



Experience with CMS Offline and Computing in LHC Runs 2010-2011.

Jean-Roch Vlimant for the **CMS Collaboration**.
UC Santa Barbara, Fermilab, CERN.

EPS-HEP 2011, Grenoble.



CMS Computing World Wide

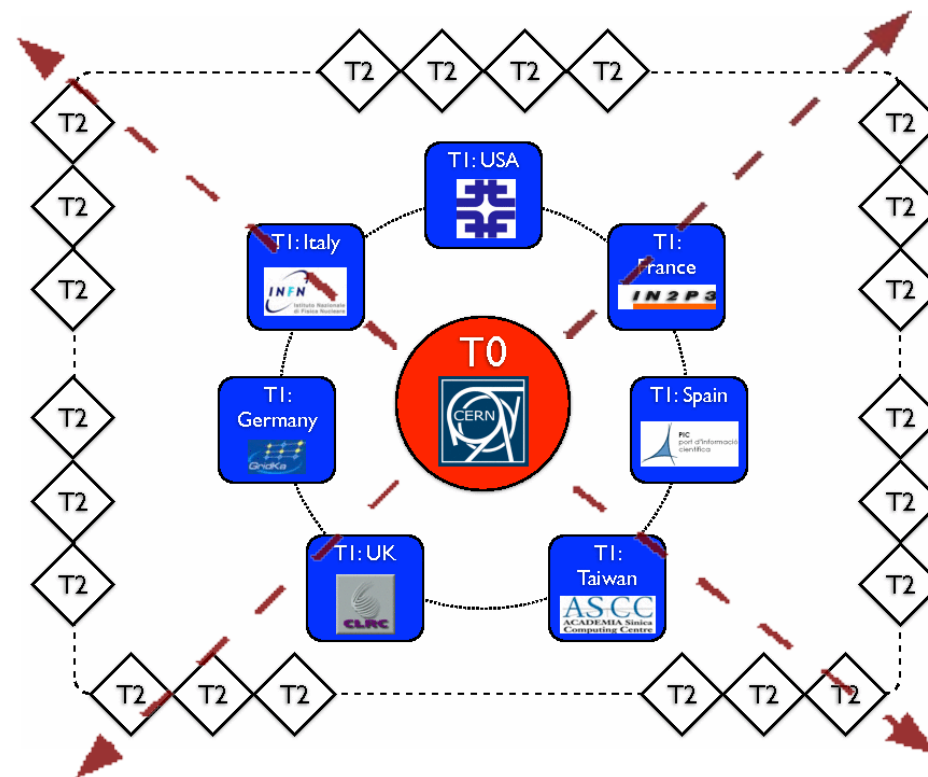




Data Distribution Infrastructure

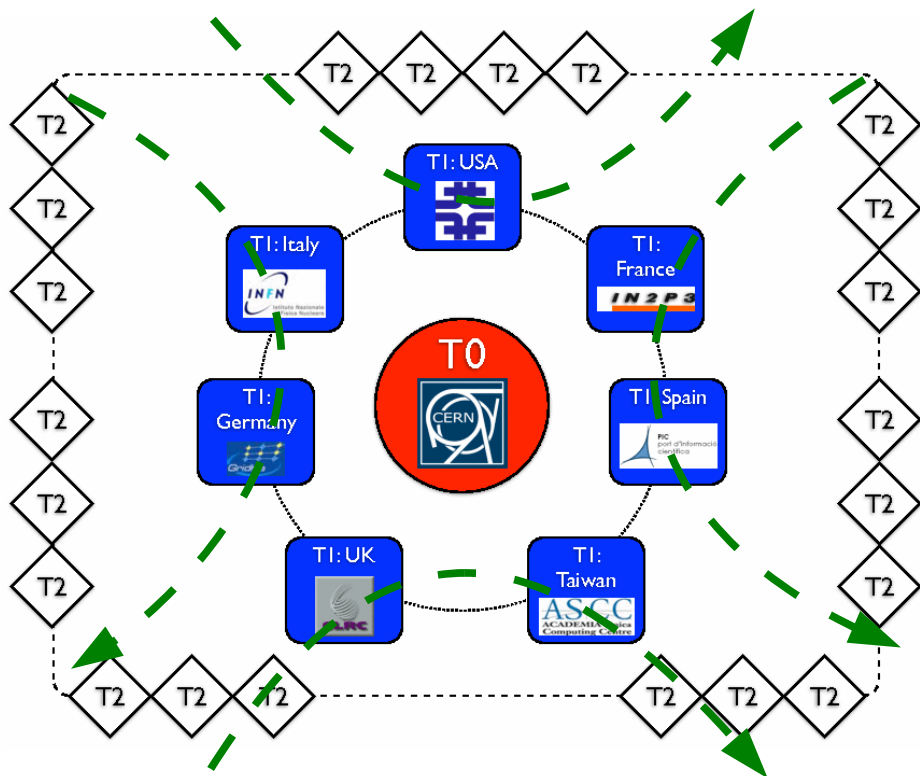


- RAW Data
 - Archived at CERN
 - Duplicated to Custodial Tier-1
- Reconstruction
 - Express at Tier-0
 - Prompt at Tier-0
 - Reprocessing at Tier-1
 - Detailed event description (RECO) archived at Tier-1
- Analysis
 - Skims produced at Tier-1
 - Analysis event content (AOD) distributed to Tier-2
 - Analysis jobs ran at Tier-2





Simulation Distribution Infrastructure



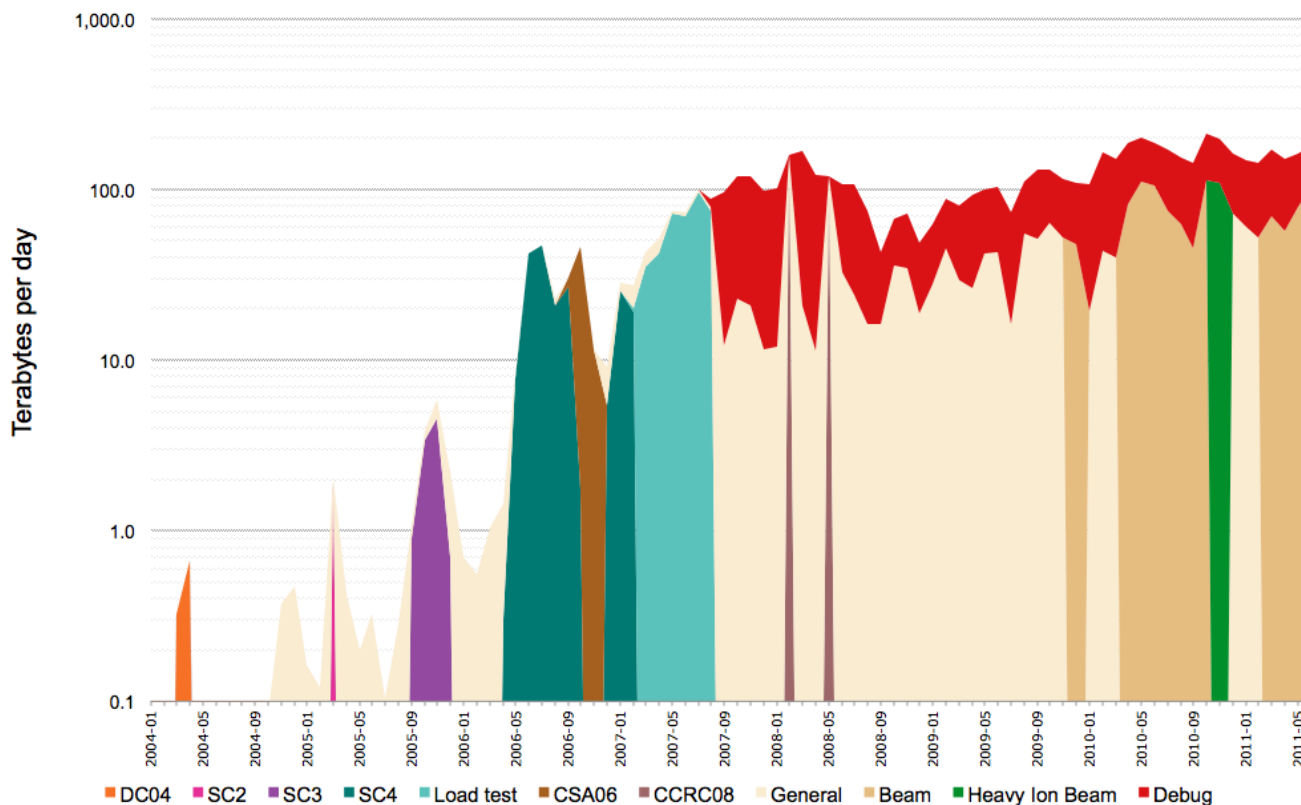
- Event Generation
 - Performed at Tier-2 and Tier-1
 - Archived back in Tier-1
- Reconstruction
 - Performed at Tier-2
 - Reprocessing at Tier-1
 - RECO archived at Tier-1
- Analysis
 - Analysis samples distributed to Tier-2
 - Analysis jobs ran at Tier-2



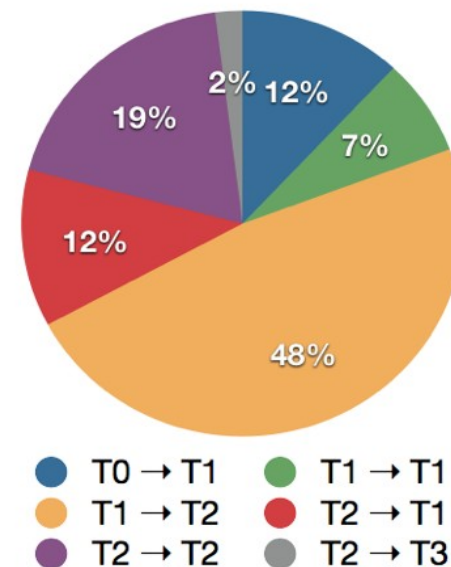
Traffic of CMS Data



Average data transfer volume



Breakdown of 2010 CMS data traffic in *PhEDEx*



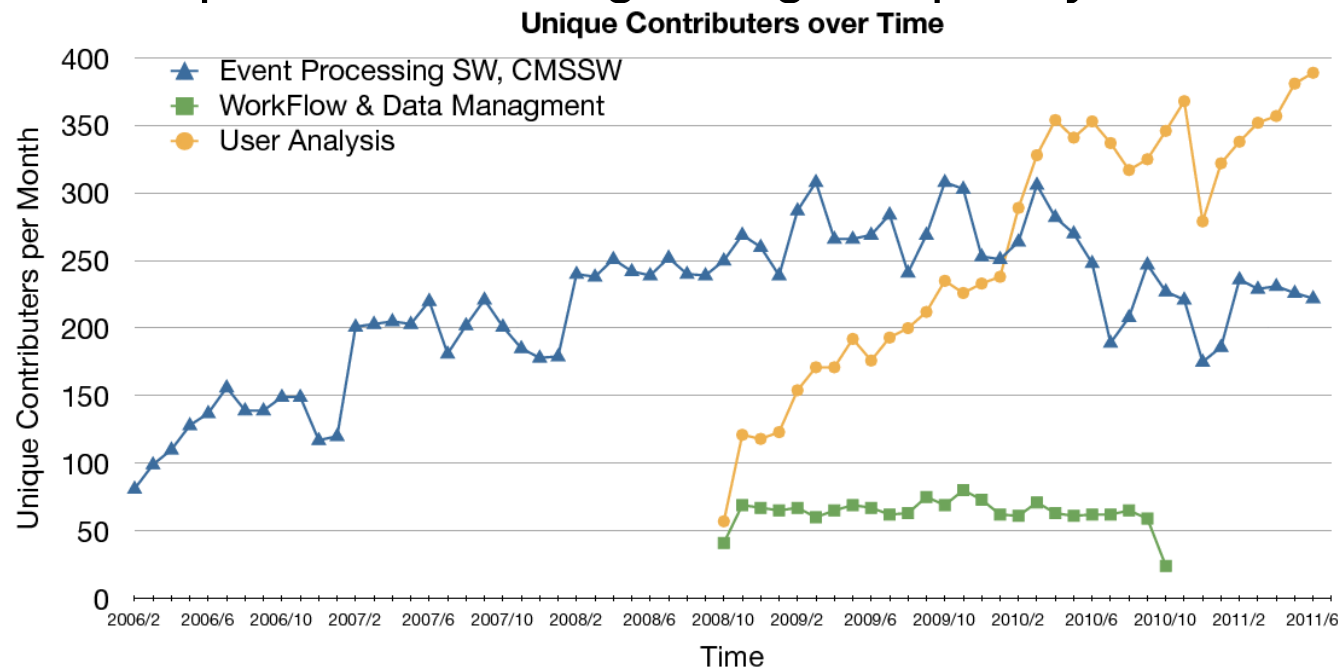
- Data transfer system handles large amount of data efficiently.
- T2 ↔ T2 transfer reduces the latency on availability of analysis samples.



Software Development



- Standard integration mechanism
 - ✓ Daily build, Performance monitoring, Sanity check.
- Frequent production of test samples
 - ✓ Software closure tests, Physics performance monitoring, Base for development.
- Active core team of software performance
 - ✓ Constant improvement with growing complexity.

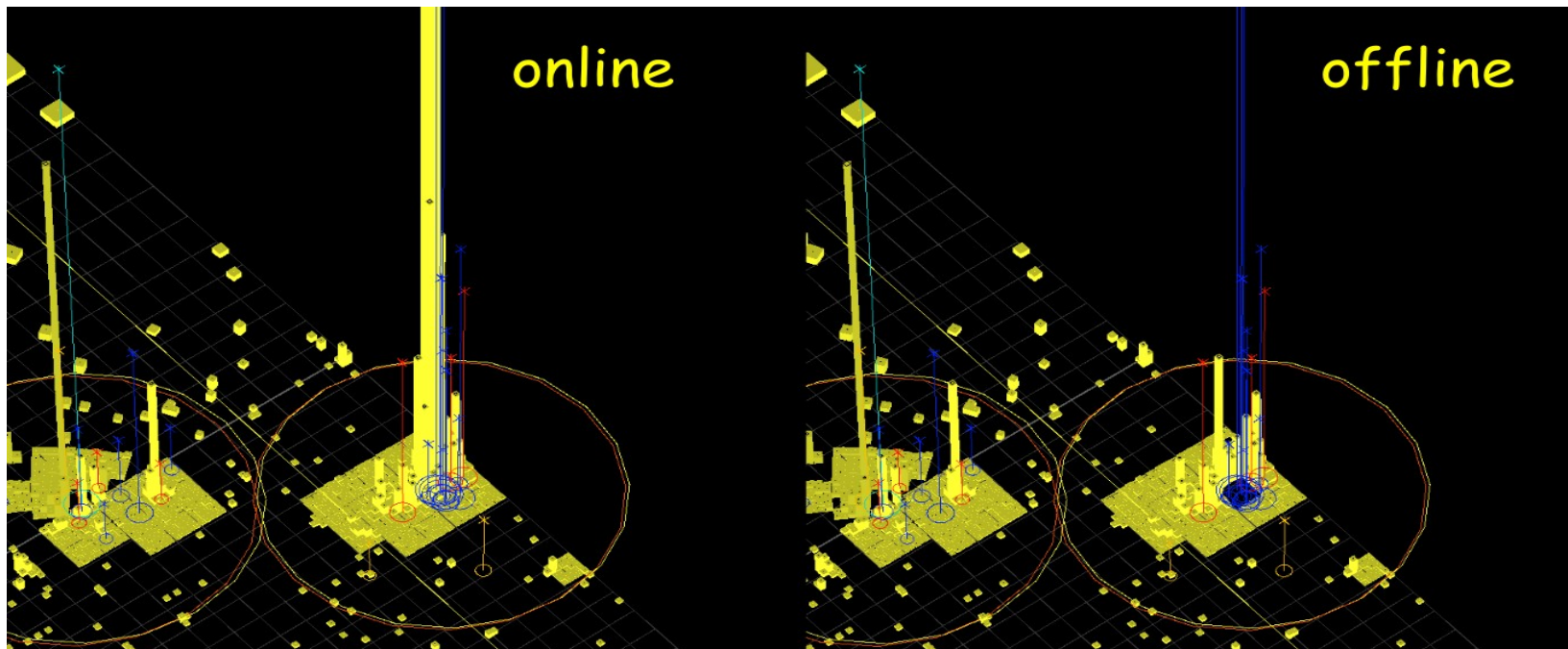




CMS Event Display

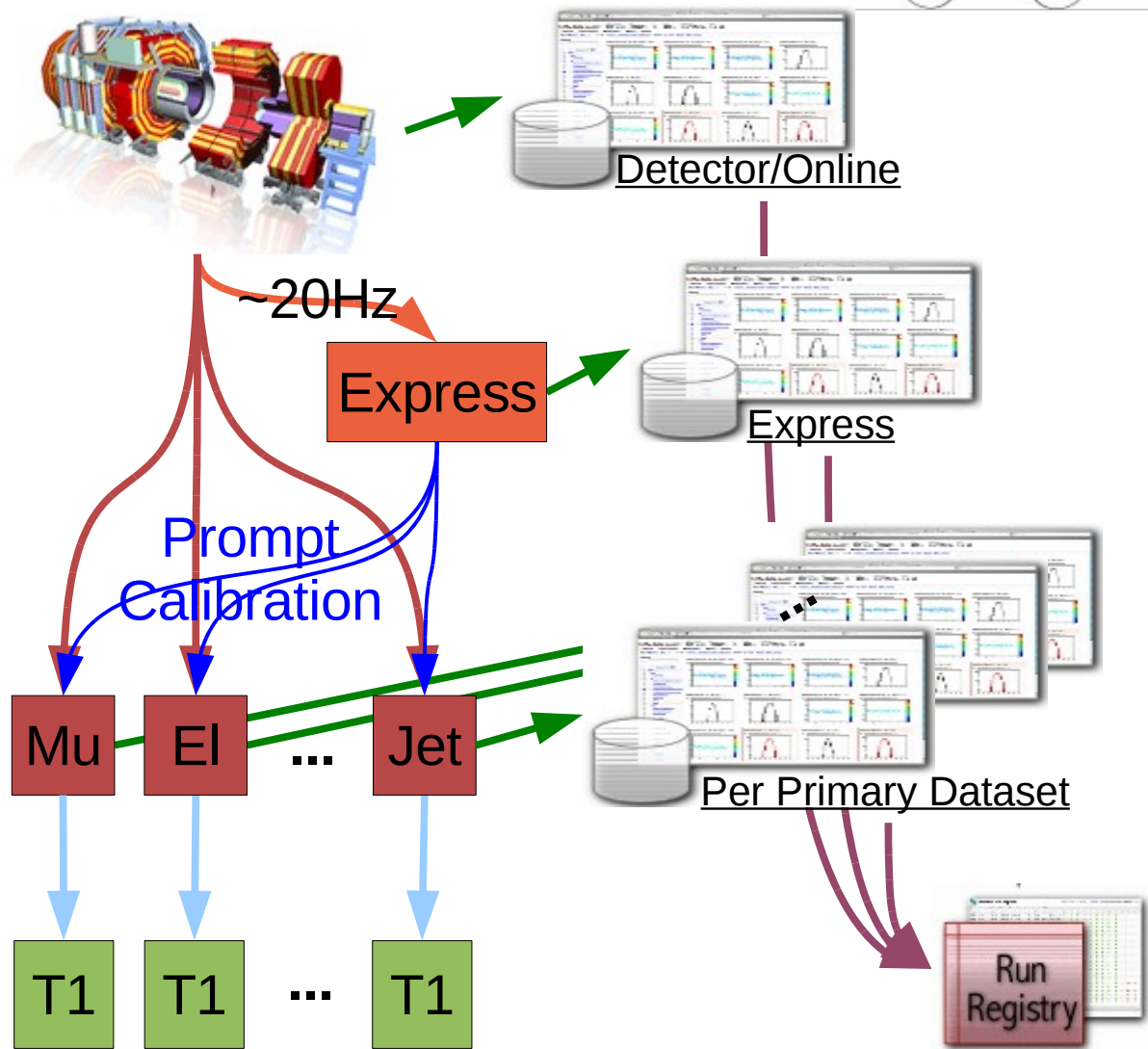


- Very detailed event display, consolidation of several projects.
- Available under multiple forms
 - Light-weight : portable.
 - Full-framework : running reconstruction software “live”.
 - Simulation geometry : full detail.
- Permits fast understanding of event details, software features, physics content.





Data Quality Monitoring



- Several Levels of monitoring
 - Detector/Online
 - Express reconstruction
 - Prompt reconstruction
- Web gui for visualization
- Easy access to information
 - Shifter/Expert layouts
 - Comparison
 - Summary plots
- Data certification
 - Lumi-section mask
 - From express early in early LHC runs, on prompt with increased luminosity
- Software Validation
 - Use the same infrastructure



CMS Software Updates



Simulation

- Full and Fast applications
- Improving generator interfaces
- Geometry, Material budget

Reconstruction

- Algorithmic
- Calibration
- Event size on tape

Updates targeted and scheduled for production of consistent Simulation and Data samples for Physics conferences.

Reprocessing of data is a matter of weeks.

Updating simulation samples is the most expensive part; reprocessing is measured in months.

Fast simulation is under-used for central production.

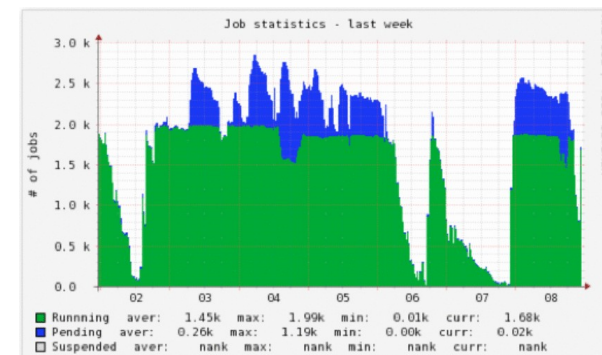
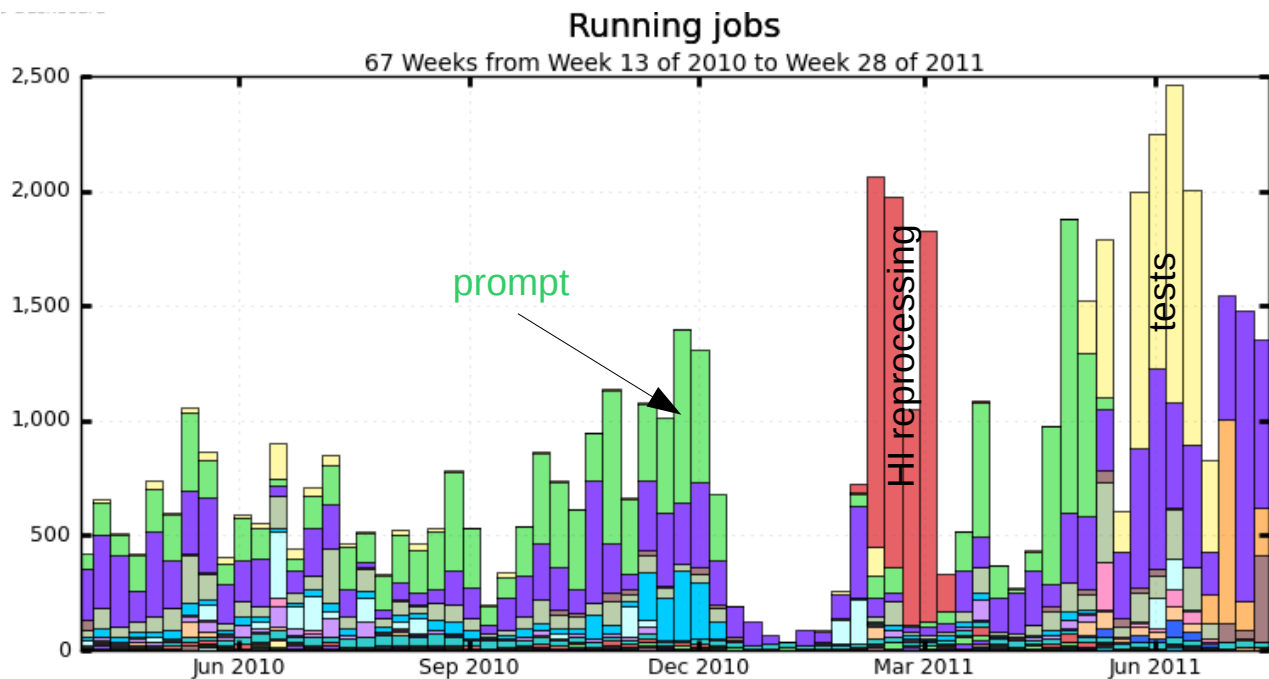


Tier-0 in 2010-2011



Prompt reconstruction

- 2010
 - 911M events, 3 major releases deployed ($\sim 10^{-8}$ event failure)
 - ~ 500 kB/event distributed centrally (detailed)
- 2011
 - ~ 1 B events, 2 major releases deployed
 - ~ 100 kB/event distributed centrally (analysis oriented)



- Full occupancy of the Tier-0 farm for days in the last LHC fills.
- On-going work for optimized utilization.



Data Availability in 2010-2011

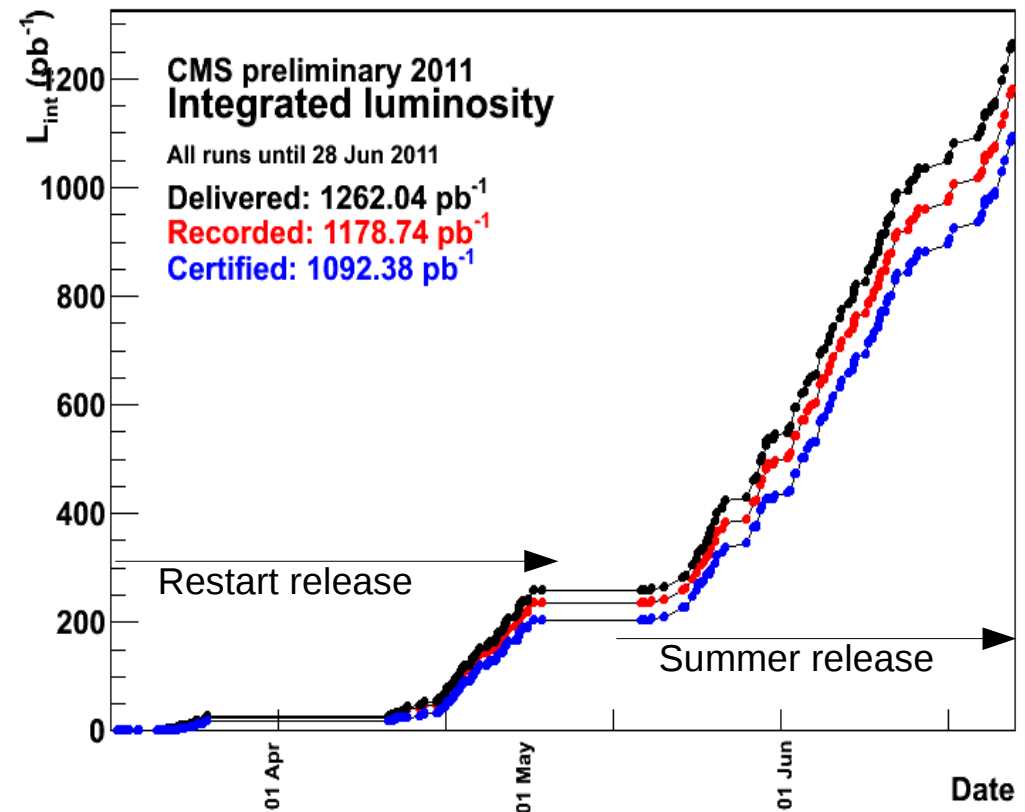


Reprocessing in 2010 ($\sim 50\text{pb}^{-1}$)

- Detailed event distributed centrally ($\sim 500\text{kB}/\text{event}$)
- 4 major releases + 2 calibrations
 - 2 passes for ICHEP
 - 1 pass for all 2010 data
 - 1 pass towards Moriond

Reprocessing in 2011

- Analysis event distributed
- 1 release
 - 1 pass for all 2010 data
 - 1 pass at Tier-0 switch
- Constantly provided a consistent dataset of CMS data to analysis.
- Promptly provide certification of available dataset.

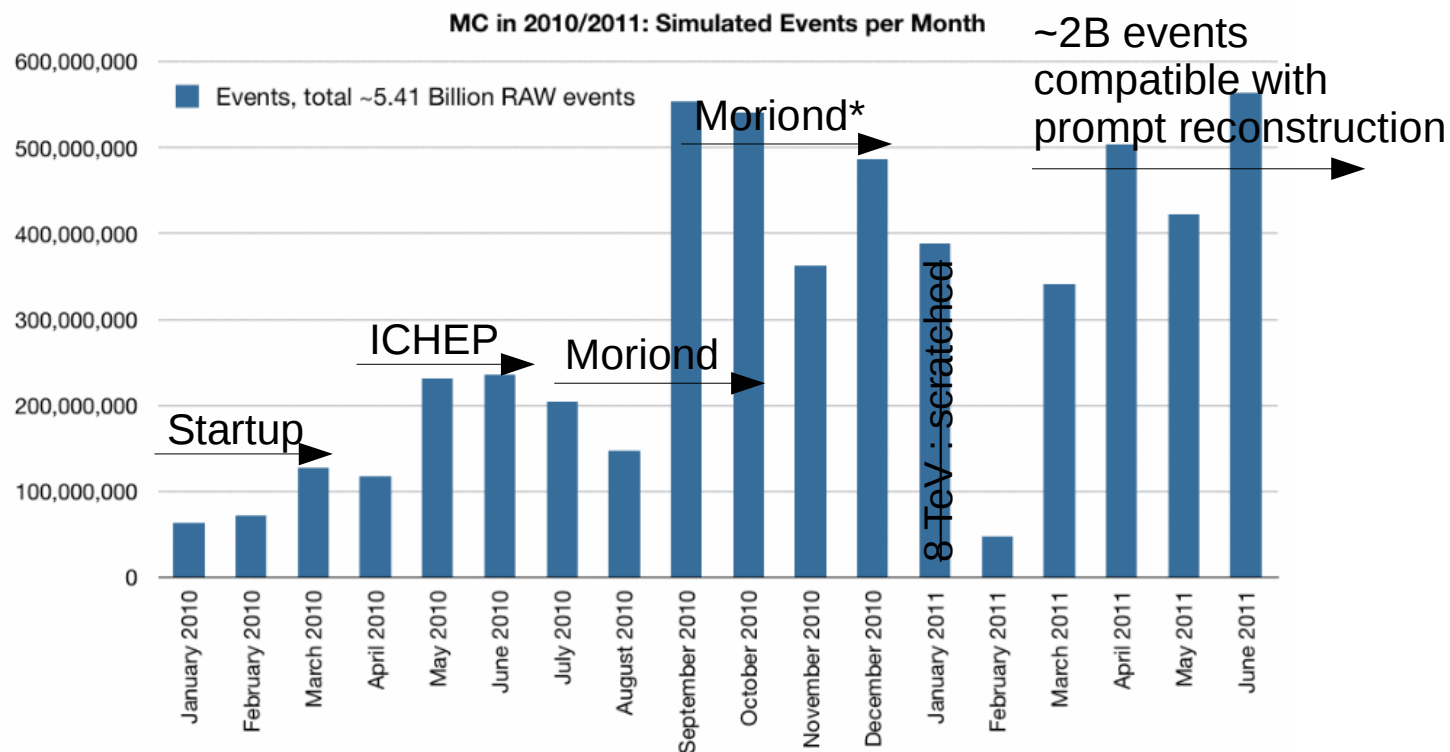




Simulation in 2010-2011



- Monte Carlo samples generated from scratch in two main campaigns.
- Pile-up added on top of detector simulation : less CPU expensive.
- Reprocessings with release compatible with prompt reconstruction.
- Always providing dataset consistent with data to be analyzed.
- ✓ Full availability made possible with contingency in schedules.

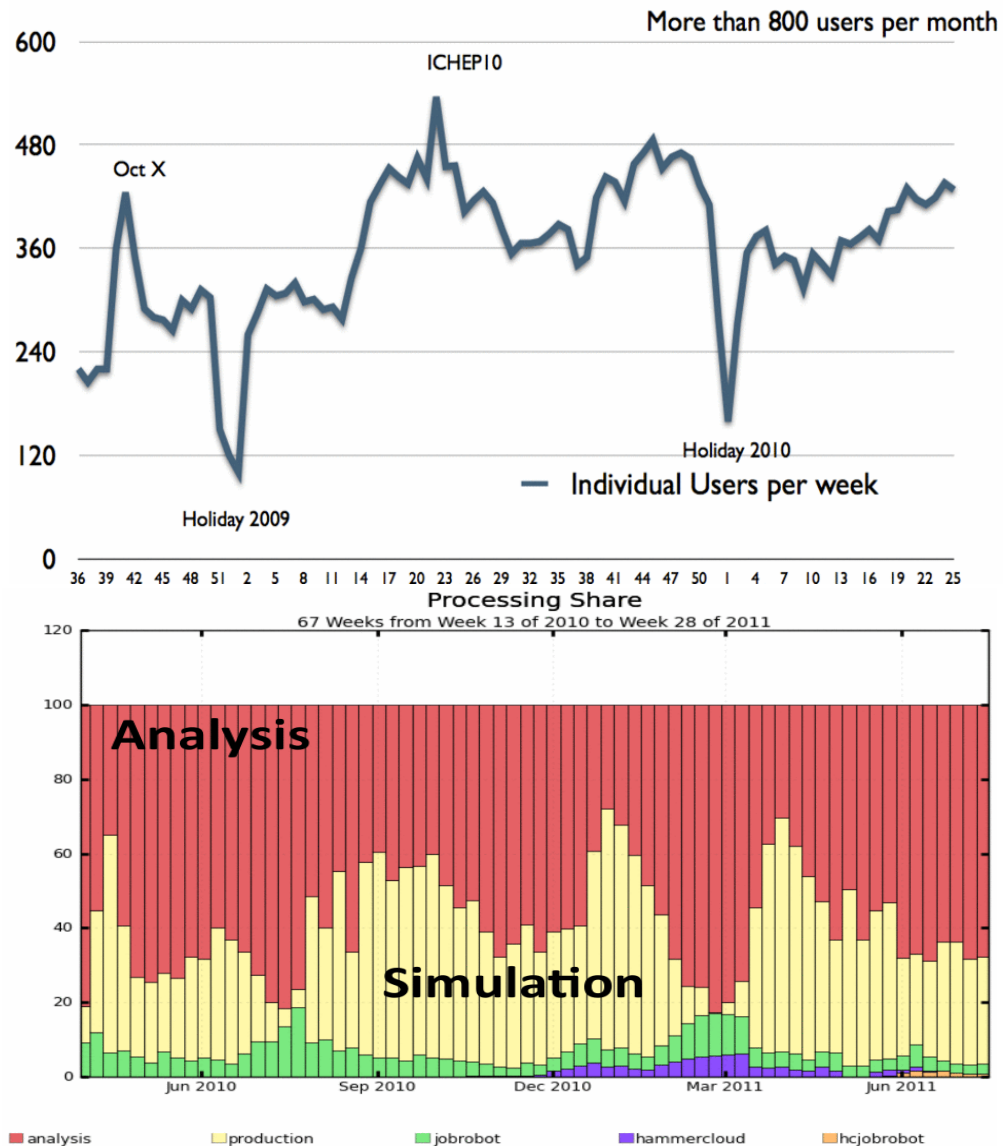




Analysis in 2010-2011



- Physics Analysis Specific Releases produced
- Analysis user friendly environment created
- Wide usage of the GRID
- Analysis at Tier-2
 - Shared with MC production.
 - ~40 / 52 ready for computing.
 - Increasing share due to partial shift of simulation production to Tier-1.





Summary



- The Computing Project efficiently and successfully handled the data from the CMS detector, and billions of simulated events.
- The Offline Project well managed the software necessary for processing of the CMS data.
- Infrastructure and mode of operation were adapted to change of regime of data taking and distribution.
- Transition to stable data taking operation with high pile-up at the LHC.

