LHCC-2005-023

  (http://doc.cern.ch//archive/electronic/cern/preprints/lhcc/public/lhcc-2005-023.pdf)

- Computing model of CMS
  - Tier-1: Skimming of AOD $\longrightarrow$ transfer skims to Tier-2
    Reconstruction/reprocessing/...
  - Tier-2: Run ORCA on AOD, produce 'ntuples'
    Central MC production
  - Tier-3: Analyse ntuples, make plots
    Edit/compile/run/debug ORCA analysis code

- Users of Tier-2:
  - Scheduled MC production
  - Analysis group(s)
  - 'Local' community (+ opportunistic (parasitic?) LCG usage)

- Some Numbers:
  - 50kB AOD event size
  - 1.5×10^9 raw events recorded by CMS, skim down to ca 1%
  - 0.25 kSI2k sec analysis time per event

Analyses and Requirements/Year

- ECAL: Pauss / Dissertori

Analysis	Events (data + MC)	Disk [TB]	CPU [kSI2k]
$H \to WW \to \ell\ell\bar{\nu}\bar{\nu}$	$10^6 + 10^6$	0.1	5
incl. leptonic W/Z	$10^6 + 10^6$	0.1	5
incl. γ	$10^5 + 10^6$	0.1	5
topol. searches	$10^6 + 10^7$	1.0	20

- Pixel/tracker: Eichler / Langenegger(-2008)

Analysis	Events (data + MC)	Disk [TB]	CPU [kSI2k]
...	$10^8 + 10^8$ (??)	10.	60
$H \to b\bar{b}$	$3 \times 10^7 + 3 \times 10^7$	3.	20
$B^0_{s(d)} \to \mu^+\mu^-$	$3 \times 10^7 + 3 \times 10^7$	3.	20

- CPU
  - ▷ Process all events of one year within one week? (luxury, imvho)
  - ▷ for 'big' samples $\longrightarrow$ ca. 3×10^7 s
  - ▷ $\longrightarrow$ 50 CPU
  - ▷ $\longrightarrow$ time to think $\longrightarrow$ 10-20 CPU

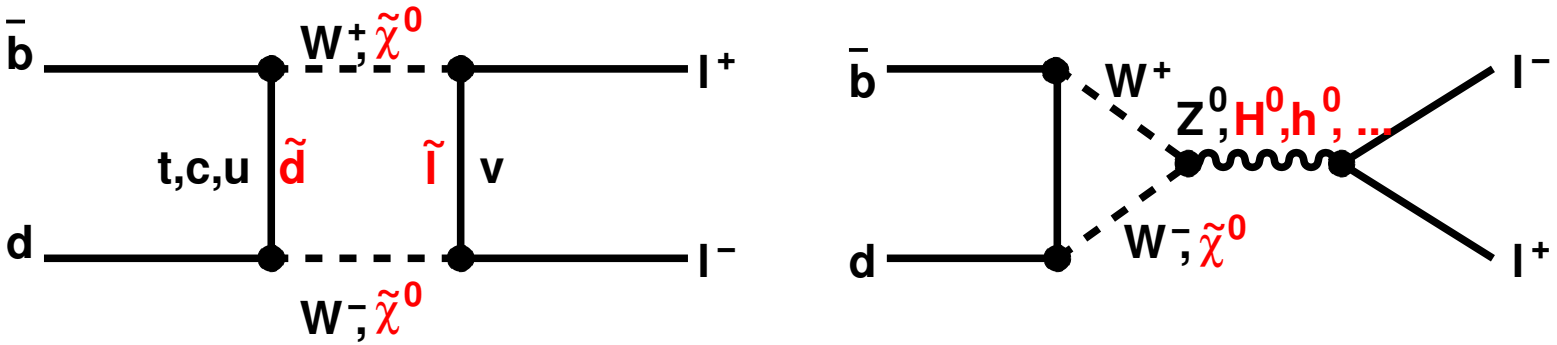
Answers to Questions

- Strategy
  - T3 (own linux cluster) at ETH
  - estimate CPU and disk T2:T3 $\sim$ 80:20
- Active users in group? ca 10
  - 'power' users, PhD students running jobs, running jobs, . . .
- CTDR estimates for users:
  - 15 kSI2k, 3.5TB

  Seem OK
- Estimate of resource requirements
  - $0.8 \times 10 \times 15$ kSI2k 120 kSI2k
  - $0.8 \times 10 \times 3.5$ TB 25 TB
- Time extrapolation
  - Users will stay constant
  - Resources: Linear scaling with integrated lumi (constant trigger rate)
- Accuracy of these numbers? Factor 2–3. Maybe.

Example: One simple analysis

- SUSY parameters with the first LHC data: $B^0_{s(d)} \to \mu^+\mu^-$
  - $\tan\beta$



- This analysis chosen, because
  - straightforward and simple
  - related to pixel detector
  - learn grid-related issues
  - . . . physics potential (hopefully)

User Perspective

- Disclaimer:

  I have no interest in computing.
  All I want is physics results.

- So far ca. 1000 grid jobs in LCG-2
  - usually batches of 100 jobs
  - during the last two months
  - no grid tools, private (small) PERL scripts

  cf BABAR: 'production' meant 10k–100k jobs (CPU and I/O limited)

- Generator MC
  - No CMS s/w needed
  - Statically linked F77 executable (runs everywhere)

- Private AOD production (DST)
  - Full chain of CMS sw required
  - No pile-up included (yet)

- No AOD analysis yet over grid

Setup

- At ETH: 'T3' (dual CPU server with CMS s/w)
  - edit/compile/validate AOD analysis code
  - AOD analysis
- User interface (job submission) for grid
  - lxplus
  - ui-lcg.projects.cscs.ch
- Generator jobs
  - Copy executable from 'storage element'
  - Run exectubale
  - Store output .ntpl to 'storage element'
  - Runtime: some hours
- DST production
  - Copy .ntpl from 'storage element'
  - Run sim/digi/reco/dst-writing (local) executables
  - Runtime: 15–24 hours for full chain with 1k events

Problems and Mistakes

- Mistakes
  - two weeks until 'all scripts' worked
  - expiry of proxies: grid-proxy, then my-proxy . . .
  - overloaded resource broker (.ntpl in outputsandbox)
- Problems
  - Middleware not robust
  - H/w of resource broker at CERN $\longrightarrow$ there are alternatives
  - H/w of monitoring server at CERN $\longrightarrow$ there are alternatives
  - Missing CMS software installations $\longrightarrow$ mail Gian Luca
  - Bugs in CMS software $\longrightarrow$ savannah, better: mail Gian Luca
  - Submission time (ca 10 seconds/jobs)
- Benefits
  - Potentially many CPUs
  - Patience
- But: BABAR (batch/OBJY at SLAC) in 2000 was worse

Some jobs run, and some don't . . .

Batch	Job	Successful	WN	SE-retrieval	SE-storage
202xx	GEN/phoenix	100			
204xx	GEN/phoenix	100			
205xx	GEN/phoenix	0			100
206xx	GEN/grid	14	83		3
211xx	GEN/grid	12	71	17	
212xx	GEN/grid	0	98	2	
204xx	DST/phoenix	10	16		74

- A selection of errors:

  error while loading shared libraries: libg2c.so.0: cannot open shared object file: No such file or directory

  370 pi+ 1 211 163 -0.364 0.00 Killed

  4173 -1 Killed

  CE Accesspoint not found

  /storage/exp_soft/cms/cmsset_default.csh: No such file or directory.

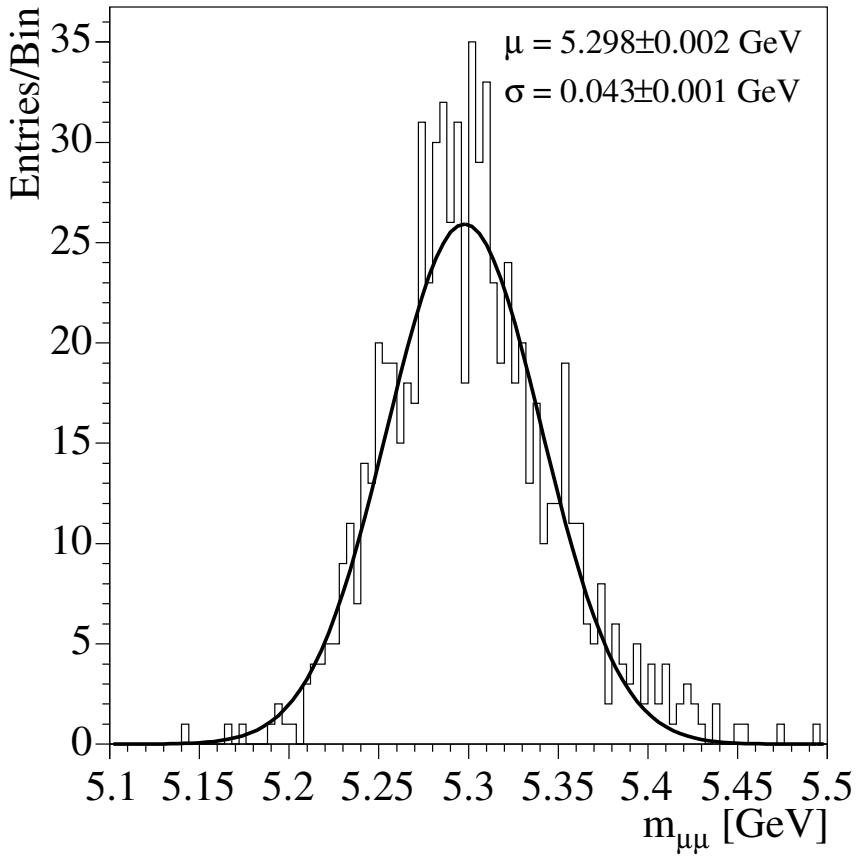
  ./cmsset_default.csh: No such file or directory.

  0: Event not found.

- Many small details need to be right at many places
  - Well possible that some of these errors due to my stupidity

. . . The Histogram

- Full simulation for $B_s^0 \to \mu^+\mu^-$
  - DST production on phoenix
  - DST analysis at ETH



(Don't ask about the central value)

Conclusion: Hardware! ('More, more, more, . . . ')

- The message:
  - No CH physics at LHC without CH computing (away from CERN)
- LHC without CH Tier-2
  - Not an option
- Time scale
  - Ready and ramped-up when first data comes in
- Independent of grid/middleware/. . .
- Hardware
  - diskspace and CPU and the best I/O in between