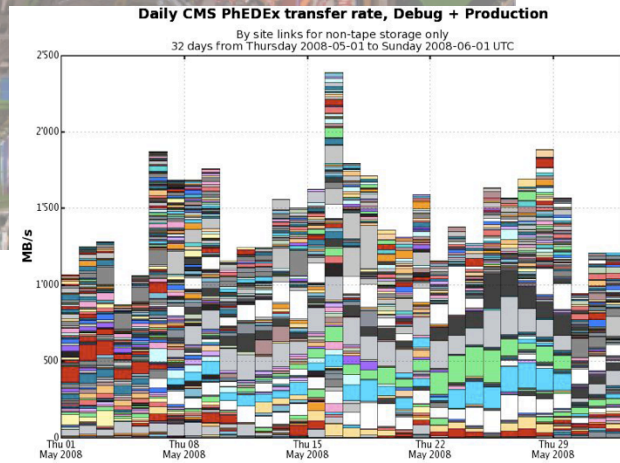
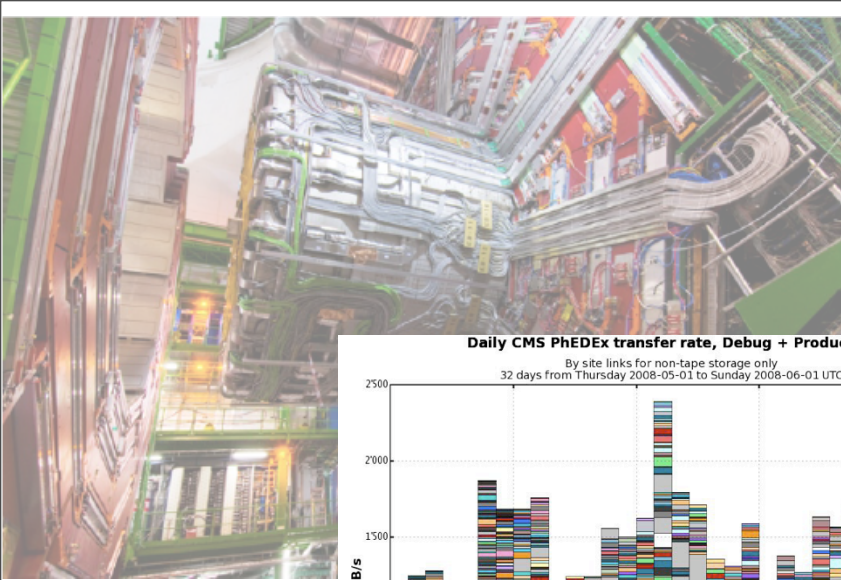


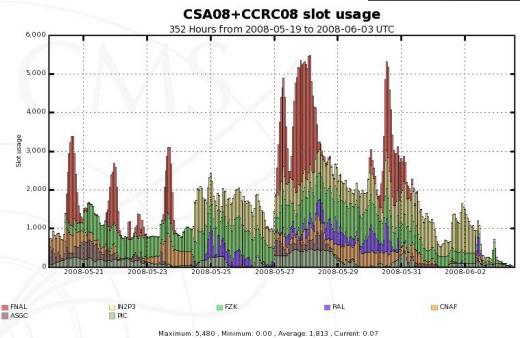
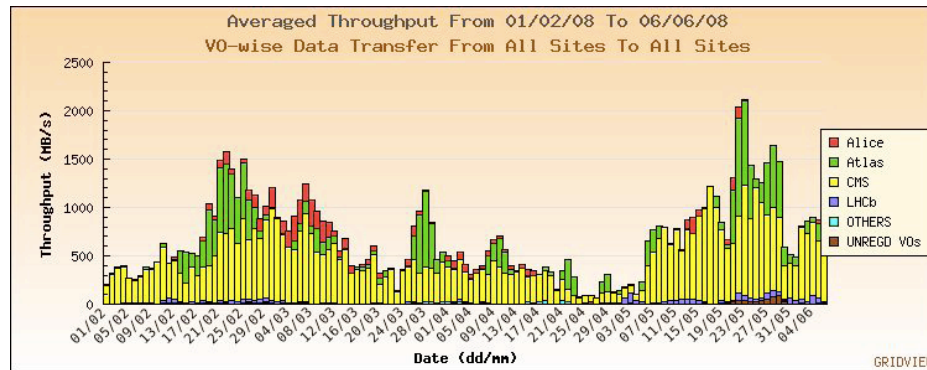
CMS: CCRC'08 and readiness for data taking

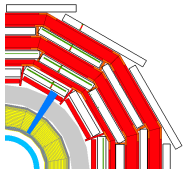


Patricia McBride
Fermilab/CMS

July 1, 2008

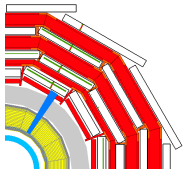
LHCC/LCG mini-review



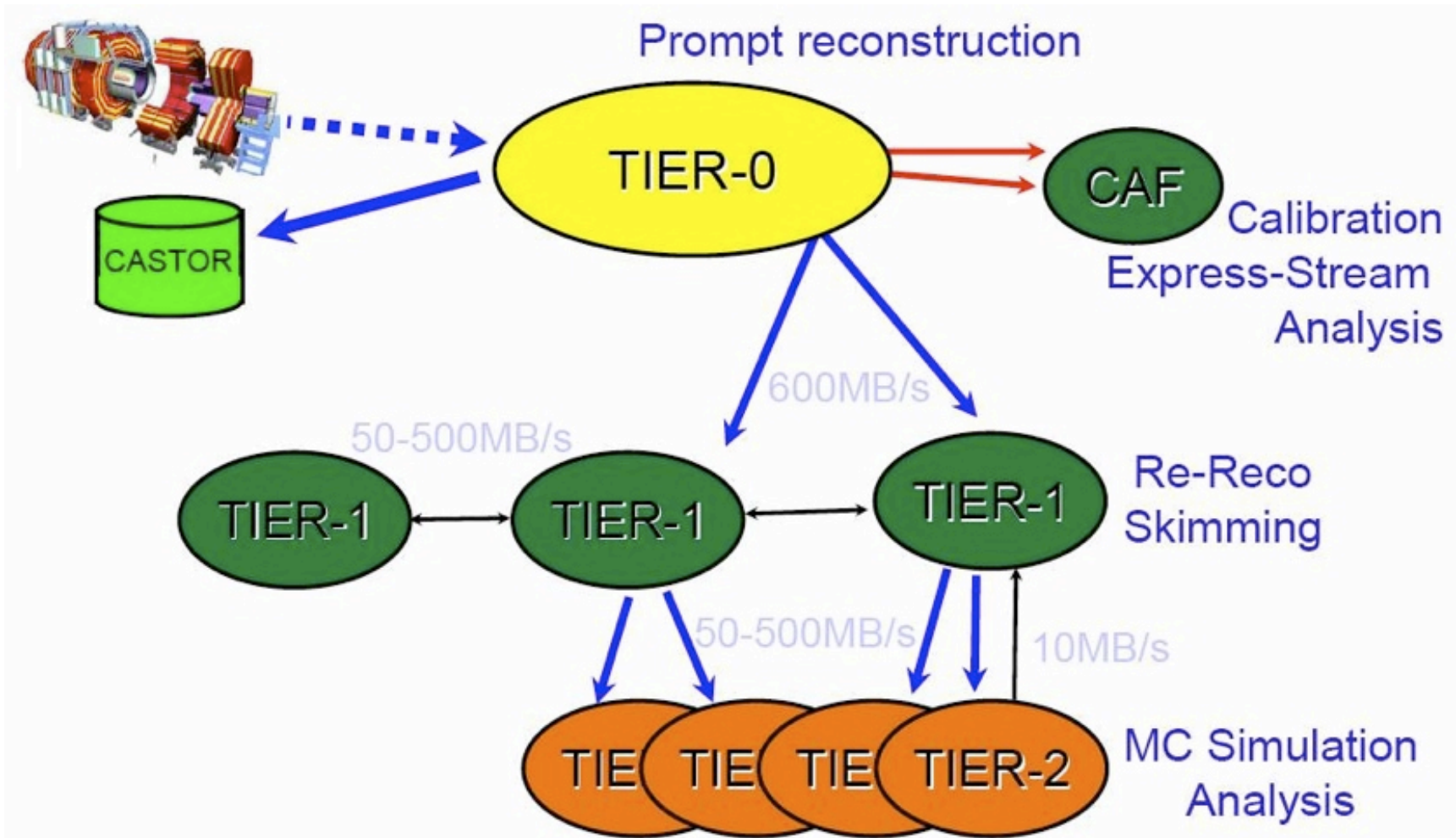


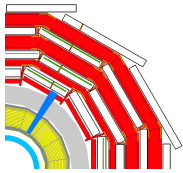
Outline

- Introduction
- CSA08
- CCRC (May)
- Global Runs (Cosmics @ P5)
- Readiness



CMS Computing Workflow

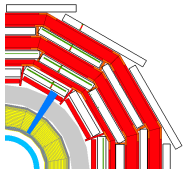




Tests of CMS Computing (CCRC/CSA08)

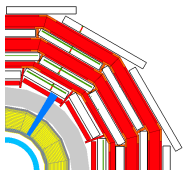
- Concurrent *CSA08* and *CCRC08* challenges in May (superimposition)
- *CSA08* driven by commissioning and physics goals
 - ◆ Higher priority
 - ◆ MC pre-production, reconstruction, calibration & alignment, re-reconstruction
- *CCRC08* driven by computing goals
 - ◆ Complement and add to computing aspects of *CSA08*
 - ◆ Include Data skimming, user analysis
 - ◆ Scale and end-to-end tests of the computing system
 - ◆ Many functionality/performance tests run in *CCRC08* phase I (Feb)
 - ◆ Run CMS computing workflows concurrently
 - ◆ Run concurrently with other LHC VOs

All the details can be found in D. Bonacorsi summary at the *CCRC* workshop:
<http://indico.cern.ch/contributionDisplay.py?contribId=1&confId=23563>



CSA08

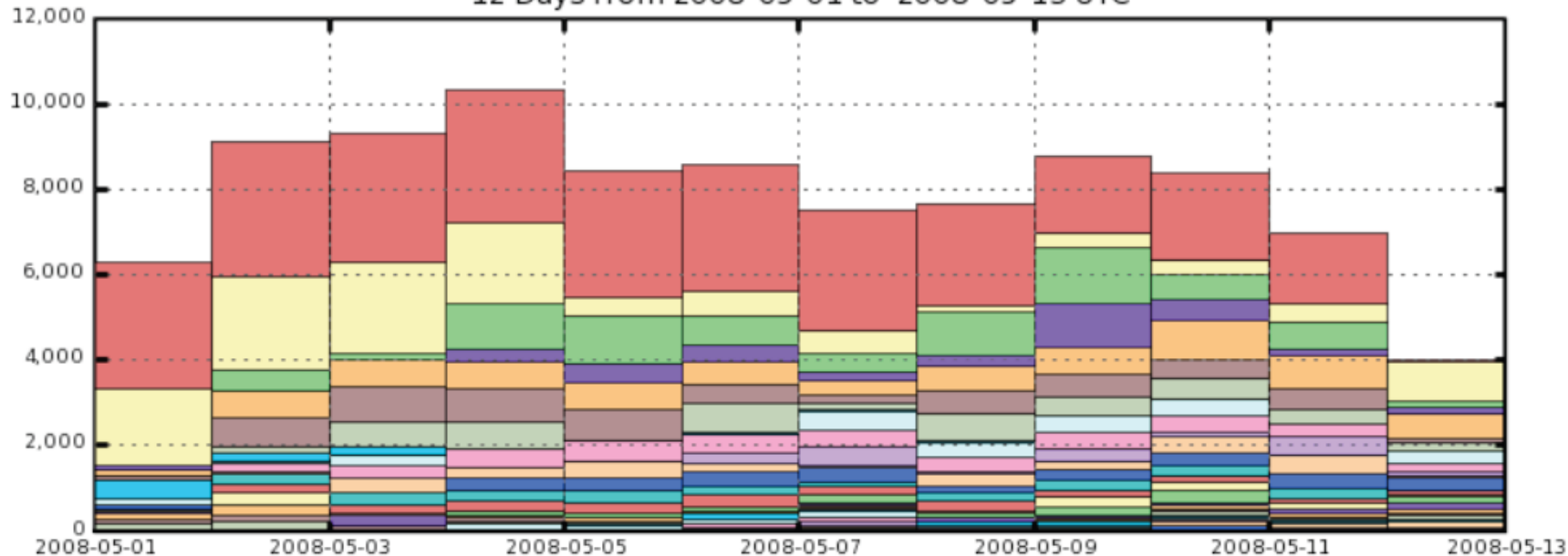
- **Aim of CSA08: Test the full scope of offline data handling and analysis activities needed for LHC data-taking operations at CMS in 2008**
- **Focus is on offline detector commissioning, with calibration and alignment forming the central component**
- **For the first time, a full-scale challenge with large statistics under conditions similar to LHC startup**
 - ♦ Initial mis-alignments and mis-calibrations (before collisions)
 - ♦ Event signatures and rates typical for low luminosity
 - ♦ Full complexity of almost 20 concurrent alignment & calibration end-to-end workflows (with interdependencies)
- **Operations during CSA08 were “run-like”**
 - ♦ Regular use of CMS centre
 - ♦ Daily run meetings, database operations, constants validation & sign-off
 - ♦ Schedule matched to pace of reconstruction: alignment and calibration in “real-time”, with resulting constants used to re-reconstruct the data ready for physics analysis



CSA08 Monte Carlo pre-production (full sim)

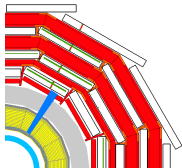
Approx batch slot usage

12 Days from 2008-05-01 to 2008-05-13 UTC

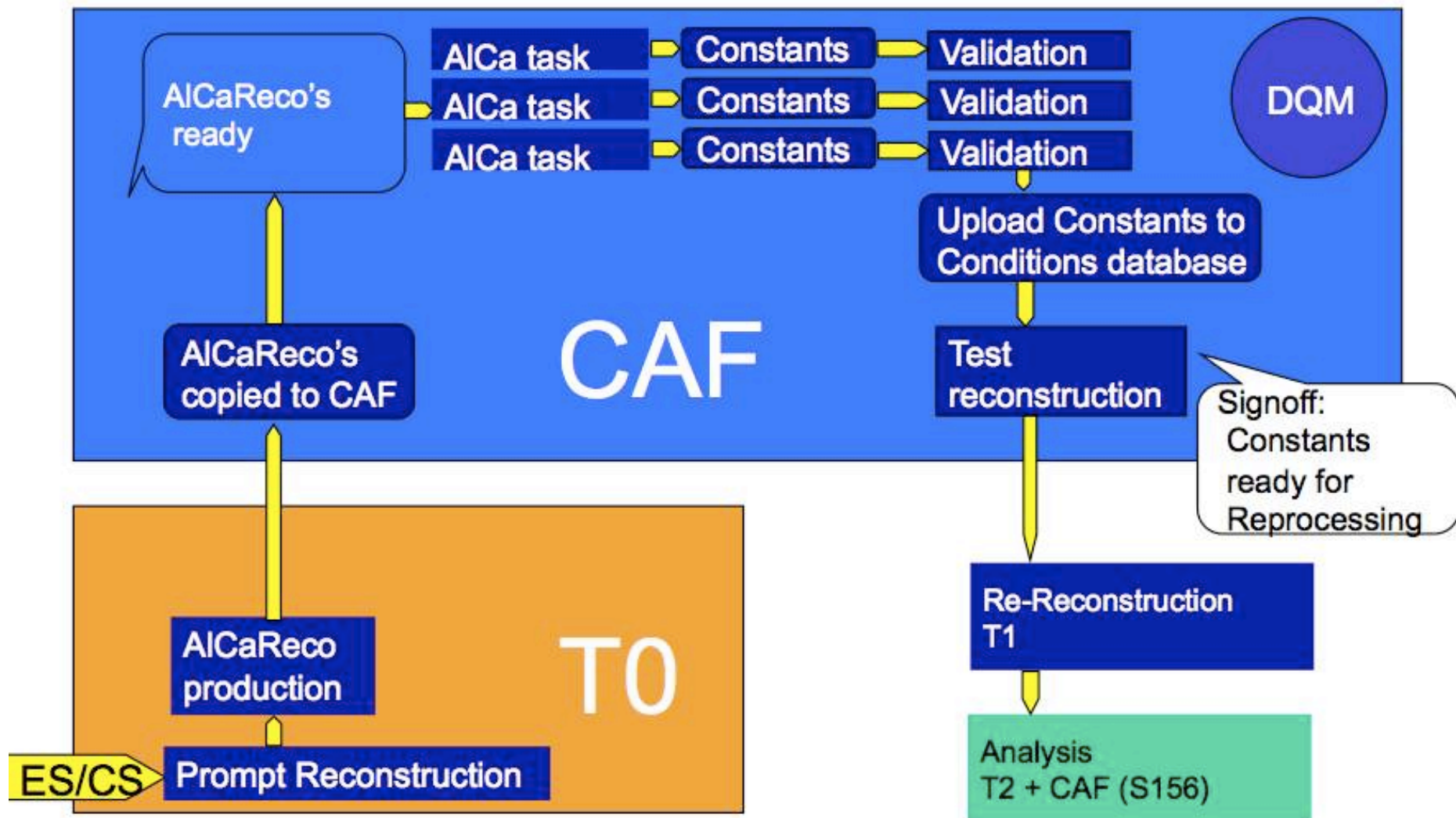


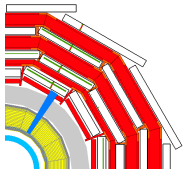
- | | | | | |
|---------------|-----------------|---------------------|---------------|-------------------------|
| T1_US_FNAL | T0_CH_CERN | T2_US_Purdue | T1_DE_FZK | T2_US_Wisconsin |
| T2_US_Florida | T2_US_MIT | T1_UK_RAL | T1_FR_CCIN2P3 | T1_IT_CNAF |
| T1_ES_PIC | T2_US_Nebraska | T1_TW_ASGC | T2_US_Caltech | T2_ES_CIEMAT |
| T2_US_UCSD | T1_DE_FZKK | T2_FR_CCIN2P3 | T2_DE_RWTH | T2_UK_SouthGrid_RALPPD |
| T2_DE_DESY | T2_IT_Legnaro | T2_HU_Budapest | T2_ES_IFCA | T2_IT_Pisa |
| T2_AT_Vienna | T2_BE_IHHE | T2_UK_London_Brunel | T2_IT_Bari | T2_UK_London_Imperial |
| T2_TW_Taiwan | T2_CN_Beijing | T2_CH_CSCS | T2_BE_UCL | T2_UK_SouthGrid_Bristol |
| unknown | T2_UK_London_IC | T2_FR_GRIF_LL | Wisconsin | Purdue |
| Nebraska | | | | |

On average ~8000 concurrent jobs, at all Tiers: T0/T1/T2
 MC production is routine and on-going in CMS

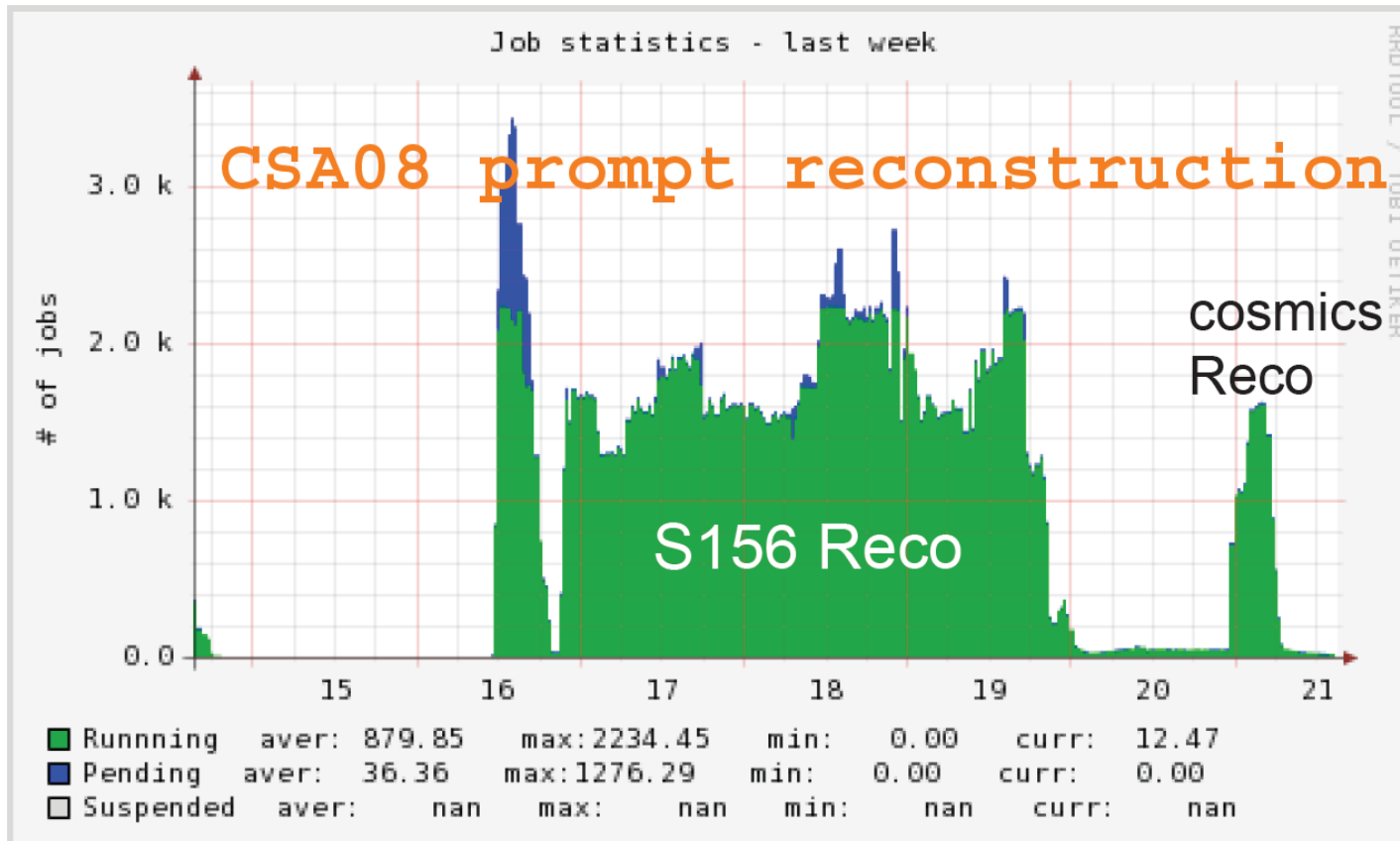


CSA08 T0/CAF workflow: Alignment and Calibration

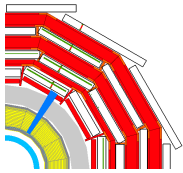




CSA08 Processing at the Tier-0



150M events reconstructed in less than 4 days!



CSA08 Schedule

Week 1 (5-11 May):

S43 prompt reconstruction at T0 started on time, completed 15th May

Week 2 (12-18 May):

S43 prompt calibration/alignment exercises on CAF On time

S156 prompt reconstruction at T0 → started 16th May, completed 19th May

Week 3 (19-25 May):

S156 prompt calibration/alignment exercises on CAF On time

S43 re-reco at T1s using constants from S43 exercises → started on time

Week 4 (26th May - 1st June):

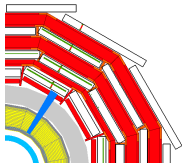
S156 re-reco at T1s using constants from S156 exercises On time

Physics analyses using prompt data at CAF On time

All CSA08 production operations completed on time by deadline

midnight 2nd June - Physics analyses continuing at CAF/T2s

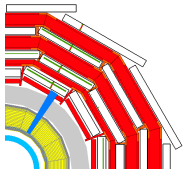
CSA08 was a successful test for calibration/alignment and computing!



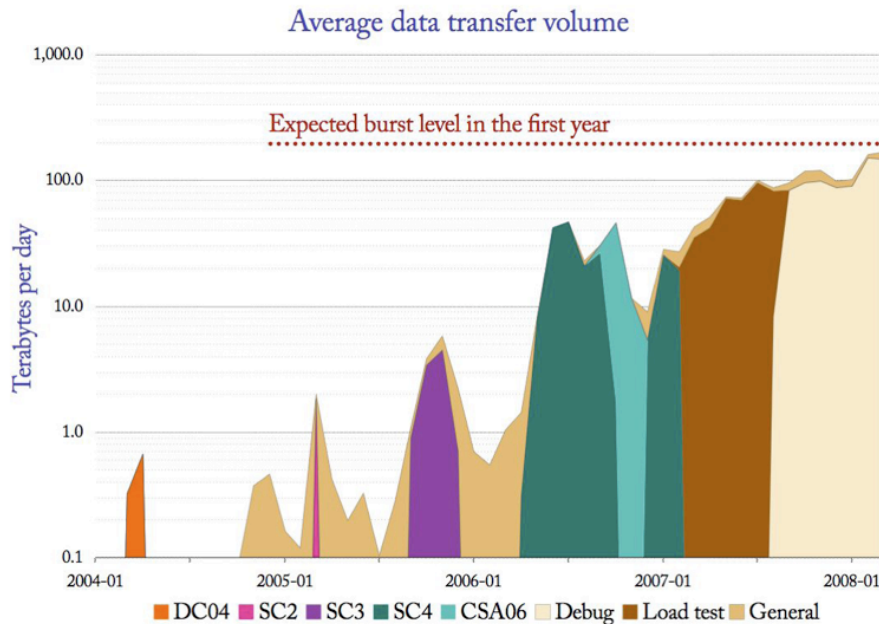
CCRC (May 2008)

	Week 18	Week 19	Week 20	Week 21	Week 23	CSA/CCRC Ends
Tier-0	PreProduction	S43 Prompt Reco and dataset transfer to CERN	S156 Prompt Rec	CCRC08 end-to-end tests		
CAF		DataSets arrive	S43 alignment and calib	S156 alignment and calib		
			S43 User Analysis	S156 User Analysis		
Tier-1	PreProduction			S43 ReReco	S156 ReReco	
				CCRC08 scale tests, Skimming		
Tier-2	Phase 0 -Prep	PreProduction		Other MC Production		
		Phase 1 - Centrally Organized Activities	Phase 2 - Chaotic analysis	Phase 3 - Final phase		
				CSA analysis		

- ◆ **Tier-0:**
 - preproduction MC, prompt-reco for CSA08
 - CCRC end-to-end tests (not yet complete)
- ◆ **CAF:**
 - Calibration and alignment exercises, user analysis for CSA08
- ◆ **Tier-1s:**
 - pre-production MC, re-reco for CSA08
 - CCRC Skimming
- ◆ **Tier-2s:**
 - pre-production MC for CSA08, other MC production (on-going)
 - CCRC T2 Analysis - in phases
 - CSA08 analysis
- ◆ **Data transfers (CCRC)**



CCRC Data Transfers

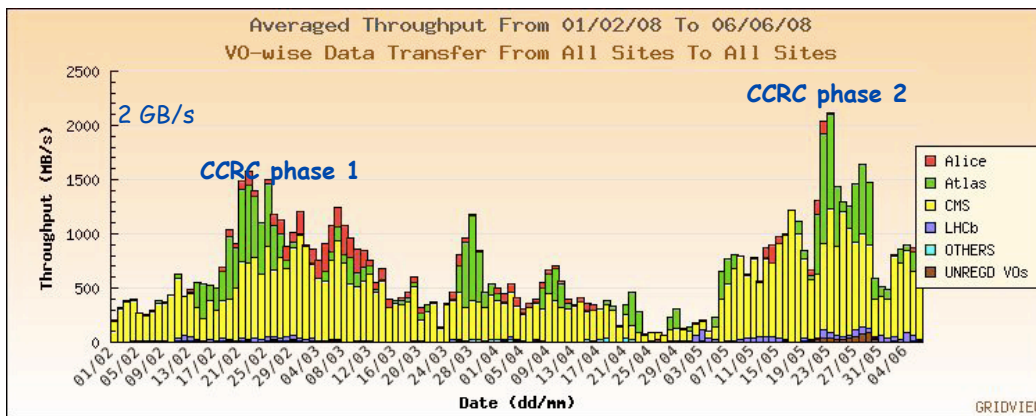
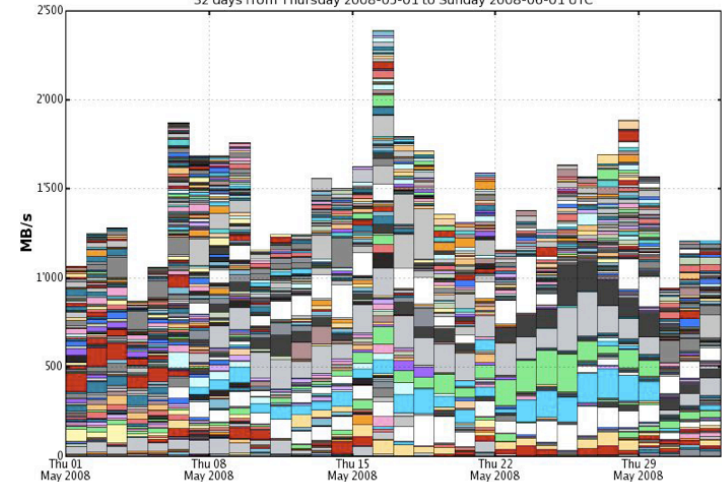


- We continually work to improve data transfers.
 - ◆ for CCRC used CSA08 data, CSA07 data and load tests

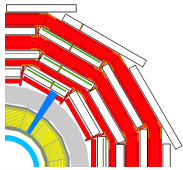
On average we transferred ~ 120 TB/day
Total of 3.6 PB in May

Daily CMS PhEx transfer rate, Debug + Production

By site links for non-tape storage only
32 days from Thursday 2008-05-01 to Sunday 2008-06-01 UTC



Data Transfers out of CERN



CCRC Data Transfers

Impressive list of few hundreds of links.



◆ T0-T1 Transfers:

Demonstrated sustained CERN export to all 7 T1s simultaneously with multi-VO traffic at 2008 nominal rates > 600 MB/s (RAW+promptRECO data in the computing model)

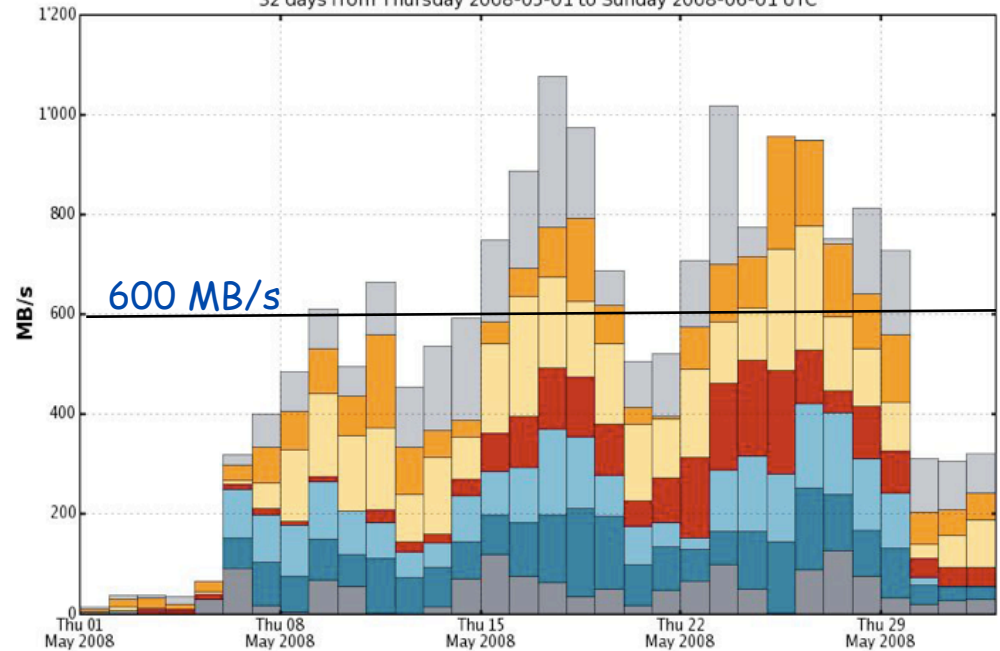
There is also a dedicated effort to commission links:

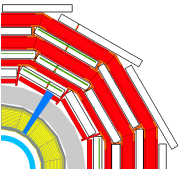
- T1 <-> T1 (full mesh)
AOD synchronization
- T1 -> T2 (test full mesh)
data access for physics
- T2 -> T1 regional
MC production

Continually raising the bar. Will expect sites to have commissioned links.

Daily CMS PhEDEx transfer rate, Debug + Production

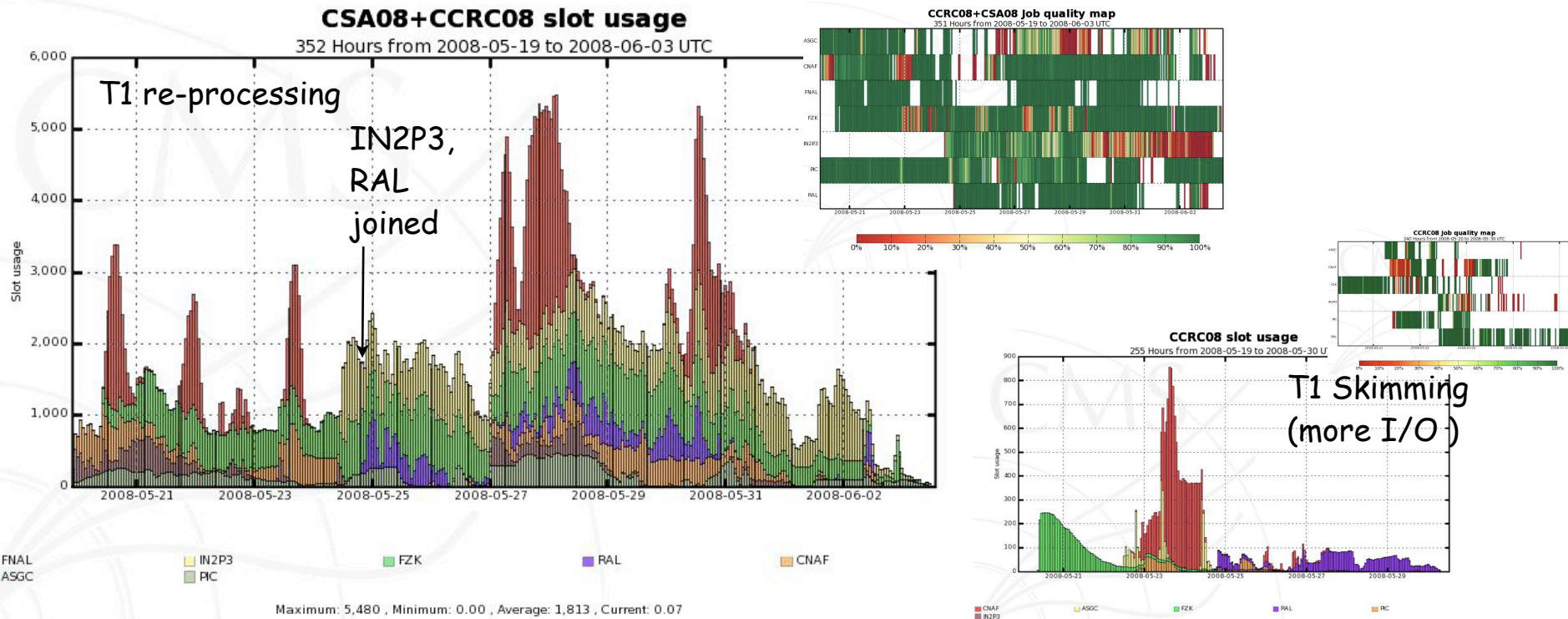
country to site links for non-tape storage only
Sources matching 'CERN.*' and destinations matching 'T1.*'
32 days from Thursday 2008-05-01 to Sunday 2008-06-01 UTC

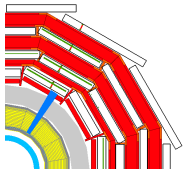




Re-processing and skimming at the Tier-1s

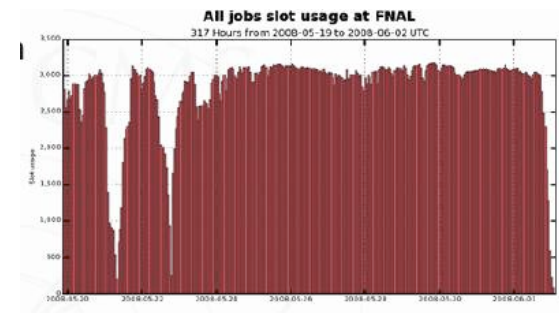
- ~ 200k jobs all 7 Tier-1 sites
- Reprocessing at the Tier-1s ~ 250M re-reconstructed events (CSA08)
- ~ 800 slots FZK, ~ 400 PIC, ~ 500 IN2P3, ~ 300 ASGC, ~ 300 CNAF, ~ 3000 FNAL
- Large scale and sustained activity for two weeks



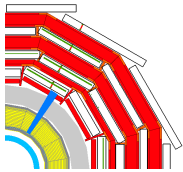


Summary: CCRC at the Tier-1s

- **Very successful exercises**
 - ◆ Large scale, re-reconstruction and skimming concurrently with data transfers
 - ◆ Very good performance of computing system and sites
 - ◆ Sites found it extremely useful (scaling, tuning, operating points)
 - ◆ Very useful for integration of production systems
 - Issues being address via focused campaigns (developers/ops/integration team)
 - Monitoring needs further integration and development
 - Data pre-stage from tape needs automation
 - gLiteWMS bulk job submission at EGEE sites. Run dedicated ProdAgent instances per Tier-1. JobTracking being parallelized
 - Pilot-based job submission at FNAL



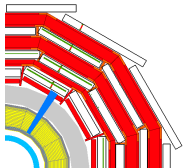
- **Tests of concurrent processing with other experiments pending**



CCRC: Testing Analysis at the Tier-2 Centers

<https://twiki.cern.ch/twiki/bin/view/CMS/CCRC08-AnalysisAtTier2>

- The objective: to test the performance and readiness of the Tier-2's for data analysis.
- The tests were done in 4 phases :
 - ◆ Phase 0 -- Preparation Phase
 - engage the sites and CMS integration team in testing the sites/software
 - ◆ Phase 1 -- Centrally Organized Activities
 - Understand a site's performance characteristics;
 - Exercise physics analysis workflows using CRAB.
 - ◆ Phase 2 -- Chaotic job submission from all T2s to all T2s.
 - Encourage user involvement to submit test jobs to other sites
 - ◆ Phase 3 -- Final Phase
 - Stop watch phase (measure performance and latency)
 - Local Scope DBS validation tests
- Detailed Summary: <https://twiki.cern.ch/twiki/bin/view/CMS/CCRC08-AnalysisAtTier2>



CCRC analysis at T2s; controlled submission

"Physics Group" analysis: job reads a RECO dataset and writes a user roottuple.

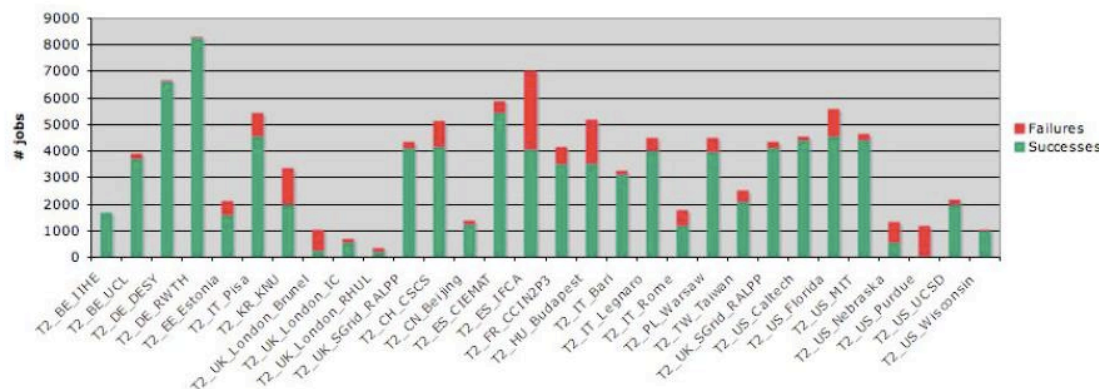
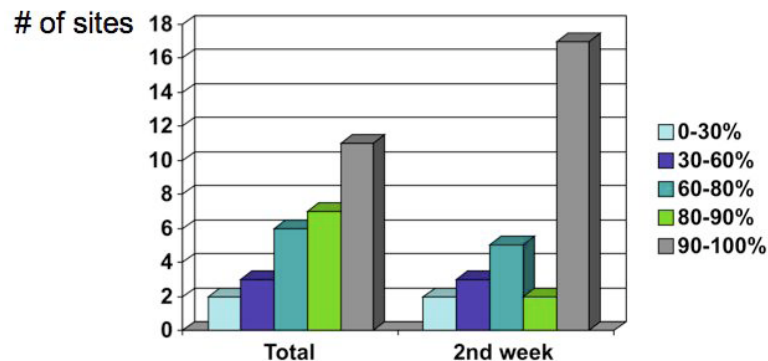
Jobs submitted by experts.

Tier2's successfully associated to each "Physics Group"

Started with 12 sites on May 6, ended with 29 sites on May 16

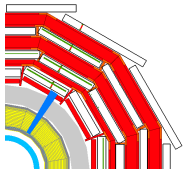
~ 105k submitted jobs May 6-16

Analysis-like 3-4 hrs jobs which ran and did remote stageout with SRMv2
O(10MB) transferred/job, to selected T2's: CIEMAT, Legnaro, RWTH



Aborted jobs reported to GGUS.

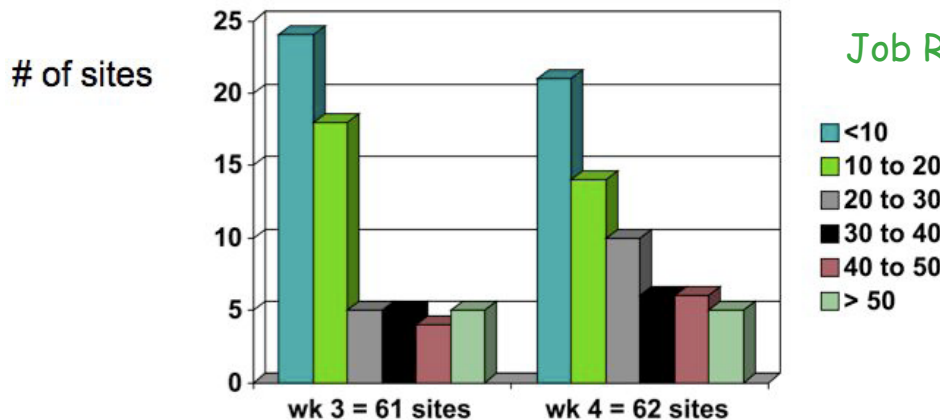
Often fast response.



CCRC Analysis at T2s; chaotic submission

- We encouraged people at all T2 sites to submit test jobs to other sites, with the output staged out to their "home tier-2".
- The primary goal here was to see many users would/could submit a significant number of successful CRAB jobs across many sites.
 - ◆ 450 active users in May
 - ◆ 1.5M jobs across 30+ Tier2 sites

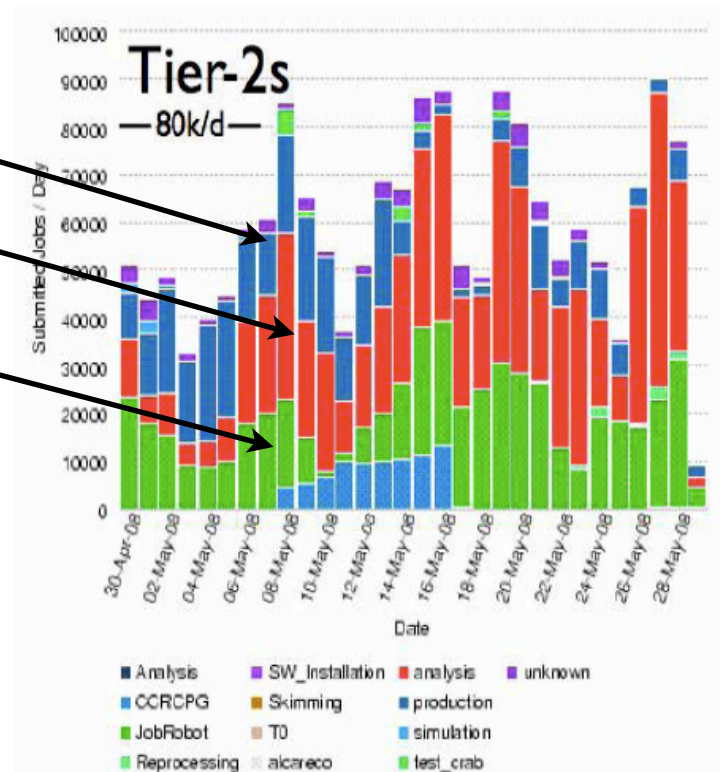
T2s during CCRC

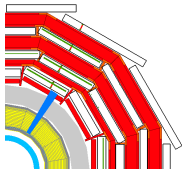


MC Production

Analysis

Job Robot





Short summary of job submission during CCRC

- from Daniele Bonacorsi

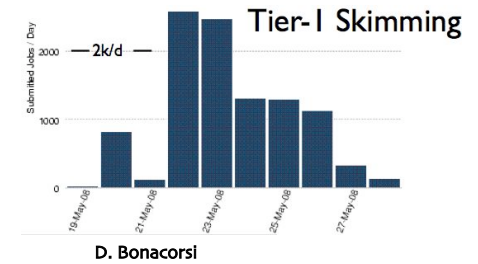
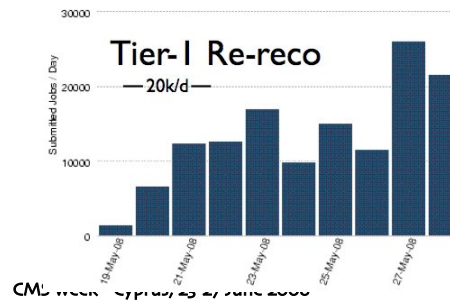
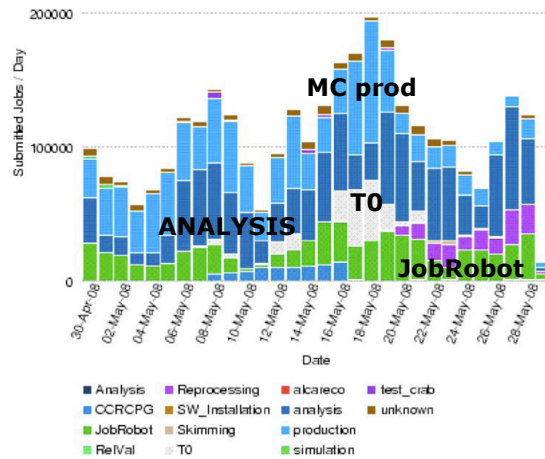
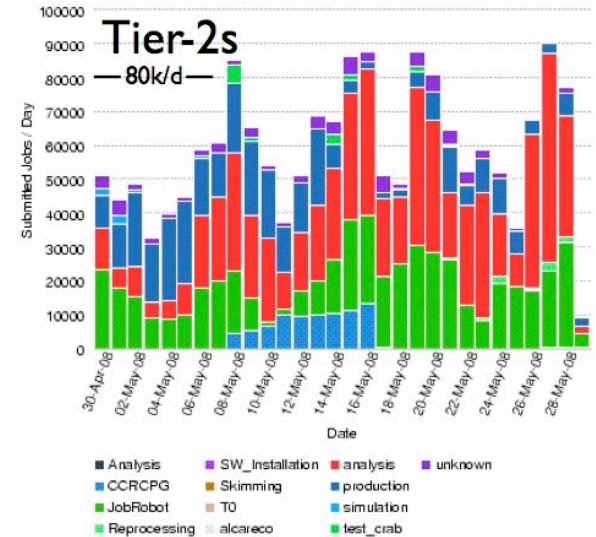
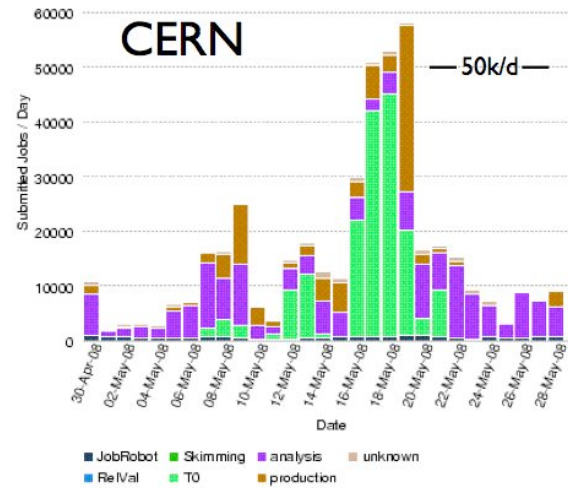
CMS WMS system:

Routine submission of ~100k jobs/day by CMS to all Tiers

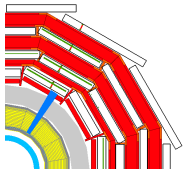
Peaks of 200k jobs/day seems okay.



CMS routinely submits the required # jobs/day

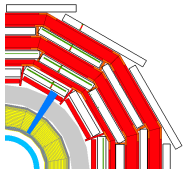


83



Summary of CCRC/CSA08 for CMS

- **Very successful CCRC08/CSA08 computing exercises**
 - ◆ Extremely interesting for sites, operations and computing systems
 - ◆ All data-and workflows of CMS computing system exercised at large scale
 - ◆ Data and workflows run concurrently
 - ◆ Cosmic data taking processed and transferred in parallel
- **Workload and data transfer systems scaling well**
 - ◆ >100k jobs/day submitted and >100 TB/day data transfers routinely sustained
 - ◆ Load generators and computing commissioning task forces essential
 - ◆ Issues identified and addressed via focused Dev/Int/Ops campaigns
 - ◆ Development of control system for fully automation of Tier-0 workflows well advanced. Plan for Tier-0 end-to-end tests
- **Will keep running scale tests**
- **Computing system in very good shape for data taking, processing and analysis**

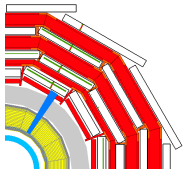


CMS Global Runs for Commissioning

- **CMS commissioning at P5 started in May 2007:**
 - ♦ “local” runs: single detector commissioning
 - ♦ “global” runs: integration runs for ~10 days per month.

- **Latest major global runs:**
 - ♦ GREN: Global Run End of November
 - ♦ GRUMM: Global RUn Middle of March
 - ♦ CRUZET1: Cosmic RUn at ZEro Tesla (5-9/5/2008)
 - ♦ CRUZET2: same as above- phase 2 (11-14/6/2008)

- **Services, workflows and dataflows are setup as close as possible to collision mode**
 - ♦ Test of TO workflows - data operations teams at CERN/FNAL



Data Flow and Processing for Global Runs

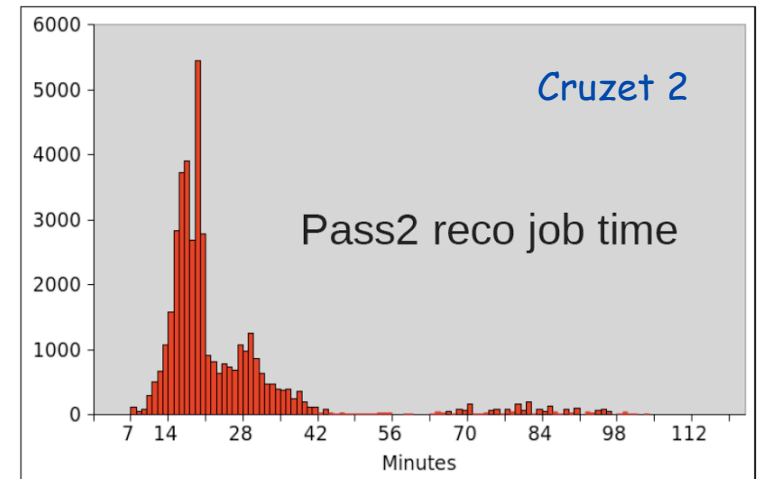
Increasing level of complexity with time:

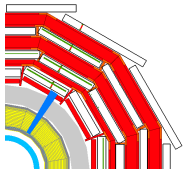
- ◆ single stream
- ◆ multiple streams (“Physics”, calibrations, DQM and HLTDebug)
- ◆ Primary Datasets (repacked at T0 using HLT info):

- /Cosmics/CRUZETx-vy
- /BarrelMuon/CRUZETx-vy
- /EndCapsMuon/CRUZETx-vy

“Prompt” reco:

- cronjob (starting every 15 mins/1h) checking availability of repacked RAW data, starting RECO and publishing datasets
- DQM offline included in first processing step

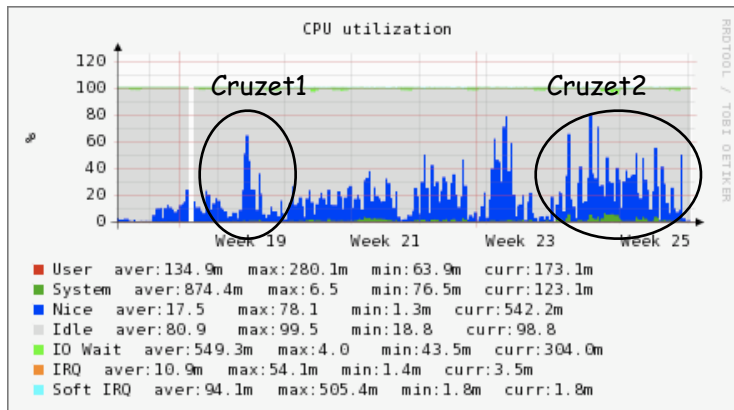




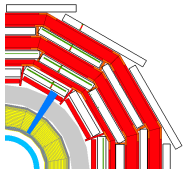
CMS Analysis Facility (CAF) for Commissioning

- **CMS CAF is heavily used for commissioning:**
 - ◆ 608 CPU slots; 541 TB CASTOR disk space
 - filled at a rate of ~10TB/day of GR
 - ◆ 150 GB AFS shared space/group
- **Main Usage:**
 - ◆ prompt analysis (data on disk)
 - ◆ “Local” detector DQM
 - ◆ Offline DQM harvesting

Period	CPU	Disk
2007	128 slots	35 TB
Feb-March 08	228 slots	172TB
April 08	328 slots	263 TB
May-June 08	632+48 slots == 1.4 MSI2K	540 TB
End-June 08	CAF-T1 : 1.4 MSI2k CAF-T2 : 1.0 MSI2k	CAF-T1: 1600 TB CAF-T2 : 0.2 TB
2008 Pledge	2.1 MSI2K	1800 TB



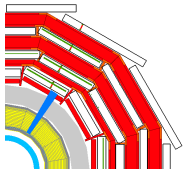
CPU usage on CAF is ramping up:
average 40%, peak 80%



Plans for Commissioning CMS

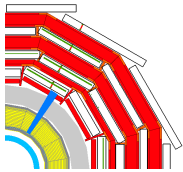
■ Summer Schedule (plan so far)

- ◆ End-to-end tests of Tier-0
- ◆ Site commissioning
- ◆ MC production with V2_1
- ◆ Tests at Tier-1s (with other VOs)
- ◆ Analysis at Tier-2s
- ◆ Support of Global Runs
 - CRUZET 3: July 7-14 with tracker
 - ➔ Exercise prompt-reco at T0
 - CRUZET 4: early August with Ecal Endcap and pixel. Possibly followed by continuous running while magnet ramps up and then few days at 3.8T (CRAFT).
 - ➔ Exercise final Tier-0 production system
- ◆ CMS continues to run - assume >50% operation until collisions at 900 GeV and then until collisions at 10 TeV.



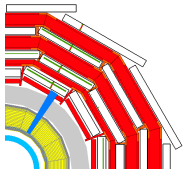
Where are we....

- Demonstrated all key use-cases in recent tests
 - ♦ I've skipped many details...
- TO workflows improving constantly
 - ♦ some issues with Castor (most addressed quickly)
- Distributed data transfers okay
 - ♦ some issues with latency of transfers (improve tools, keep testing)
- Reprocessing and skimming ran together with data transfers at T1s
- Skimming okay; not yet at full scale
- Analysis test gave very promising results
- Still need to work on optimization and automation
- More work needed on monitoring
- Training of users/operators a goal for the summer
 - ♦ Data operations team is well advanced;
 - ♦ planning for 24x7 facilities operation



Offline Software Activities

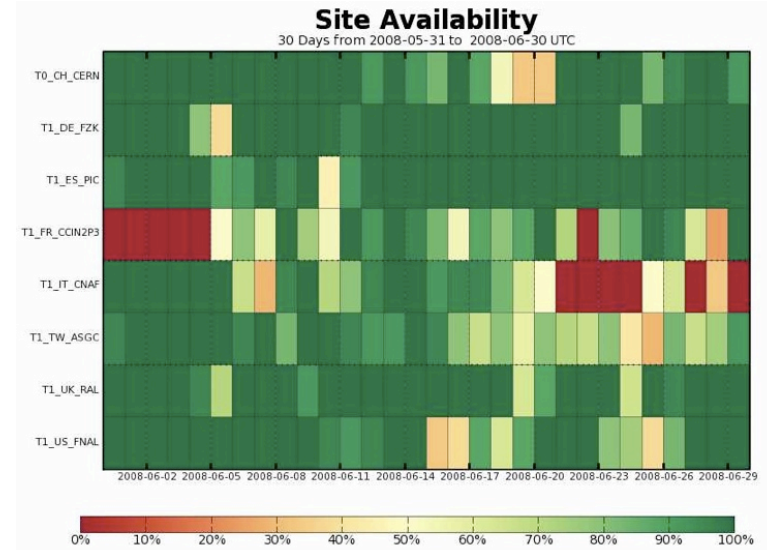
- Complete functionality required for data-taking nearly ready (CMSSW 2_1_0 expected to be released on July 5)
 - ◆ CMSSW : final geometry, new magnetic field map, new Geant4 physics models, DQM and certification, common analysis tools etc
 - ◆ Production tools: data repacking, prompt reconstruction & prompt calibration workflows which run at Tier 0
- Campaigns on-going to optimize software quality and performance at all levels - needed for smooth operations
 - ◆ to improve memory footprint and cpu performance
 - ◆ to review event content and reduce its size



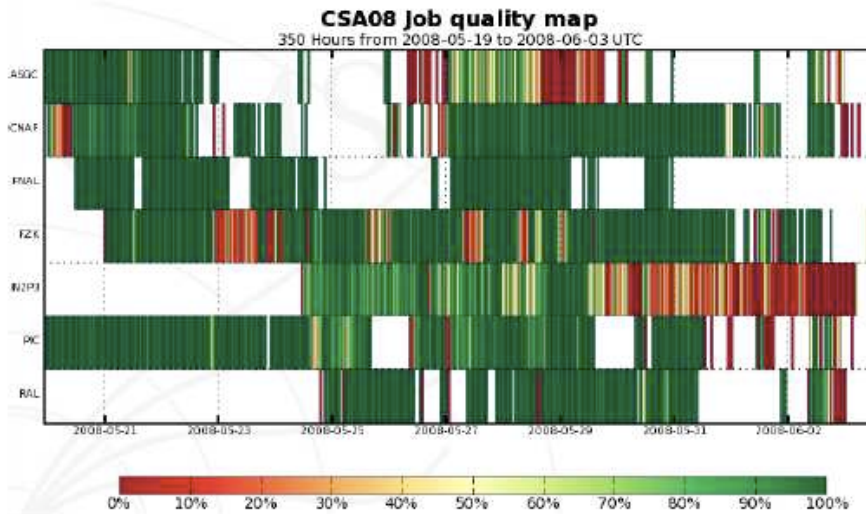
Site availability - T0 and Tier 1

CMS SAM tests run on all sites every 2 hours. The measured availability of the T0/T1 sites is reasonably well understood (upgrades, etc.) but not good enough yet.

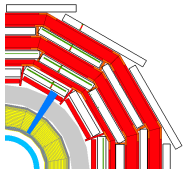
Situation has improved in past 12 months. Still it is not easy to be green!



T1 processing during CCRC/CSA08



<http://lxarda16.cern.ch/dashboard/request.py/samvisualization>



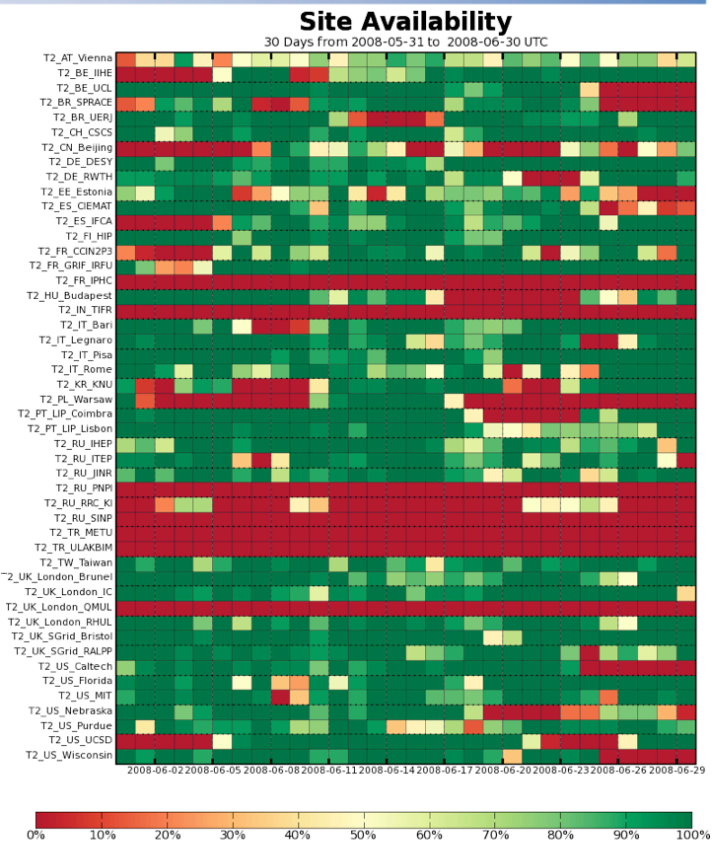
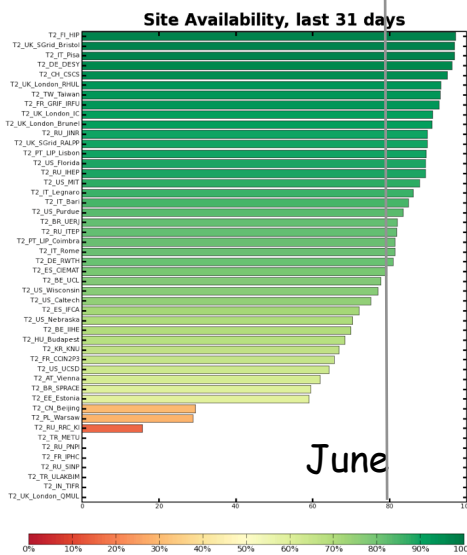
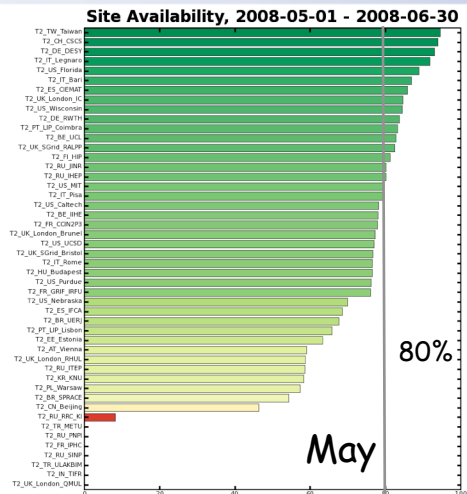
T2 site commissioning

. A CMS campaign on monitoring and site commissioning has been initiated to debug problems on the T2 sites.

. Some CCRC tests were aimed at debugging the sites. Site managers participated in monitoring and debugging their systems. The challenges have been very useful exercises in this regard.

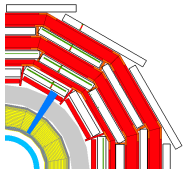
. Some sites will need some extra TLC to bring into the system.

. The program of site commissioning will propose a scheme to declare a site commissioned for CMS MC prod and/or analysis.



What do we mean by a CMS commissioned T2?

- will evaluate availability, reliability, site size
- T2s must be available for MC production, also needed for support of physics/detector groups, local support of users.



Computing Model Evolution

- The CMS computing model has been very robust, but designed primarily for stable running
- We have formed a new task force that includes representatives from physics, trigger, commissioning, offline and computing to discuss the evolution of the computing model during 2008 and 2009.
 - ♦ Includes evaluation of resources at the centers
 - uses pledges for Sept '08 => hence concentration on site commissioning/testing/monitoring
 - ♦ For early running we imagine a slightly different model: RECO available at the CAF and Tier-2s.
 - For 2008 disk resources seem okay to host at least 2 copies of the primary datasets at the T2 centers

CMS computing will be ready for collisions!