SOFTWARE, COMPUTING AND CHALLENGES AT CMS TIERS

N. De Filippis[1)]

1. INFN Sezione di Bari, Bari, Italy

On Behalf of the CMS Collaboration

Abstract — The Computing, Software and Analysis Challenge 2006 (CSA06) was a realistic test of the CMS Computing Model up to a scale comparable to 25% of the complexity expected in 2008. Pre-production of large Monte Carlo samples was executed before starting CSA06; these events passed through prompt reconstruction at the CERN Tier-0 and they were subsequently distributed at the participating Tier-1s. Skimming jobs running at the Tier-1s selected events that were propagated to the targeted Tier-2s for the subsequent end-user analysis. Calibration and re-processing exercises took place at various centers. The CSA06 exercise heavily used Grid tools provided by the LCG project. An overview of the final results of the CSA06 is presented in this work.

1 CSA06 definition

The CMS Computing Software and Analysis Challenge was a 50 million event full exercise to test the workflow/dataflow of the CMS computing model [1] at a 25% scale of the complexity expected in 2008.

It consisted of several tasks:

- Monte Carlo simulation of large datasets;
- prompt reconstruction at Tier-0;
  - reconstruction at 40 Hz using the new CMS software, CMSSW;
  - application of calibration constants from off-line database;
  - generation of Reco, AOD, and AlCaReco reduced format datasets;
- distribution of all FEVT/AOD/AlcaReco to all participating Tier-1 centres;
- calibration jobs on alignment/calibration datasets at some Tier-1s and at CERN;
- re-reconstruction performed at a Tier-1s;
- skimming jobs at some Tier-1s with data propagated to Tier-2 centres;
- physics jobs at Tier-2s and Tier-1s on AOD and Reco.

Some of the most important performance metrics of the CSA06 were:

- Number of participating Tier-1s - Goal: 7 - Threshold: 5;
- Number of participating Tier-2s - Goal: 20 - Threshold 15;
- Weeks of running at sustained rate - Goal: 4 - Threshold: 2;

- Tier-0 Efficiency - Goal: 80 % - Threshold: 30 %, measured as unattended uptime fraction over 2 best weeks of the running period;
- Running grid jobs (Tier-1+Tier-2 overall) per day (2 hour jobs typically.) - Goal: 50 K - Threshold: 30K;
- Grid job efficiency - Goal: 90 % - Threshold: 70 %;
- Data serving capability at each partecipating site: Goal 1MB/s/execution slot - Threshold : 400 MB/s (Tier-1) or 100 MB/sec (Tier-2)
- Data transfer Tier-0 to Tier-1 to tape - Individual goals (threshold at 50 % of goal); for CNAF it was: 25 MB/s;
- Data transfer Tier-1 to Tier-2 - Goal: 20 MB/s into each Tier-2 - Threshold: 5 MB/s;
- Overall "success" is to have 50 % of participant at or above goal and 90 % above threshold.

In the following sections all the steps of the CSA06 are briefly summarized and some results are reported.

2 PRE-CHALLENGE PRODUCTION

Before the effective execution of CSA06 the pre-production of 50 million event simulated samples started the last week of July. Four official teams were organized, one running in American sites with a OSG grid infrastructure [2] and three in Europe which ran in a LCG [3] environment.

Massive Monte Carlo production was done with ProdAgent [4] tool that was developed with the purpose to make production automatic as much as possible and to handle thousand of jobs in parallel. ProdAgent consists of several agents running in parallel to perform tasks such as the job creation, submission and tracking. The output files are registered in the central Data bookkeeping service [5] at CERN while the location of the blocks of files is stored in the Data Location System [6] in term of storage elements where the data blocks exist.

One of the LCG teams (LCG3) in charge for running CSA06 production was from INFN-Bari. The local setup at Bari for running the official production consisted of two user interface machines hosting the ProdAgent software and the related databases, one dedicated slave machine used for the real-time mirror of the master databases hosted on the UIs and for remote monitoring purposes and load balancing. Another machine was setup for tests and the injection of files for the transfer and one running the agents for the transfer tool itself.

A web interface for monitoring purposes was developed in ProdAgent (fully integrated in ProdAgent at this time) in order to follow the progress of the production in real-time and to produce plots with statistics per dataset and per site. A screenshot of that is reported in Fig. 1.

About 66 million of simulated events were produced overall between the end of August and middle September at Tier-1s and Tier2s; they were mainly minimum bias, $t\bar{t}$, electro-weak and HLT soup events, as detailed in Table 1, which are background for most of the analyses to be performed at CMS.

Table 1: Final statistics of unmerged and merged events for the samples produced by the INFN-Bari team.

Data sample	N. of unmerged events (million)	N. of merged events (million)
Minimum Bias	8.5	7.3
T Tbar	1.4	1.3
Electroweak soup	2.2	1.6
Jet calibration soup	1.6	1.3
HLT soup	0.2	0.2

A screenshot of the plots provided by the monitoring interface that show the distribution of the jobs over the sites for the Minimum Bias sample is reported in Fig. 2.

The overall efficiency of the production was of about 88% with a significant contribution of inefficiency related to some problems with the not robust enough procedure to stage-out the output files and of all the issues related to the local configuration of sites. Merge jobs caused in general an overload of the storage systems due to the intensive file access, irrespective of the storage technologies used, CASTOR [7], dCache [8] or DPM [9].

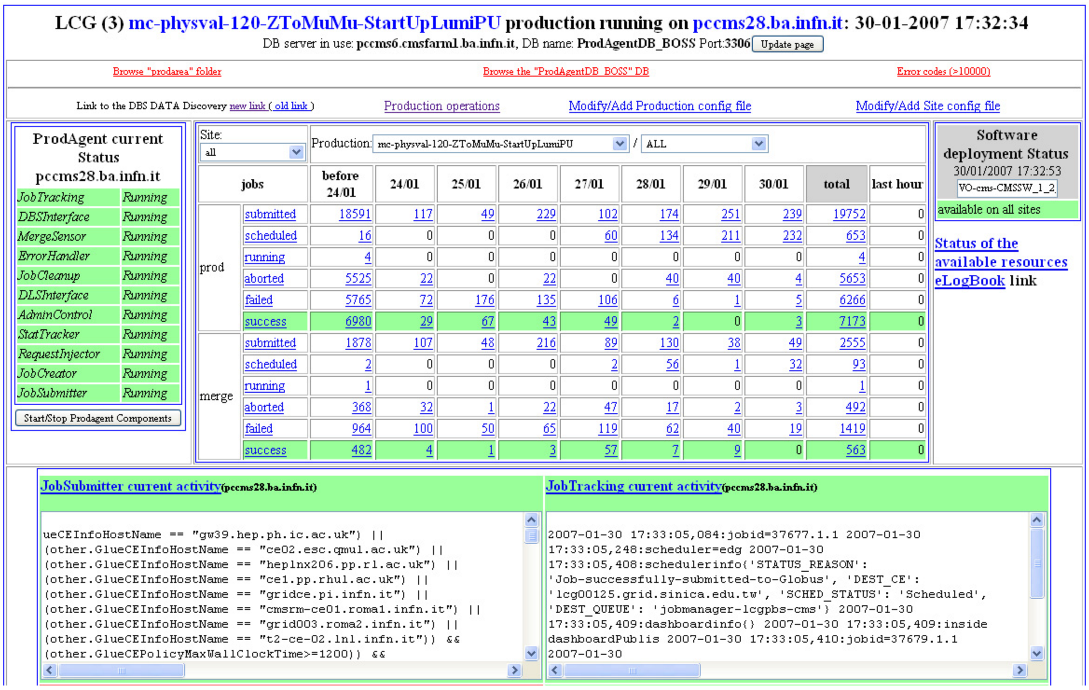


Figure 1: Screenshot of the monitoring web interface to ProdAgent.

3 RECONSTRUCTION AND ALIGNMENT EXERCISE

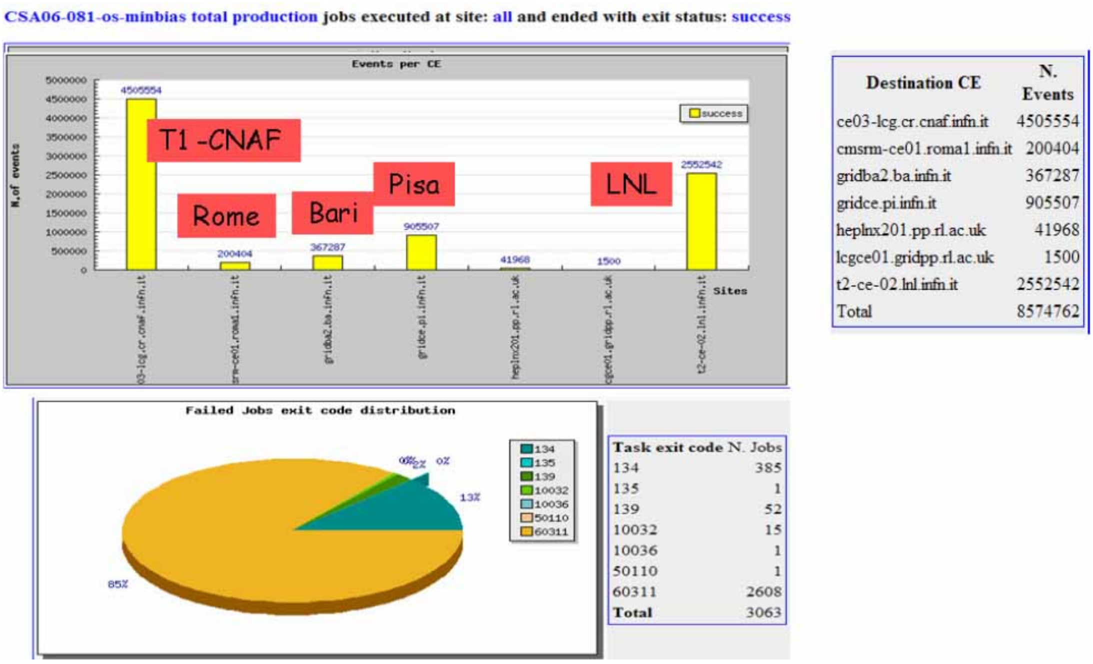
At the end of Monte Carlo production the prompt reconstruction of the simulated data started at CERN, including detector-specific local reconstruction and clustering, tracking, vertexing, standalone muon reconstruction, jets global reconstruction (with tracker), electrons, photons, b and τ tagging; the computing time for reconstruction not exceeded 25 s/event.

Processing for CSA06 was officially launched at Tier-0 October 2; after having run for about 23 days, 120 million events at 100% uptime, it was decided to increase scale for last days so all the signal samples were reprocessed in 5 days using CMSSW106 and maximum CPU usage. 160 Hz processing rate was reached, peaking at 300 Hz signals; 1250 CPUs for prompt reconstruction were used, 150 CPUs for AOD and AlCaReco production (separate step). Calibration/alignment constants, as obtained from offline database were applied to all events as part of prompt reconstruction.

A total of 4 weeks uptime (goal) was reached and 207 million events were processed and automatically transferred to Tier-1s. More than 1 PB of data were transferred using the CMS official tool for data movement, PhEDEx [10], reaching a maximum rate of 550 MB/s from CERN.

An alignment exercise was also executed at the first prototype of the Central Analysis Facility at CERN, which consisted of:

- the alignment of a portion of the CMS tracker with the High Impact Parameter algorithm [11] by using $Z \rightarrow \mu\mu$ events in a misalignment scenario close to the first data taking conditions;



Destination CE	N. Events
ce03-lcg.cr.cnaf.infn.it	4505554
cmsrm-ce01.roma1.infn.it	200404
gridba2.ba.infn.it	367287
gridce.pi.infn.it	905507
heplnx201.pp.rl.ac.uk	41968
lcgce01.gridpp.rl.ac.uk	1500
t2-ce-02.lnl.infn.it	2552542
Total	8574762

Task exit code	N. Jobs
134	385
135	1
139	52
10032	15
10036	1
50110	1
60311	2608
Total	3063

Figure 2: Distribution of the minimum bias jobs over the sites. A chart with the failure exit codes per site is also shown.

- Intercalibrate ECAL crystals by ϕ symmetry in minbias events.

Specialized reduced Reco data format (AlCaReco) were used for calibration/alignment stream from Tier-0 and a mechanism to write constants back into offline database was tested.

The dataset used for the tracker alignment exercise was $Z \rightarrow \mu\mu$. 1 million events were used and 10 iterations with the High Impact Parameter algorithm; jobs ran in parallel on 20 CPUs on a dedicated queue at Tier-0 and new costants were inserted into the offline database successfully. The convergence of the tracker detector position residual distribution was verified after the 10 iterations.

4 SKIMMING PROCEDURES

Skimming procedures ran at Tier-1s to reduce primary datasets to manageable sizes and to select interesting events and data information dependent on the subsequent analyses to be run at Tier-2s. Several data formats are used with the purpose of saving a fraction of the information of the simulation and reconstruction steps out of the full event content of the input files; files in Reco format (and RECOSIM with the addition of some monte carlo generation and simulation information) included reconstructed clusters, tracks, vertices, calorimetric clusters and HLT objects; clusters and hits were dropped in the AOD output format (such as in AODSIM where some Monte Carlo generation information were saved).

Skimming selection were prepared by the analysis teams at Tiers and were mostly based on a selection of Monte

Carlo candidates to fit with signal of events looked for such as Z $\rightarrow \mu\mu$ or W $\rightarrow \mu\nu$. About 10 filters were prepared by the Italian people to cope with the analyses prepared in Italy.

Skimming operations were run with ProdAgent by the 4 teams organized for production and started October 9. The production team from INFN/Bari run 13 skim filters over the Electroweak soup dataset at CNAF, RAL, IN2P3, ASGC Tier-1s.

The skimmed output files were transferred at Tier-2s for end-user analyses. Some problems occurred with the transfer from many Tier-1s mainly related both to bad configuration of sites or of the PhEDEx agents running with the FTS system [12] and to Tier1s problems with the FTS server and the storage system.

The final piece of CSA06 functionality at Tier-1 was the demonstration of re-reconstruction. The FEVT files generated at Tier-0 were re-processed reading the updated calibration/alignment constants from the frontier/squid database [14]. The production teams used ProdAgent to submit about 100 re-reconstruction jobs of 1000 events each at each Tier-1.

5 ANALYSIS EXERCISES

The last step of the challenge was the preparation and the run of some analysis codes to extract physics results by running on skimmed files at Tier-2s. The Italian teams at Tier-2s performed some analyses that were summarized below.

5.1 Tracker misalignment studies at Bari

The team at Bari performed an analysis with the purpose to study the effect of the CMS tracker misalignment on the performances of the track reconstruction [13]. Three configurations were investigated:

- the ideal scenario with a perfect tracker geometry;
- the short term misalignment scenario when the uncertainties on the position of the sub-structures of the CMS tracker are between $10\,\mu\mathrm{m}$ for pixel detectors and $400\,\mu\mathrm{m}$ for microstrip silicon detectors in the endcaps. Detector position and errors are read from the offline database at CERN by caching the needed information locally via frontier/squid software [14]. Because of the misalignment the analysis requires to refit tracks by using the Kalman filter formalism [15] with a misaligned tracker geometry.
- the long term scenario when the alignment uncertainties are supposed to be a factor 10 smaller because of the improvement obtained by using alignment algorithms with a high statistics of tracks.

Global efficiency of track reconstruction and track parameter resolution for muons and fake rate were compared in all the cases. The following data samples used were: $Z \rightarrow \mu\mu$ and $t\bar{t}$ (the second for computing the fake rate).

The global efficiency of track reconstruction of muons coming from Z decay is shown in Fig. 3 (left) as a function of the pseudorapidity in the tracker acceptance; the effect of misalignment is relevant in the short term scenario and causes a partial inefficiency of the track reconstruction; that can be recovered if the intrinsic position resolution of the tracker detector is combined with the alignment uncertainties to make larger the error on the position of the reconstructed hit (called alignment position error, APE) so improving the track fit at the expense of a larger rate of fake tracks.

The residual of Z mass obtained as the invariant mass of muons coming from Z decay in the case of perfect tracker geometry and in short-term and long term misalignment scenarios is shown in Fig. 3 (right); the Z mass resolution is degraded of a factor 2 because of the tracker misalignment in the short term scenario.

About 2.5 thousands jobs (45 at most in parallel) ran with a grid efficiency of 90 % and an application efficiency of 80%, by accessing detector position and errors from the offline database via frontier. The performance of the access via frontier at Bari is reported in Fig. 4 where the number of requested kbyte to be read by each event is reported as a function of time. Because of a wrong configuration of the local cache of squid the information about alignment uncertainties were read from the CERN database rather that from the local cache as reported in blue; after increasing the local cache the behaviour obtained was that expected and jobs accessed alignment data via the local cache as shown in the distribution at subsequent time periods.

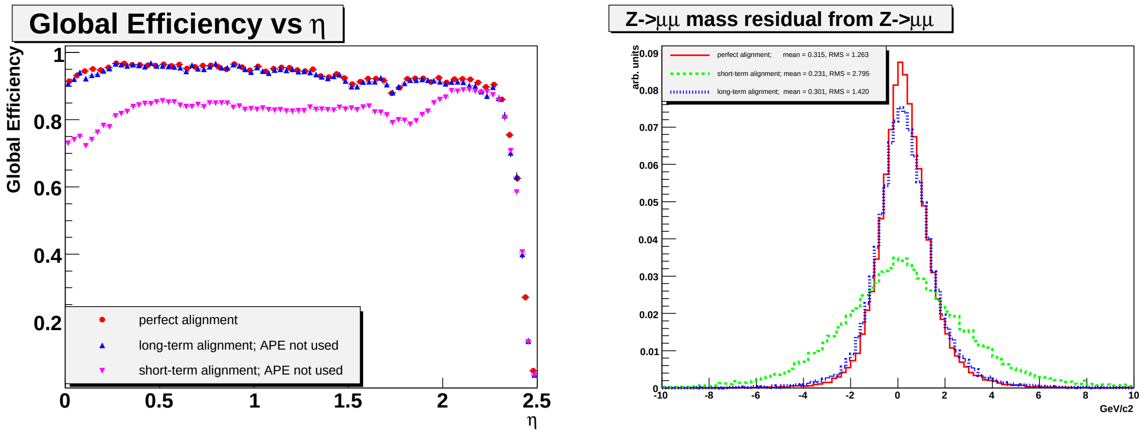


Figure 3: ((left) Global track reconstruction efficiency vs pseudorapidity for muons coming from Z decay in the case of perfect tracker geometry and in short-term and long term misalignment scenarios when the APE is not used; (right) Residual of Z mass obtained as the invariant mass of muons coming from Z decay in the case of perfect tracker geometry and in short-term and long term misalignment scenarios.

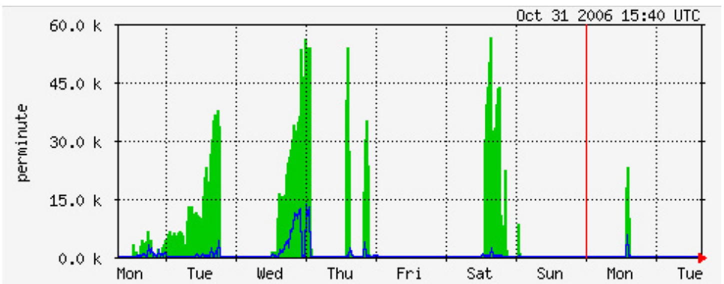


Figure 4: Access via frontier at Bari: number of requested kbyte as a function of time.

5.2 Reconstruction of $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ events at Rome

The main purpose of the exercise at Rome was the selection of $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ out of the "Electro Weak Soup" and their use to evaluate the performances of the electron reconstruction; the demonstration of the workflow for running offline Ecal calibration in a remote Tier-2 centre, starting from an AlcaReco stream produced at the Tier-1 centre was verified.

A skim filter designed to produce as output two streams for the same events was developed: one stream was the standard electron AlcaReco stream, the other one was the standard RECOSIM output. The events were selected on a MonteCarlo basis, requiring only the Drell-Yan $\mathrm{e}^+\mathrm{e}^-$ events in a mass range from 50 to 130 GeV/c^2. The electrons were also required to have $\mathrm{p_T}$ greater than 5 GeV/c and $|\eta| < 2.7$. This selection has an efficiency over the Electro Weak Soup of 3.7 %.

Skim jobs have been processed at the CNAF Tier-1, starting from less than 3 million EWK soup events and the two created streams were transferred to the Rome Tier-2. Analysis jobs ran at Rome Tier-2 using CRAB on both the RECOSIM and the AlcaReco streams.

The electron reconstruction efficiency versus p_t and η are shown in Fig. 5.

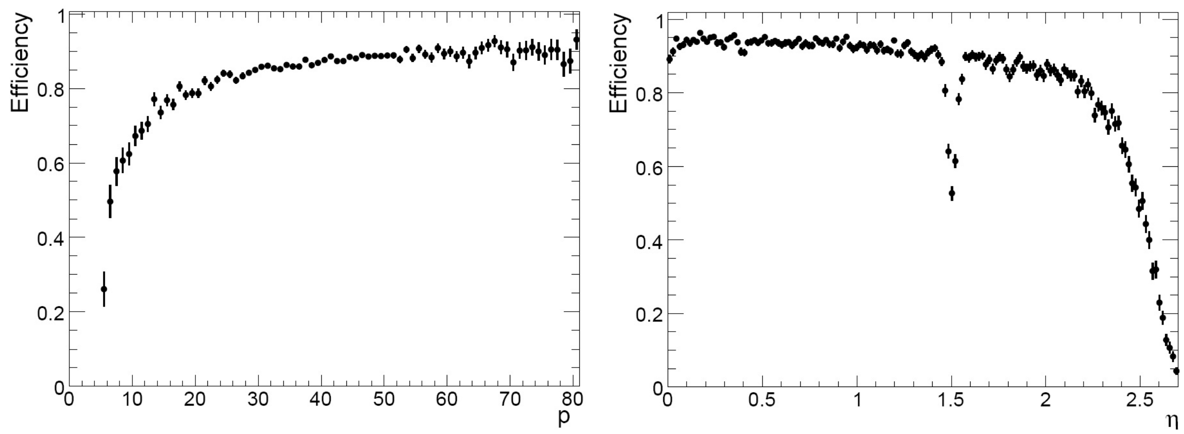


Figure 5: (left) Electron reconstruction efficiency as a function of p_T and (right) η.

The quality of the track reconstruction was evaluated looking at the E/p distribution, using the track parameters both at the vertex and at the outermost layer. At the beginning of the exercise the parameters of the reconstructed electron track were not stored after the smoothing step, and for this reason the quality of the E/p distribution with the track parameters at the vertex is rather poor. A refitting of the electron track was executed offline so improving the quality of the E/p matching.

5.3 $\mathrm{W} \rightarrow \mu\nu$ selection at Legnaro/Padova

Several data samples with different signal composition were analyzed to study the $\mathrm{W} \rightarrow \mu\nu$ pre-selection. The efficiency and resolution of the muon reconstruction obtained with the CMSSW software were studied as a function of $\mathrm{P_T}$ and η, the $\mathrm{W} \rightarrow \mu\nu$ signal was selected with isolation cuts and the transverse mass was compared with the same distribution obtained from the background.

The following data samples were considered:

- the 'electroweak soup' (3.4 M evts, 50% $\mathrm{W} \rightarrow \mu\nu$ and 50% DY)
- the 'soft muons' (1.8 M evts, 50% minimum bias and 50% J/Ψ, $\mathrm{P_T} > 4\mathrm{GeV}$)

The skim jobs tan at the PIC and CNAF Tier-1s and the output was sent at the LNL Tier-2 in AODSIM format (Analysis objects). For the $\mathrm{W} \rightarrow \mu\nu$ channel the skim consists in the selection of events having at least one muon (StandAlone muon reconstructed object) with $\mathrm{P_T} > 10$ GeV.

The efficiency of the 'electroweak soup' data skimming resulted to be 16% while the selection efficiency for the $\mathrm{W} \rightarrow \mu\nu$ skimming of the soft muon sample resulted to be 50%. The 'soft muons' consisted of the mixing of approx 1 Mevts of J/Ψ inclusively produced in pp collision and 1 Mevts of $\mathrm{pp} \rightarrow \mathrm{m} + \mathrm{X}$ events from a minimum bias sample, preselected with a requirement of $\mathrm{p_T}(\mu) > 4\ \mathrm{GeV/c}$ at generator level.

Figures 6 (left) and 6 (right) show the transverse momentum of muons after the skimming, obtained with the GlobalMuon reconstructor, for the 'electroweak soup', the 'soft muons'.

The transverse mass of the W boson candidates was reconstructed using the missing energy information. The neutrino $\mathrm{p_x}$ and $\mathrm{p_y}$ were assumed to be equal to $\mathrm{MET_x}$ and $\mathrm{MET_y}$. The MET was estimated with the sum of the momentum of tracks and the raw energy of the calorimeter towers. The $\mathrm{W} \rightarrow \mu\nu$ signal was compared

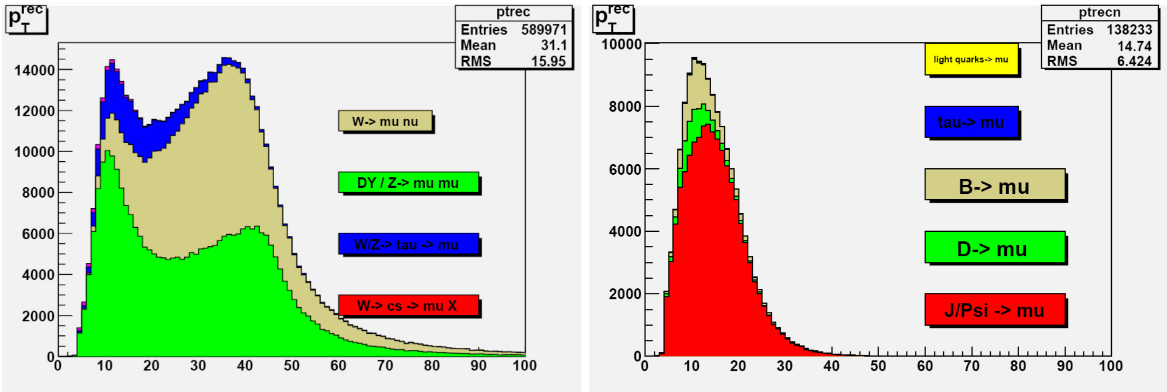


Figure 6: (left) Transverse momentum of muons after the skimming selection for the different 'soft muons' components; (right) Transverse momentum of muons after the skimming selection for the minimum bias component of the 'soft muons'.

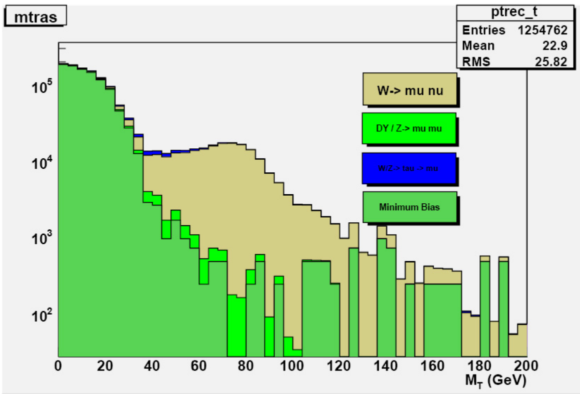


Figure 7: Transverse mass of W boson candidates after the skimming selection.

with the minimum bias and the $Z \rightarrow \mu\mu$ backgrounds with the cross section normalization. A significative signal/background enhancement was obtained with the isolation cut. The muon was considered isolated when the calorimeter transverse energy and the track transverse momentum were less than 5 GeV inside the cone $\Delta\eta < 0.5$. Figure 7 shows the transverse mass for the W boson candidates.

5.4 Minimum Bias and Underlying Event at Perugia/Pisa

The Minimum Bias (MB) event is the generic inelastic proton-proton interaction at LHC, including hard scattering and diffractive components. In order to describe the unavoidable background of the pile-up at high luminosity the features of the MB at the LHC and the cross check with the predictions of the Monte Carlo generators is fundamental. The Underlying Event (UE), instead, is the softer component of a single proton-proton collision accompanying the hard scattering, and accounting for a large fraction of the activity in terms of multiplicity and momentum of the observed particles.

Concerning the UE the topological structure of hadron-hadron collisions was used to study that activity looking

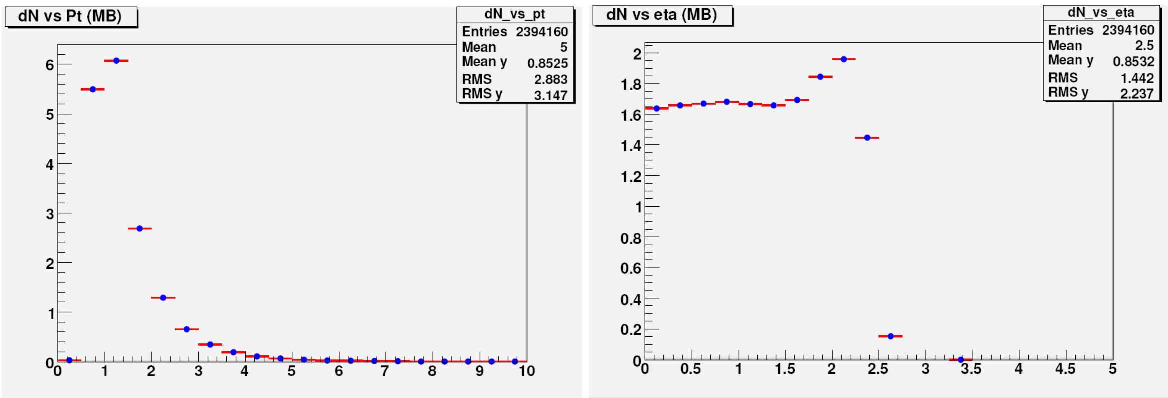


Figure 8: (right) Pt distribution of reconstructed track (left) multiplicity of reconstructed track as function of η

only at the outgoing charged particles. Jets were reconstructed from the charged particles using a simple clustering algorithm and then the direction of the leading charged particle jet was used to isolate regions of eta-phi space that are sensitive to the UE. The transverse region to the charged particle jet direction, is almost perpendicular to the plane of the hard 2-to-2 scattering and is therefore very sensitive to the UE.

The Underlying Event (UE) analysis is based on the selection of the following events: Minimum Bias, Hadronic Jets, Drell-Yan ($Z \leftarrow \mu\mu$). For each kind of events a filter was created that defines the output stream based on the RECOSIM format. In more detail the selection criteria that have been adopted for the UE analysis correspond to the following three filters:

- Minimum Bias: a Minimum Bias trigger strategy is not yet defined, as it is not defined higher level selection criteria for this kind of events. A random selection is applied in order to select the events without introducing any additional bias.
- Hadronic Jets: the selection is performed requiring at least one calorimetric jet with a Transverse momentum greater than 20 GeV/c.
- Drell-Yan ($Z \leftarrow \mu\mu$): the stream is defined requiring at least two muons in the central region ($|\eta| < 2.5$) having a minimum transverse momentum of 3 GeV/c and invariant mass above 15 GeV/c.

The first filter is also used in the Minimum Bias analysis. The previous filters were run at Tier-1 (FNAL and CNAF) and the output was transferred to the Pisa Tier-2 while the output of the Hadronic Jet and the Drell-Yan skims were transferred to the Florida Tier-2.

The analysis step was performed using CRAB. Two different analyses for the MB and UE observables were run at Pisa and Florida. Starting from a total sample of 3×10^6 MB events, 1.2×10^6 Hadronic Jets and 0.5×10^6 DY, we finally performed the analysis on 2×10^5 MB, 10^5 QCD Jet and 10^5 DY events.

In the MB analysis particularly interesting are transverse momentum and the track multiplicity distributions (Fig. 8).

In the UE analysis for the hadronic topologies, instead, the observables related to the activity in different regions with respect to the charged jets are interesting, by studying in particular the energy density ($\delta \mathrm{P_{Tsum}}/\delta\eta\delta\phi$) and the charge density ($\delta \mathrm{N_{chg}}/\delta\eta\delta\phi$).

5.5 τ validation package at Pisa

The τ reconstruction validation package was run at Pisa as an analysis exercise. The skim filter, based on Monte Carlo truth, selected events with at least one electron or muon with a transverse momentum of $\mathrm{p_T^{e,\mu}} > 15\,\mathrm{GeV/c}$ with a selection efficiency of $\approx$47% .

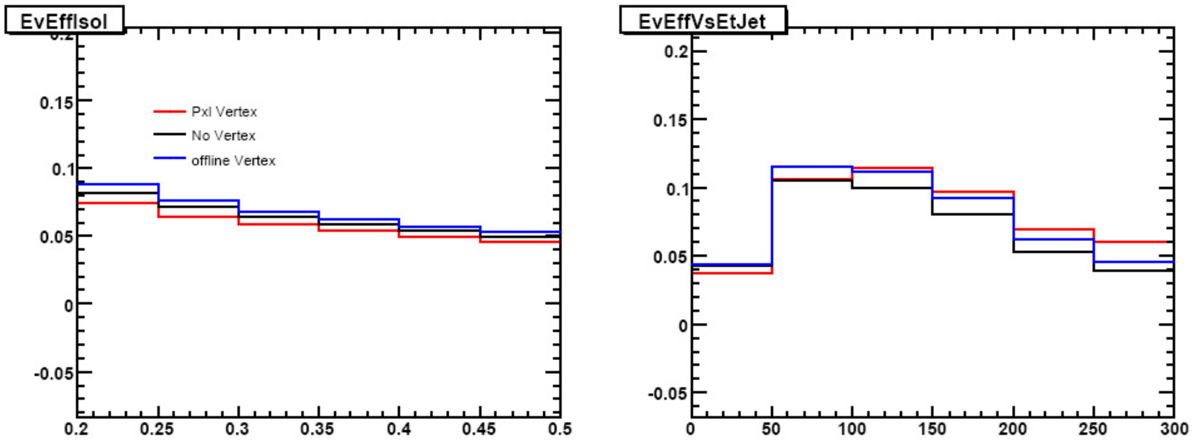


Figure 9: $t\bar{t}$ events: τ tagging efficiency vs isolation cone (left); τ tagging efficiency vs jet transverse energy (right). The various curves correspond to different selection of the primary vertex.

A total of 2.15M of $t\bar{t}$ events were produced and the RECOSIM output of the filter was moved to various T2 centers in UK, Italy, Switzerland and Germany. The jobs, submitted through CRAB from CERN, have been run at Pisa Tier2.

The τ validation package, whose scope is to monitor the performance of the various τ tagging methods, includes a number of standard histograms which will be produced periodically for each release of CMSSW and during data taking. Some of the most relevant histograms are the Efficiency vs Isolation Cone, Efficiency vs Jet transverse momentum, Efficiency to find the Leading Track, Number of Signal Tracks, ΔR between the leading track and the jet. As an example Fig. 9 shows the τ tagging efficiency versus the isolation cone (left) and versus the jet transverse energy (right). The various curves correspond to different selection of the primary vertex: for $t\bar{t}$ events without pileup, the various definitions are almost equivalent, as expected.

6 CONCLUSIONS

CSA06 was a successfull 50 million event test of the workflow and dataflow defined in the CMS computing model. 66 million of events were produced in about 1.5 months with an overall efficiency of about the 88 %.

Prompt reconstruction at Tier-0 was tested up to a rate of 160 Hz. Calibration and alignment exercises were executed successfully accessing the information from the offline database via frontier.

Reconstructed data we shipped at Tier-1s with a high transfer rate and without problems. Skimming jobs ran at Tier-1s sites and the output files were shipped to Tier-2s where analyses prepared by local physics groups ran as the last step of the CSA06 chain.

The INFN contribution to the CSA06 was impressive and really successfull thanks to few expert people and to sites administrators of Tier-1 and Tier-2 sites.

References

[1] CMS Computing TDR, CERN LHCC-2005-023, 20 June 2005.

[2] http://www.opensciencegrid.org

[3] http://lcg.web.cern.ch/LCG/

[4] https://uimon.cern.ch/twiki/bin/view/CMS/ProdAgent

[5] https://uimon.cern.ch/twiki/bin/view/CMS/DBS-TDR

[6] https://uimon.cern.ch/twiki/bin/view/CMS/DLS

[7] http://castor.web.cern.ch/castor/

[8] http://www.dcache.org/

[9] https://uimon.cern.ch/twiki/bin/view/LCG/DpmAdminGuide

[10] http://cmsdoc.cern.ch/cms/aprom/phedex

[11] "The HIP Algorithm for Track Based Alignment and its Application to the CMS Pixel Detector", V.Karimaki, T.Lampen, F.-P.Schilling, CMS NOTE-2006/018.

[12] http://egee-jra1-dm.web.cern.ch/egee-jra1-dm/FTS/default.htm

[13] http://webcms.ba.infn.it/cms-software/index.html/index.php/Main/StudiesOfCMSTrackerMisalignment

[14] http://lynx.fnal.gov/ntier-wiki

[15] R. Fruhwirth, Nucl Instr. and Meth. **A 262** (1987) 444.