# Event Production in CMS

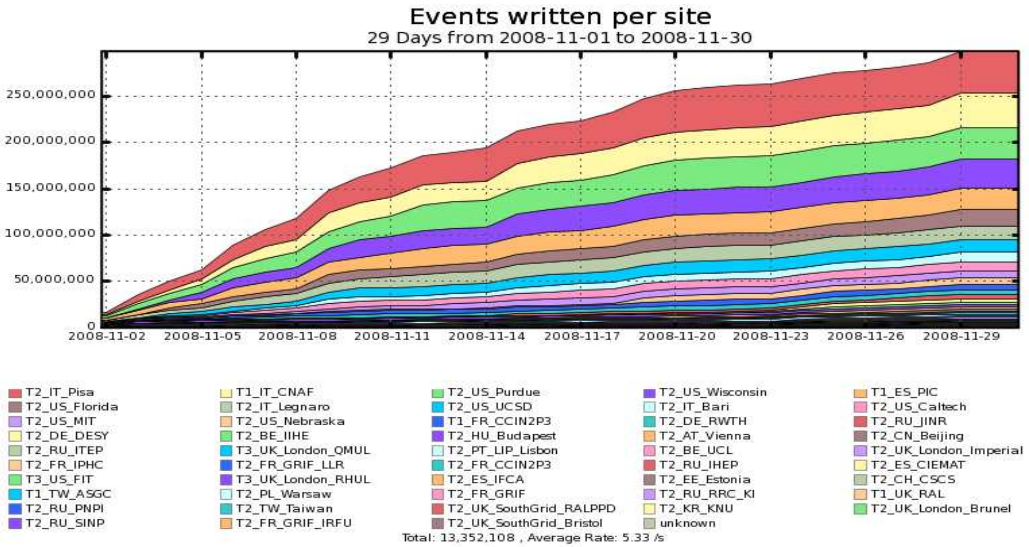
Guillelmo Gómez-Ceballos
(MIT)

December 15th, 2008

**CMS Advanced Monte Carlo Use and Tuning Strategies Workshop**

# Official MC Production (I)

- Central, T2/T1 distributed CMS production, not talking about standalone private production
- Deployment strategy: software must be available at all sites in a -unique- way (gridpacks included)
- Usual MC production:
  - run at the Tier-2 centers
  - any additional reprocessing task run at the Tier-1 centers
- Production rate: ~250M evts/month (50%/50% sim/reco)



# Official MC Production (II)

- Usual procedure:
  - Generator config file is prepared *and* validated by the requester
  - Config file is committed to CVS
  - Request is approved by the generator/Physics (or DPG) coordinators and sent to dataOps
  - Production starts
  - For large samples, aim having a fast validation with the first chunck of events
- Steps: always (1) and (2), sometimes (3) and/or (4)

  (1) GEN-SIM-DIGI-HLT-RAW: from generation to RAW format, including HLT

  (2) RECO: reconstruction, “like” data

  (3) AOD: reducing the output size to AOD objects only

  (4) Skims, AlcaReco, PAT-tuples, etc.

# Generator Code

- Must have a responsible person to keep it up-to date to include it in the official MC production chain
- Production wise, 2 types:
  - (1) Fully integrated in CMSSW
  - (2) Partially integrated in CMSSW
- (1) Fully integrated in CMSSW:
  - event generation happens within CMSSW
  - generators: Pythia, Herwig, MCatNLO, PomWig...
  - easy to deal for computing
- (2) Partially integrated in CMSSW: can't be fully within CMSSW
  - event generation, LHE files, happens outside CMSSW
  - generators: Alpgen, Madgraph...
  - LHE files used as input for cmsRun job
  - producing these files is tricky

# A Word About LHE (Ascii) Files

More details to be given in another talk later

- Output of the generator
- LHE file: standard file format to store process and event information, primarily output from parton-level event generators for further use by general-purpose ones
- Notice Alpgen output for instance has a slightly different format, but again an ascii file
- CMSSW jobs can’t be run on those files without some interface

# Production of LHE (Ascii) Files

- Present option: CmsGen
  - python wrapper used to run the generator and hide all the details of what the actual generator is doing
  - integrated into the production framework (also possible to use it interactively)
  - used for Alpgen (unweighted events file) and Madgraph (LHE file) generation
- (Near) future option: MCDB
  - LHE files will be uploaded to MCDB by user/generator experts
  - cmsRun will pick them, produce GEN files and register in DBS: this step must be done at CERN
  - SIM...RECO steps will be run as a processing task, GEN files as input
  - details to be finalized

# Caveats

- Validation of workflows:
  - situation has improved very much, but it is a complicated task
  - working config file is a must, getting better on that
  - full validation of the final output is a real issue
  - need to be able to validate via CRAB
- CmsGen:
  - gives a lot of freedom, tests can be done very quickly
  - but some info must be in a release for production
  - need to be deployed at all sites before starting the generation as any other CMS software