CMS Operations Report

January 13, 2010
Ian Fisk

CMS Computing Operations Groups

- Data Operations (coords: Gutsche / Klute)

  - responsible for central data processing and transfers: RAW data repacking and prompt reconstruction at T0, RAW data and MC re-reconstruction and skimming at T1's, MC production at T2's

  - ensure central data consistency and data distribution to T0/T1s including custodial storage of primary datasets

- Facilities Operations (coords: Kreuzer / Flix)

  - Responsible of providing and maintain a working distributed computing fabric with a consistent working environment for Data Operations and Analysis users

    - It involves coordination of facilities operation, resource management and liaison to external projects and organizations

- Analysis Operations (coords: Wurthwein, Belforte)

  - Responsible for central data placement at T2 level, CRAB server operations, validation, and support, and for metrics, monitoring and evaluation of the distributed analysis system

Operations Meetings

▶ Monday 16-18 CERN Time is Combined Computing Operations

   ▶ Half Facility Ops and Half Data Ops

   ▶ Goal is to increase the overlap of the meeting and reduce the overall time. We encourage site participation (Continuing Dedicated Asian/Russian/Turkish T2 meetings)

▶ Thursdays 14:30-16:30 CERN Time is a Joint meeting between computing and offline.

   ▶ Computing Integration, Infrastructure Development and Release planning. (Interesting, if you want to get engaged in development)

▶ During Running periods there is a daily operations at 10 CERN time meeting between Computing, Offline, and Physics Validation (Very technical and topical) - Equivalent meeting in a time zone for the Americas at 23 CERN time

   ▶ Meetings and Contact info are at http://indico.cern.ch/categoryDisplay.py?categId=1374

▶ CMS attends the daily WLCG call at 15:00 CERN time

   ▶ Continuing TWIKI reports https://twiki.cern.ch/twiki/bin/view/CMS/FacOps_WLCGdailyreports

Site Readiness and Stability

- Tier-1 Readiness November and December

 - Readiness defined as passing the CMS, SAM, Job Robot, and Transfer tests for a high percentage of a time window

Site Readiness Status for CMS Tier-1 sites
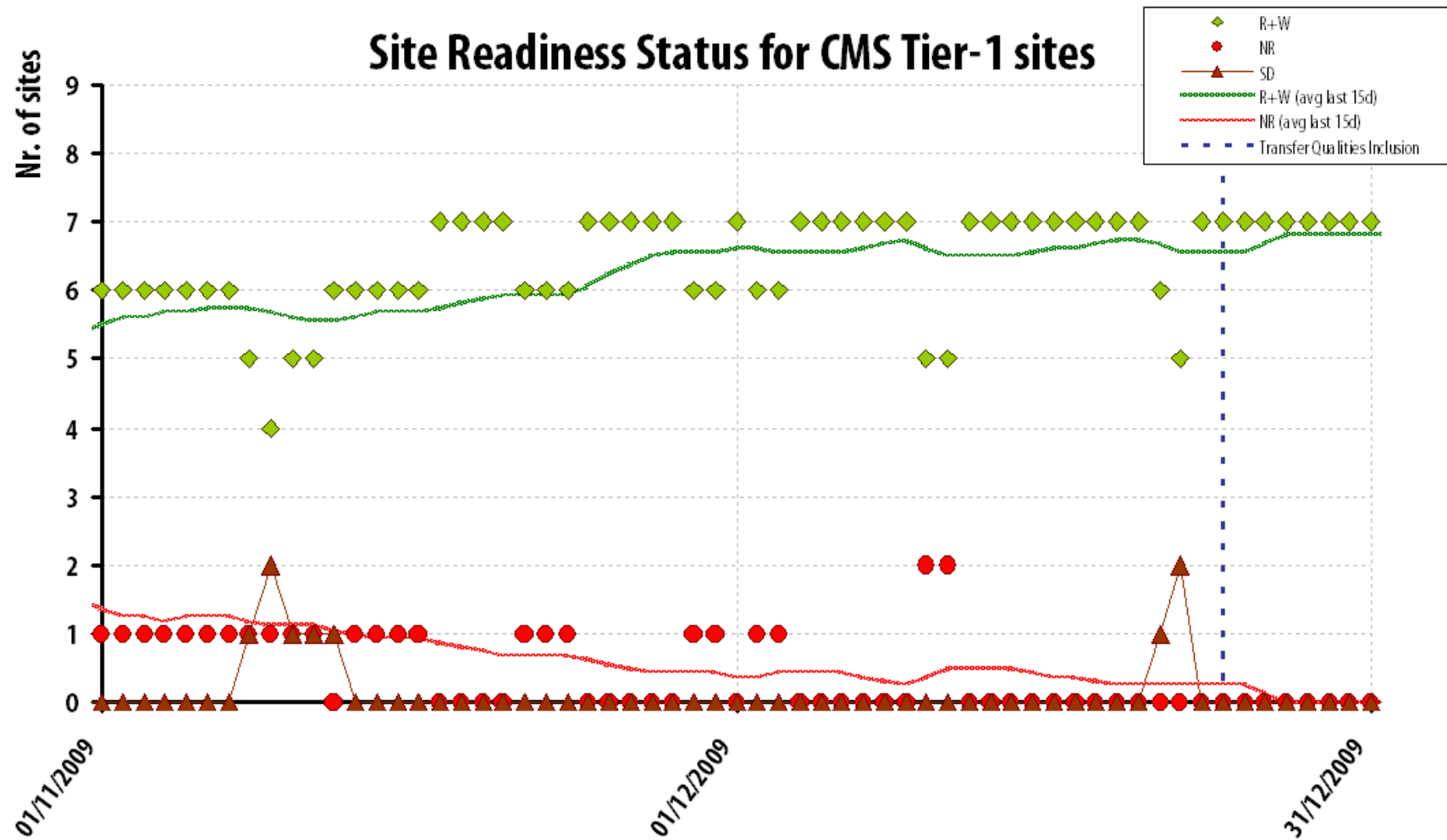
- Current plots compare favorably to view on Oct 2009
  - Readiness should be considered the starting point
    - Resource availability
      - Tapes, SL5 CPUs, Working node scratch disk, staged data, etc. all become important as we enter beam operations
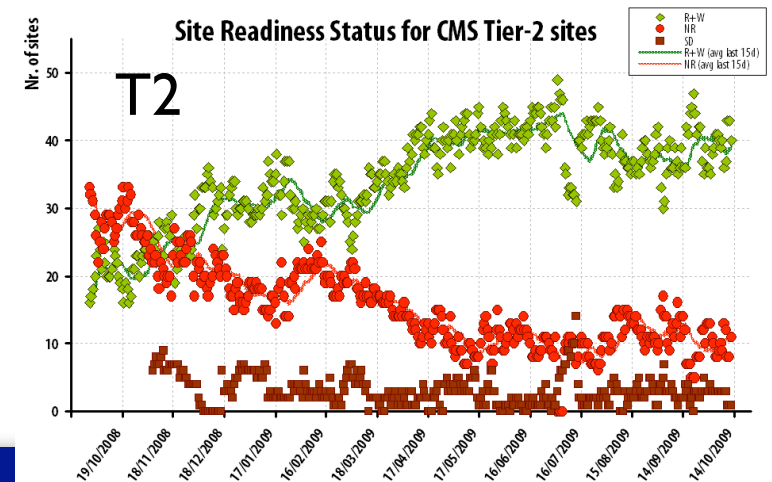   - All the s... nice job from the sites passing the readiness metric

Site Readiness Status for CMS Tier-1 sites
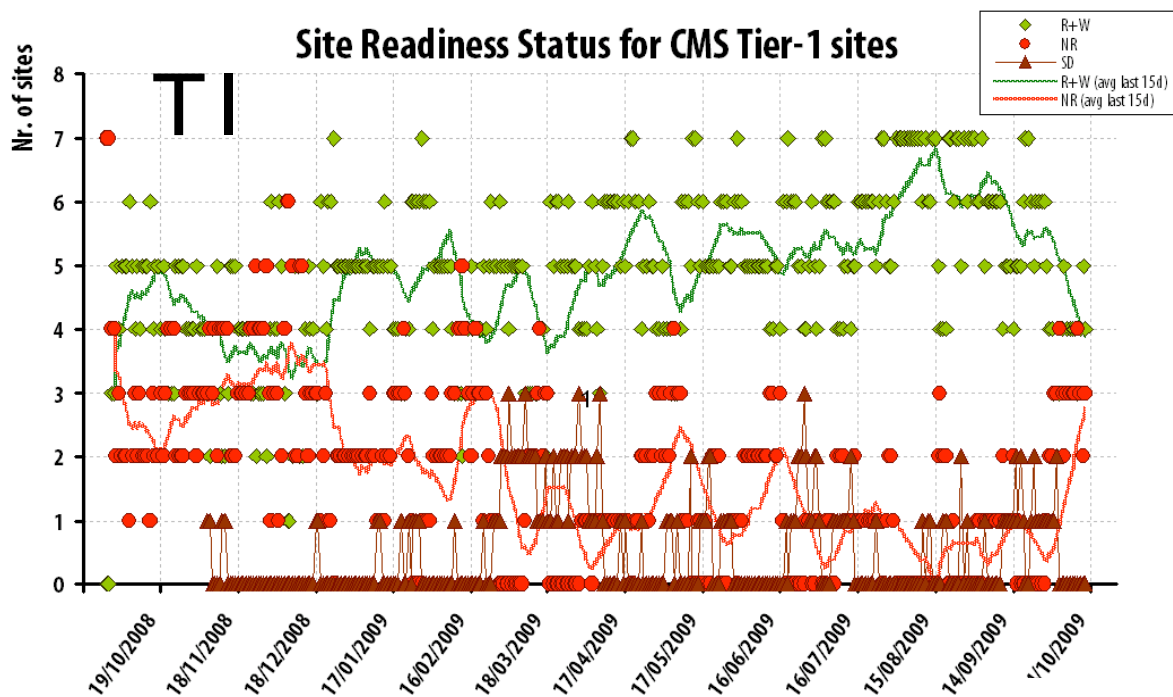
Site Readiness Status for CMS Tier-2 sites
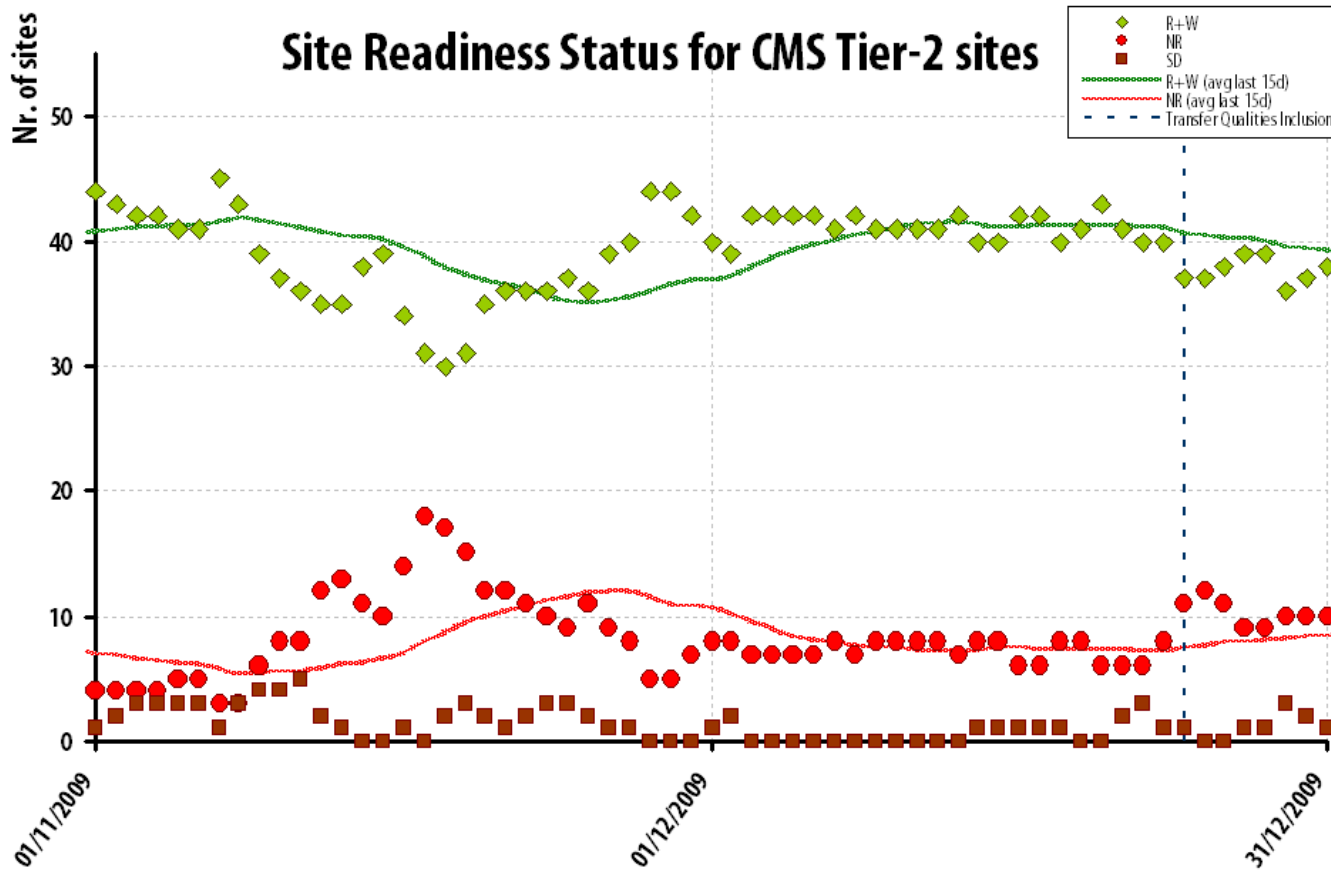
Looking back to Oct., Tier-2s have stabilized

CMS Services at Sites

- CMS software server
    - Software install automatically by CMS through Grid interface
        - Two teams (EGEE and OSG)
- Frontier / Squids
    - Used for distributed database infrastructure, stable and low effort
        - We have 1 Tier-1 and 10 Tier-2s that need to update to the latest release
- Local WN disk space (needed for caching input files - LazyDownload - and writing output), needs to be sufficiently dimensioned
    - CMS is working hard to restrict *all* file sizes to <10 GB and optimize workflows not to overfill WNs at the sites
        - Transfer limitations hit at 20GB
            - Ideally sites would have 20GB per core of local disk scratch

▶ PhEDEx Transfer service with correctly working Trivial File Catalog (TFC)

  ▶ The TFC is essentially a common name space for translating logical file names to physical file name

  ▶ Data Manager associated with the site to approve incoming transfer requests and monitor disk space

▶ CMS Needs space in the SE to store persistent experimental data for analysis and processing access.

  ▶ Additionally we need temporary space for output files before they are merged together and transferred to their final destination

   ▶ Temporary space is currently cleaned up by the local site.

Utilization

- The CMS Distributed Computing System generally performed well with the addition of collision data
  - The data rates and sample sizes are still quite low
  - The system was not resource constrained during this early period
  - The workflows and activities were generally what was expected from the computing model
    - Workflows executed much more frequently
      - Data Multiply Subscribed (More T1 and T2 subscriptions)
      - Re-processing occurred every 2-3 days
- Data Reconstruction, Skimming, Re-reconstruction at Tier-1s went nicely in parallel with distributed user analysis and MC production at Tier-2s

Data Collection Infrastructure

▶ Tier-0, Tier-1 Re-reco and Data Distribution Systems functioned with early collisions

   ▶ Events were reconstructed and exported to Tier-1 sites

   ▶ Express stream latency at target levels

   ▶ Re-reconstructed using Tier-1 centers

   ▶ Prompt Skimming system moved into production
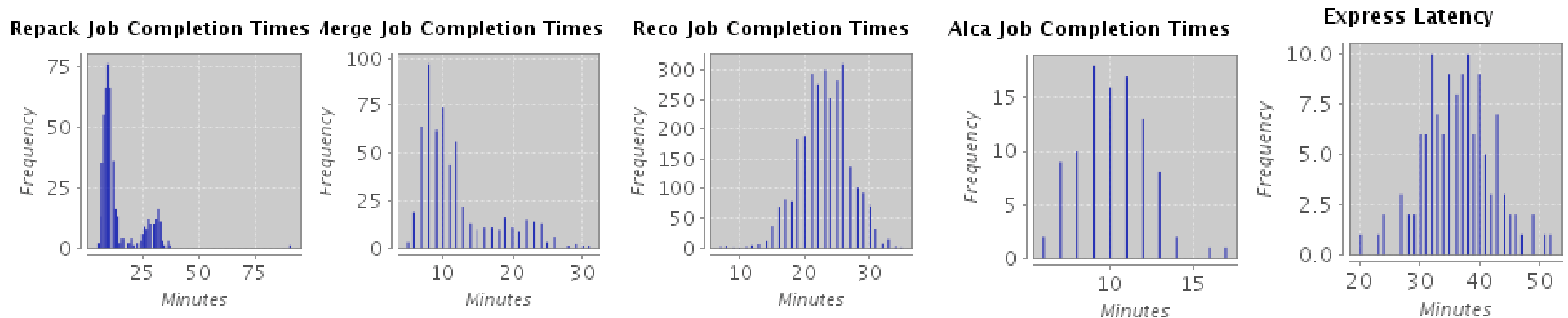
Computing Model in CMS

Processing and Analysis in CMS is performed performed on a globally distributed collection of computing facilities

Tier-1s Receive a fraction of the data divided by Dataset

Data is served to Tier-2s from multiple sources for analysis Tier-2s also produce MC

Tier-3 access is expected to be low impact
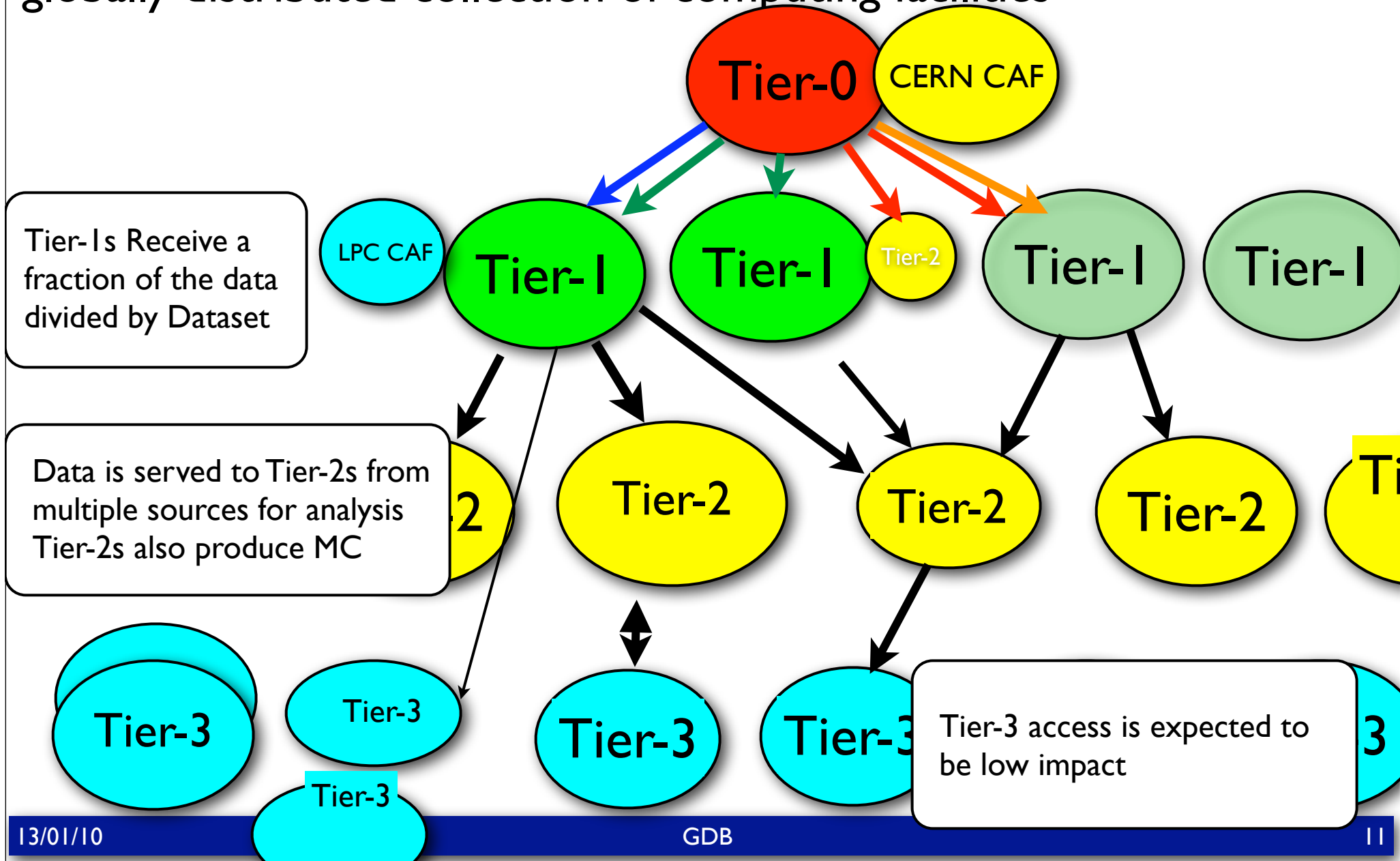
Data in first run

- The plan for higher luminosity collisions is to have ~10 primary datasets
  - There will be large variations in the volume in each, but working to ensure none is more than 30% of the nominal data rate
- In the first run there were several Primary Datasets, but only 2 with significant population
  - Minimum Bias and Zero Bias
    - Plus a back-up stream of high rate Zero Bias data
  - This impacts how we used the resources. Since we could be made multiple copies of the Minimum Bias data
    - Allows easy replication to Tier-2s

Readiness of T0 reconstruction

▶ Tier-0 Facility had been routinely exercised with cosmic data taking and simulated event samples

  ▶ Performing Stably with Cosmics

Job Type	Total Jobs	Failures	Success Rate
Express	342186	31	99.99%
Repack	134730	2	100.00%
PromptReco	38911	18	99.95%
AlcaSkim	41659	3	99.99%

▶ With Collisions (Failures concentrated in setup)

Job Type	Total Jobs	Failures	Success Rate
Express	404546	9442	97.72%
Repack	86982	69	99.92%
PromptReco	209773	2875	98.64%
AlcaSkim	17631	431	97.61%

4 9's

▶ ~3000 cores

  ▶ Local submission to farm with multiple workflows

  ▶ Good stability and performance of CMS software

▶ Received confirmation from CERN on T0+CAF pledge in 2010

CMS PhEDEx - Transfer Rate
132 Hours from 2009-12-11 07:00 to 2009-12-16 19:00 UTC

Source CERN
or Tier-1 going to
destination Tier-1
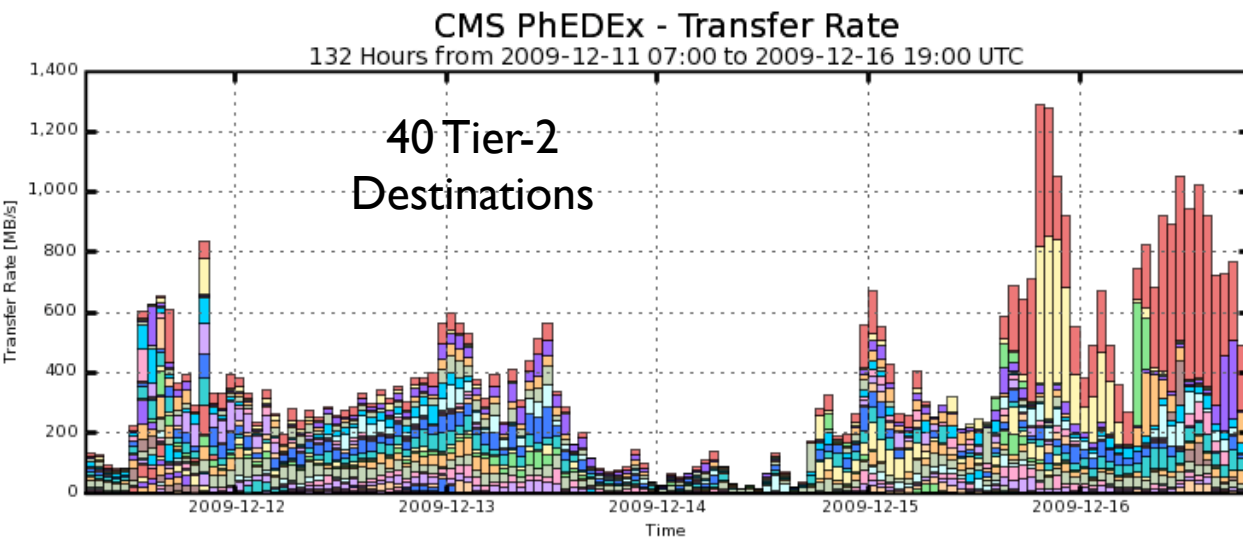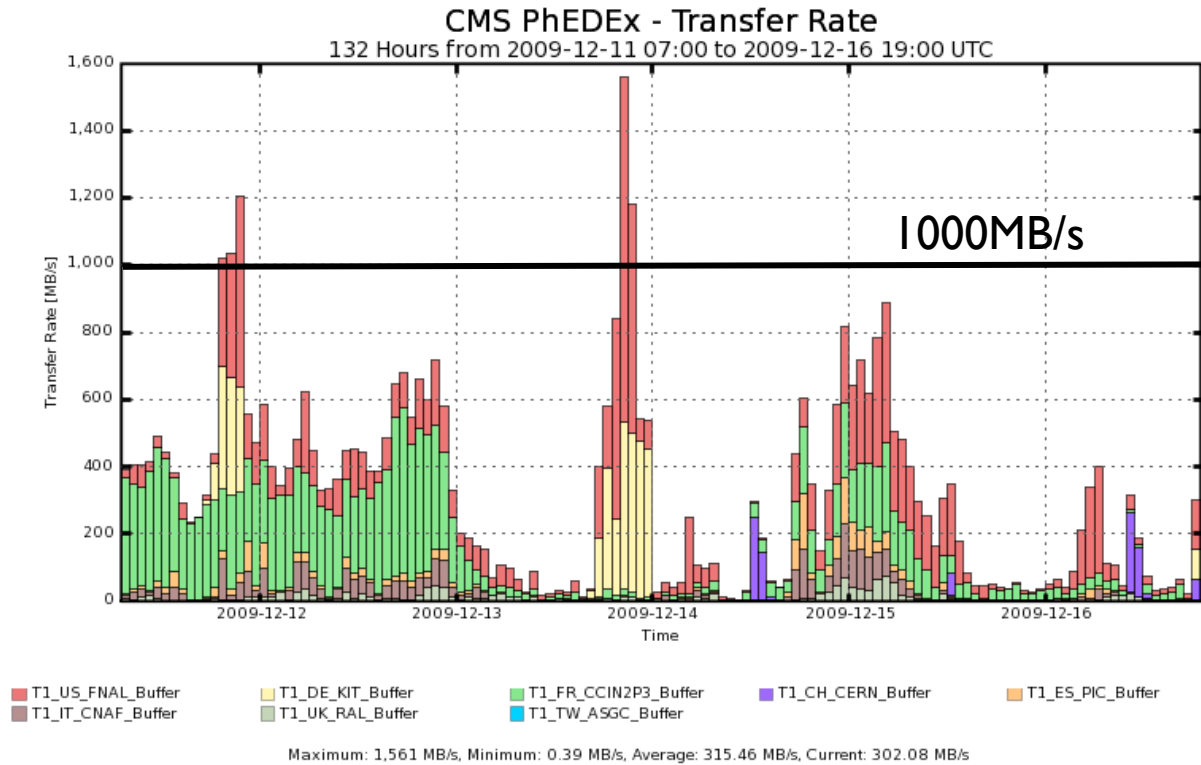
Source Tier-1
going to
destination Tier-2

We subscribed the MinBias primary dataset to PIC between the 27 to 28th of November

- Transfer system balanced the load to destination Tier-2s
- Good performance from both sites

Completed Jobs T2

100k

Dec Analysis Jobs (1/12/09 - 16/12/09)

75% success

Running Jobs T2

10k

/MinimumBias/BeamCommissioning09-PromptReco-v2/RECO-
...Bias/Summer09-STARTUP3X_V8D_900GeV-v1/GEN-SIM-RECO-
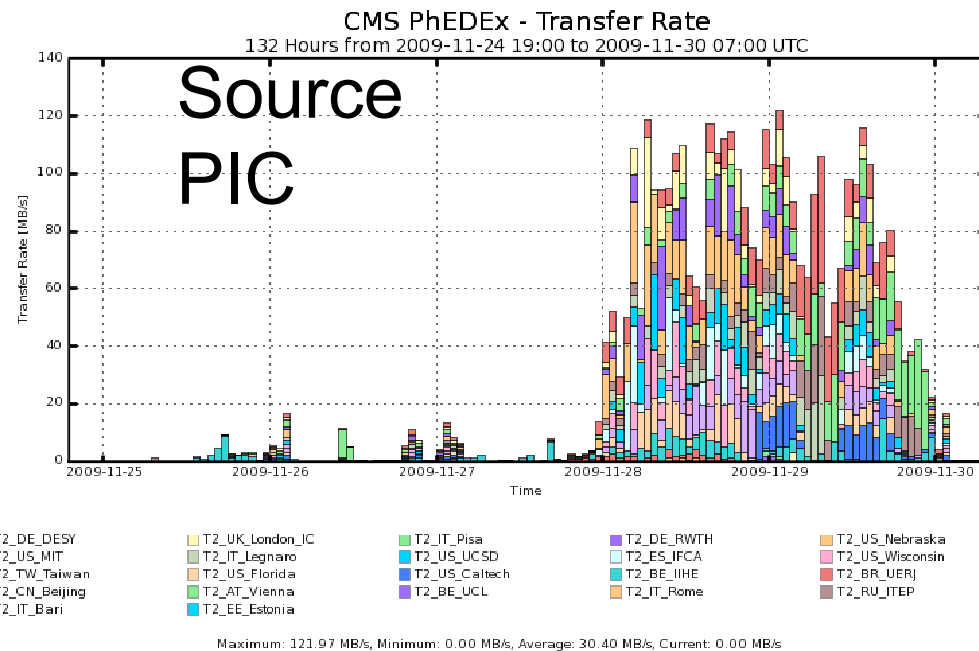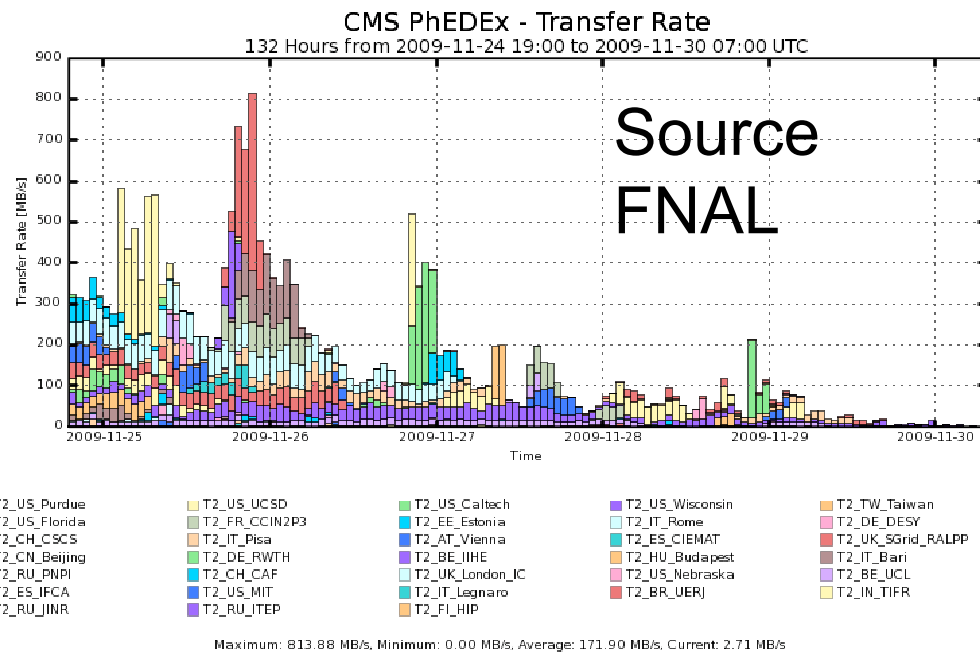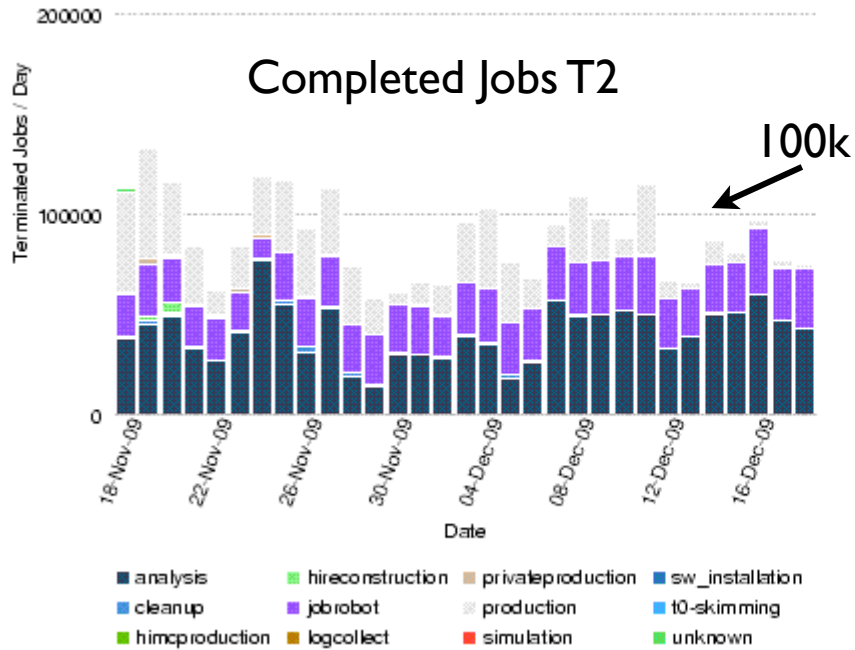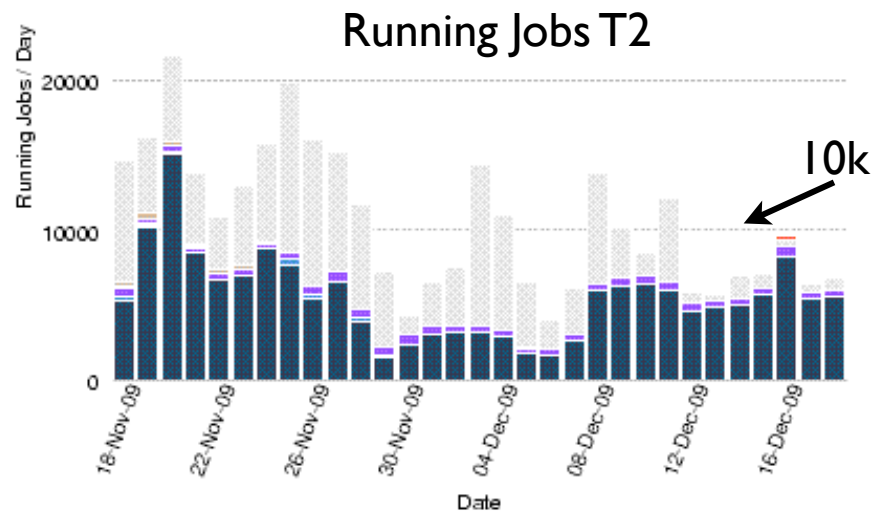/ExpressPhysics/BeamCommissioning09-Express-v2/FEVT-
/ppMuXLoose/Summer09-STARTUP3X_V8F_2360GeV-v1/GEN-SIM-RECO-
/MinBias/Summer09-STARTUP3X_V8D_2360GeV-v2/GEN-SIM-RECO-
/MinBias/Summer09-STARTUP3X_V8I_900GeV-v2/GEN-SIM-RECO-
/MinimumBias/BeamCommissioning09-Dec9thReReco-v1/RECO-
/MinimumBias/BeamCommissioning09-SD_AllMinBias-PromptSkimCommissioning_v2/RAW-RECO-
/MinimumBias/BeamCommissioning09-v1/RAW-
/QCD100to250-madgraph/Winter09_IDEAL_V12_FastSim_v4/GEN-SIM-DIGI-RECO-
/MinBias900GeV/Summer09-MC_31X_V3-v1/GEN-SIM-RECO-
/MinBias/Summer09-DESIGN_3X_V8A_900GeV-v1/GEN-SIM-RECO-
/QCD_Pt30/Summer09-MC_31X_V3-v1/GEN-SIM-RECO-
/PhysicsMuonBkg/BeamCommissioning09-PromptReco-v2/RECO-
/MinimumBias/BeamCommissioning09-rereco_GR09_P_V7_v1/RECO-
Jets-madgraph_redigi_ns_3_PATLayer1WithTtGenEvt-5806ec35aac362025b4eea830982d55c/USER-
/ZZ_4l_7TeV/ndefilip-ZZ_4l_7TeV-51eb09987b29e531a04b294fa76f903f/USER-
/ZJets-madgraph/Summer09-MC_31X_V3_7TeV-v2/GEN-SIM-RECO-
/WZ/Summer09-MC_31X_V3-v1/GEN-SIM-RECO-
/QCD_Pt80/Summer09-MC_31X_V3-v1/GEN-SIM-RECO-
p-CMSSW_2_2_7-10TeV_Skim10_glb_RECOSIM_IDEAL-28ea0c66474e22ac041d2f9cff5a280d/USER-
/MinBias900GeV/Summer09-MC_31X_V3-v1/GEN-SIM-RAW-
/MinBias/Summer09-DESIGN_3X_V8A_2360GeV-v1/GEN-SIM-RECO-
/QCD_EMEnriched_Pt20to30/Summer09-MC_31X_V3-v1/GEN-SIM-RECO-

Legend:
- analysis
- hireconstruction
- privateproduction
- sw_installation
- cleanup
- jobrobot
- production
- t0-skimming
- himcproduction
- logcollect
- simulation
- unknown

Activities over the break

- Data Processing Activities during the break

  - Re-processing and skimming of all good runs finished on 12/24 for the two large physics datasets

    - ZeroBias 22M RAW events, 1019 files processed

      - 11TB produced, 112M events in Secondary Datasets, AlcaReco etc

    - MinimumBias RAW 21.5M events, 1207 files processed

      - 10TB produced, 74M events in Secondary Datasets, AlcaReco etc distributed

      - Processed for two software releases (on SL5 and SL4)

        - Re-processing of MC datasets finished on 12/25

          - 20M MinimumBias

      - Re-processing of Cosmics MC finished on 12/25

        - 130M events

    - Almost problem-free processing of high-quality data

    - e.g. for the latest CMSSW version only one of >2000 job failed due to memory consumption all was done within 4-5 days

MC Production

▶ Smooth MC Production over break

Events written per site
21 Days from 2009-12-18 to 2010-01-08
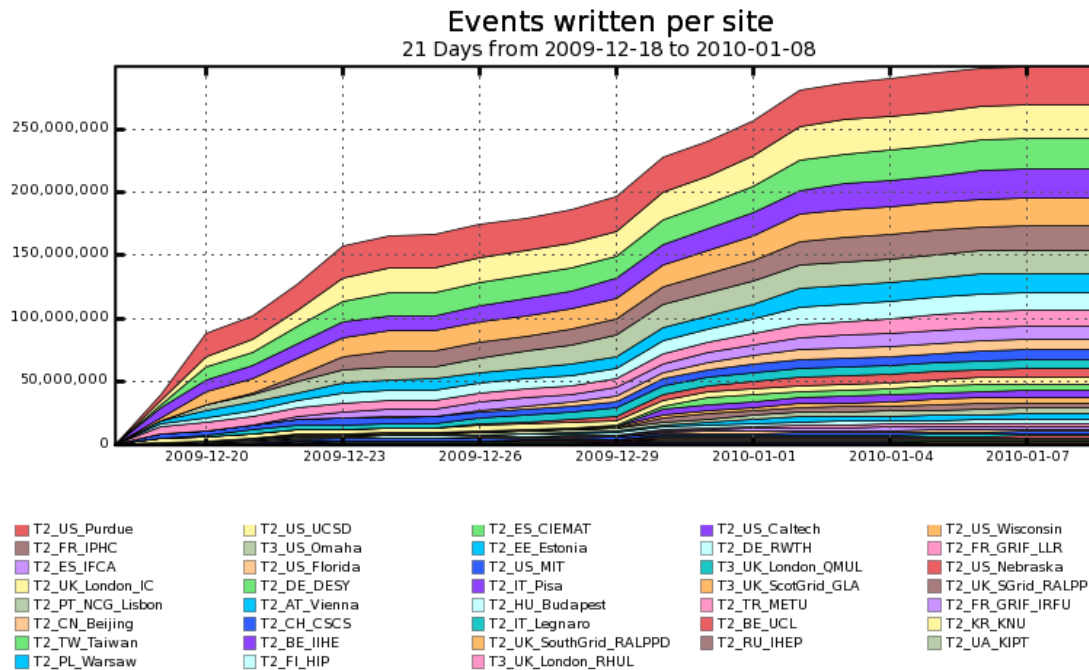
★ some 120M events produced (RAW, RECO, AOD)

✦ including special MinBias samples for comparison with 900GeV and 2.36TeV data

✦ most FullSim, some FastSim

Improving Network

- The CMS Computing TDR defines the burst rate Tier-1 to Tier-2 as 50MB/s for slower links up to 500MB/s for the best connected sites

  - We have seen a full spectrum of achieved transfer rates
    - Average Observed Daily Max peaks at the lower end

- From the size of the facilities and the amount of data hosted, CMS has planning estimates for how much export bandwidth should be achievable at a particular Tier-1

  - No Tier-1 has been observed to hit the planning numbers (though a couple have approached it)

  - CMS would like to organize a concerted effort to exercise the export capability
    - Need to work with site reps, CMS experts, FTS and Network experts
      - Area for collaboration

News

- CMS is using a Pilot Job Submission for a lot of the re-reconstruction work at Tier-1s

  - Based on an co-developed tool with OSG called the GlideIn-WMS

  - Scale and performance look good

- A GlideIn based CRAB server for analysis submissions is also in production

  - Tier-2s will also see multi-user pilot jobs

  - The majority of the analysis submissions are still gLite WMS directly submitted

- The first SL5 only production release of the CMS software was installed in late December

  - We expect data taken in Feb/March will be reconstructed only with this release, which will mean all sites participating in data analysis need to complete the move to SL5

Outlook

▶ We need to gain experience with the system with more data

  ▶ Early indications are that the system can work and the many tests with simulated activities have been representative

  ▶ Looking forward to Feb/March and driving to higher rates and more interesting activities