CMS Experiences In CSA06

A Computing, Software & Analysis Challenge in 2006

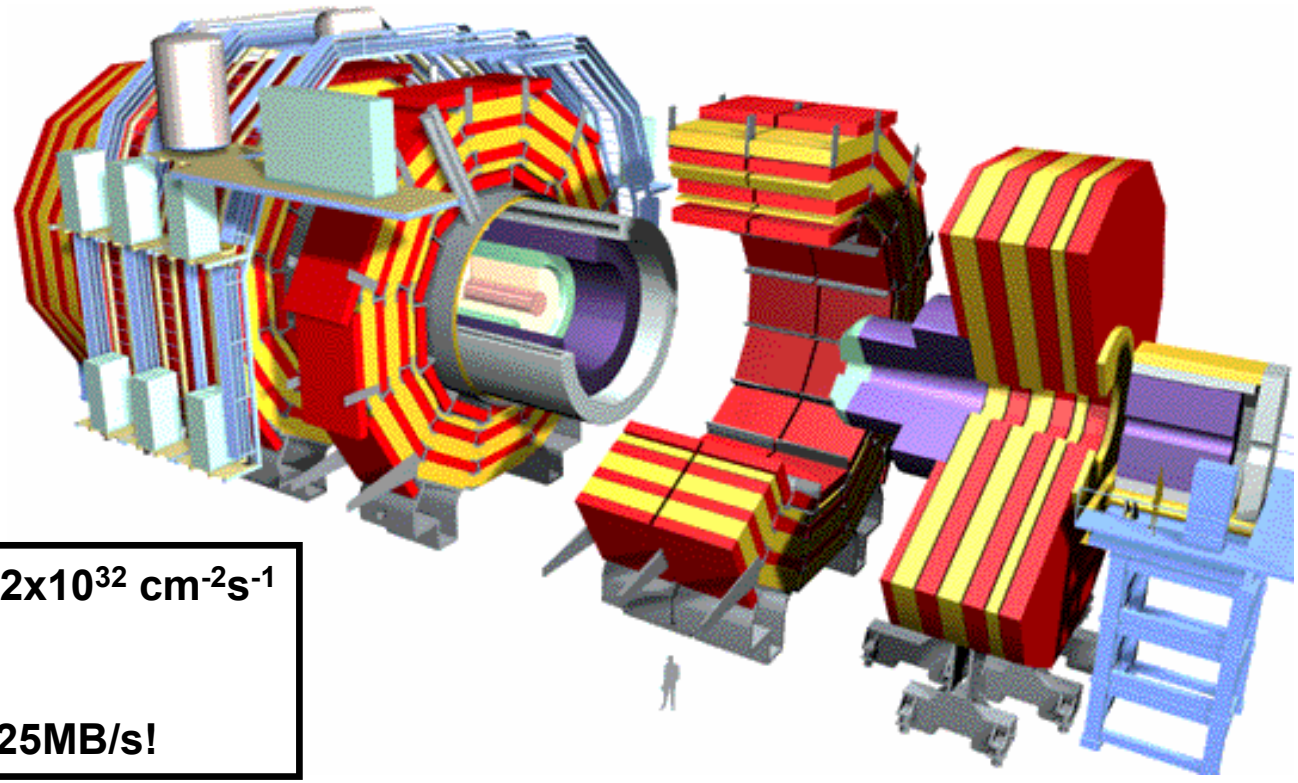
Jorge L. Rodriguez
University of Florida
Reporting for the CSA06 Group

Joint DPF JPS Meeting • November 1st, 2006 • Honolulu, Hawai'i

The CMS Experiment

CMS "Compact Muon Solenoid" 1 of 4 experiments at the LHC

- Millions of channels + High Luminosity + processing ~ 10 PB/yr
- Lots of experimenters more than 2500 located all over the world



At startup luminosity of 2×10^{32} cm^{-2}s^{-1}
Trigger rate ~150 Hz
Events size ~ 1.5MB
Data written to tape @ 225MB/s!

CMS' Distributed Computing Model

WLCG

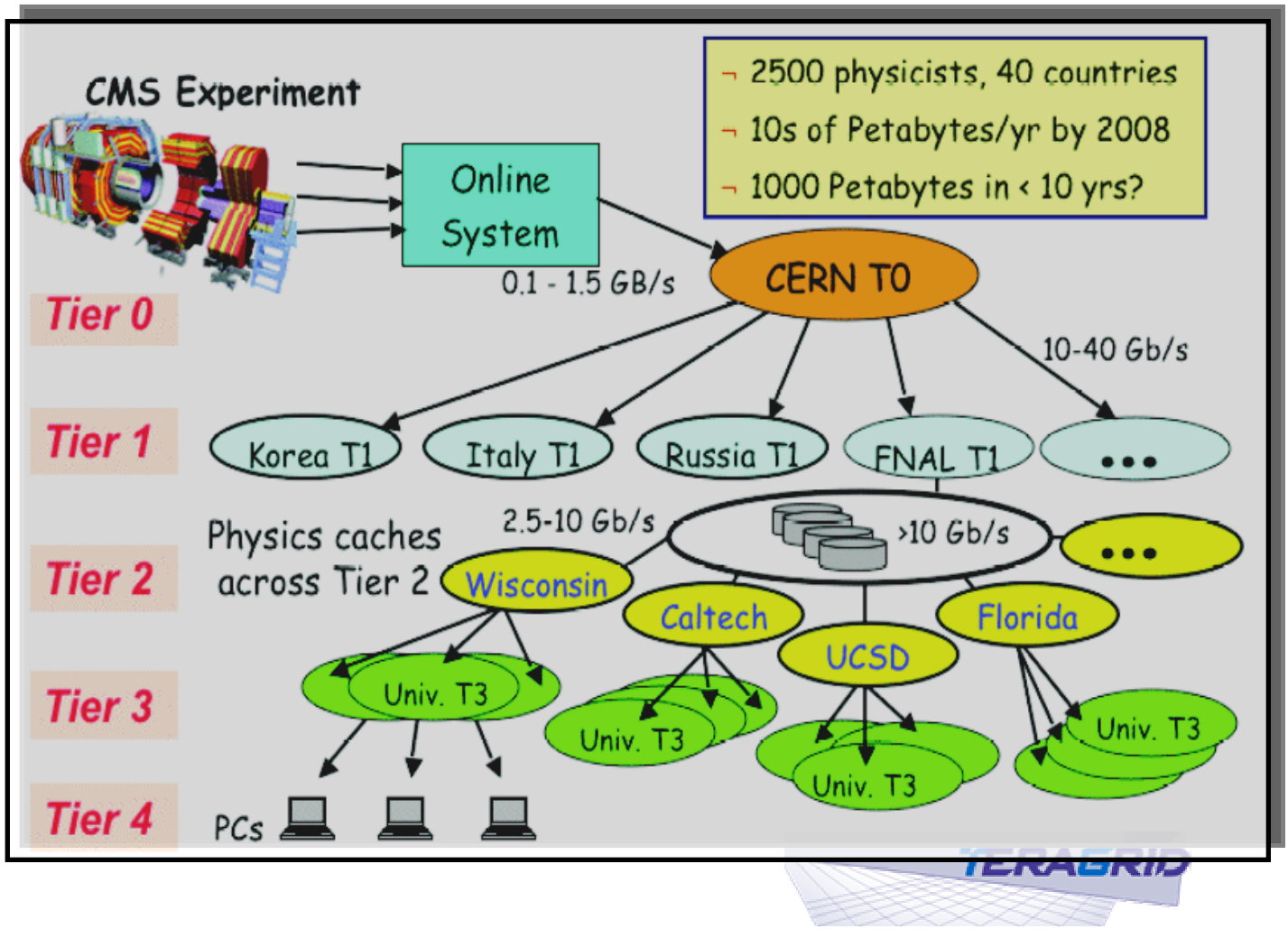
Open Science Grid

PPDG iVDgL GriPhyN Data Intensive Science

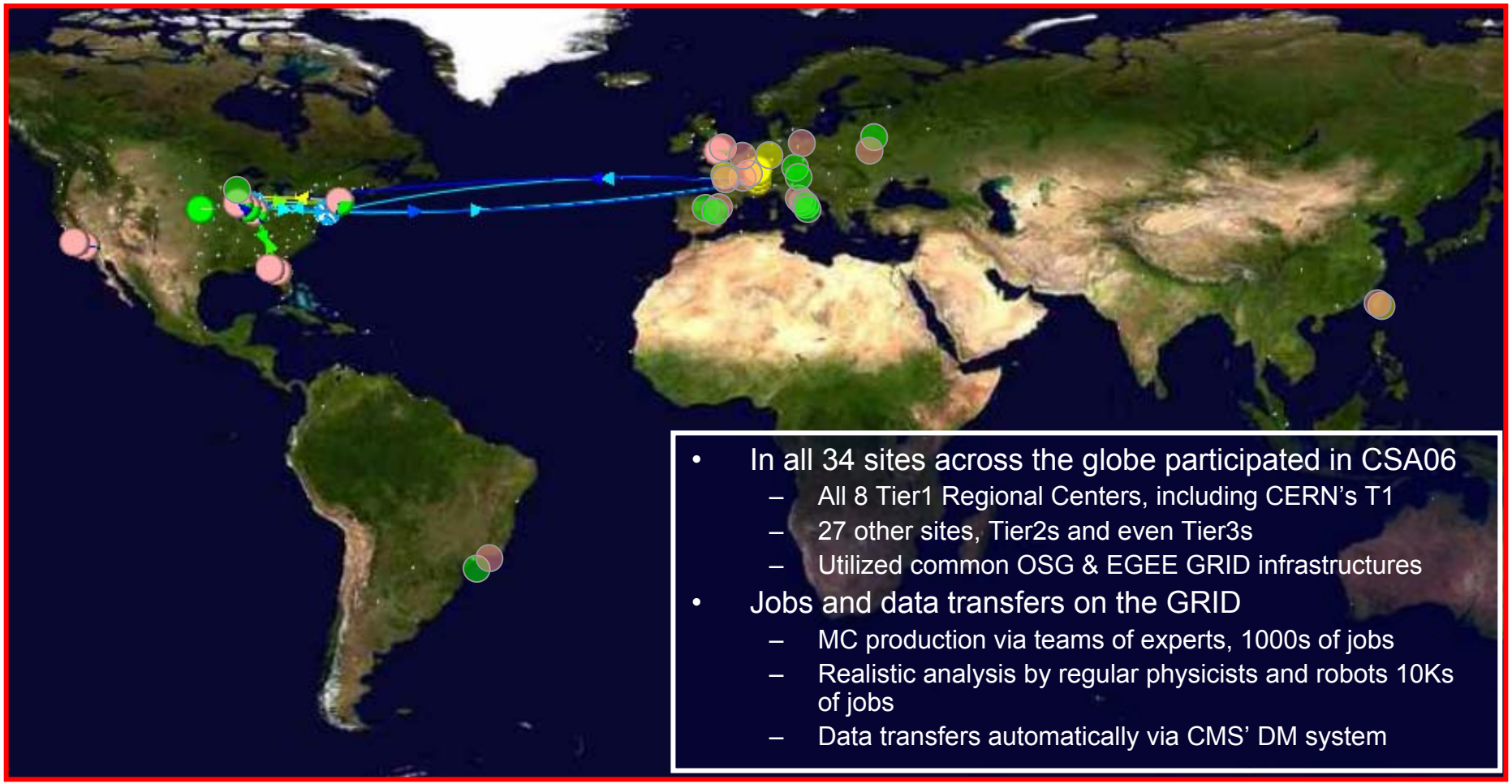
US CMS US ATLAS

eGee Enabling Grids for E-science in Europe

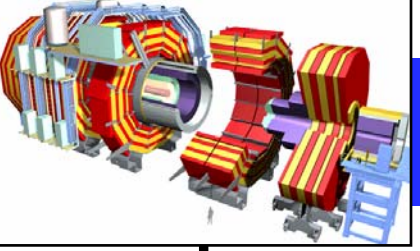
LCG



CMS Computing on the GRID

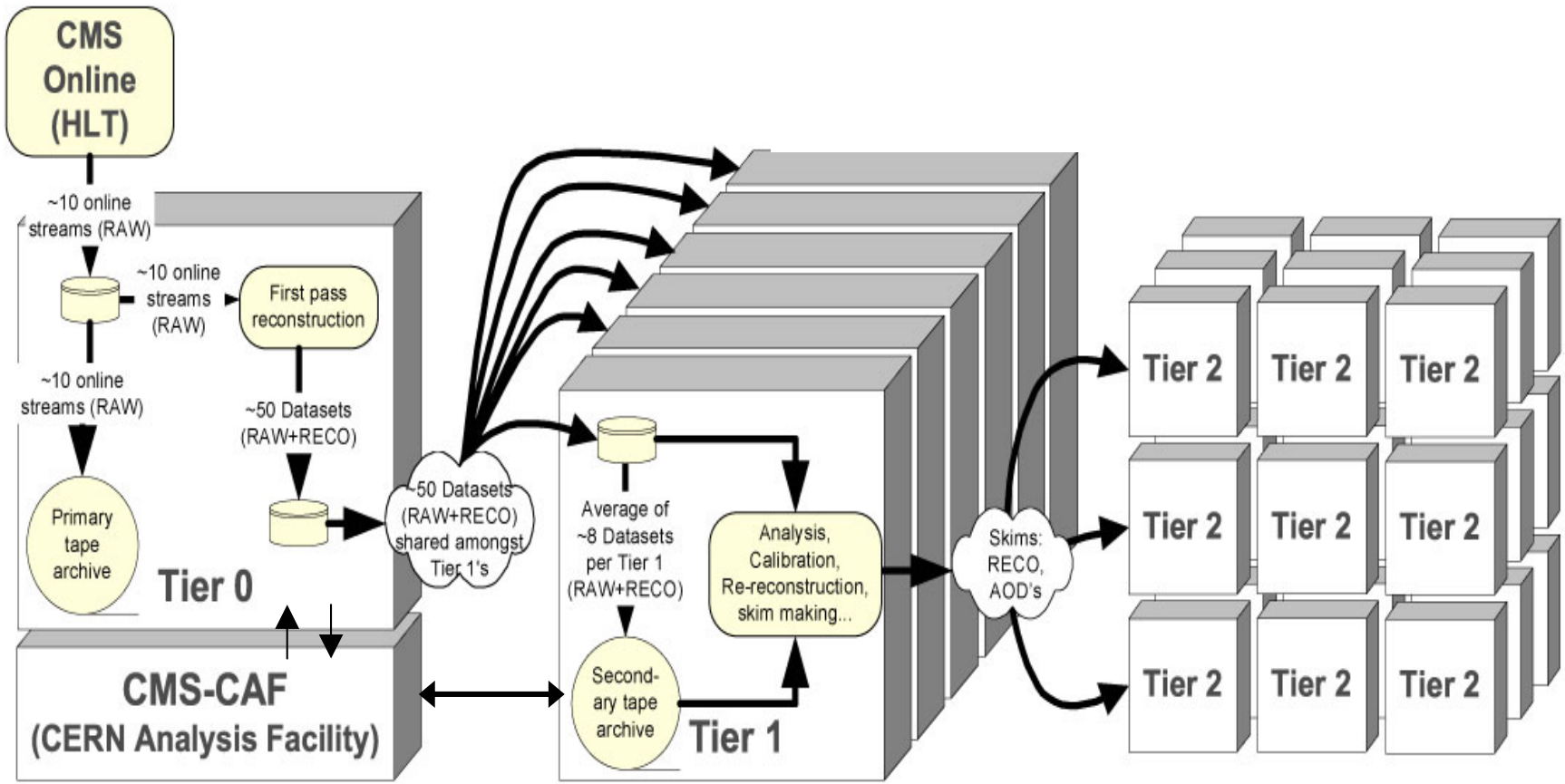


- In all 34 sites across the globe participated in CSA06
  - All 8 Tier1 Regional Centers, including CERN's T1
  - 27 other sites, Tier2s and even Tier3s
  - Utilized common OSG & EGEE GRID infrastructures
- Jobs and data transfers on the GRID
  - MC production via teams of experts, 1000s of jobs
  - Realistic analysis by regular physicists and robots 10Ks of jobs
  - Data transfers automatically via CMS' DM system



CMS Data Flow

Online System: RAW+DAQ



CMS Event Data Model

Evt Format	Data Contents	Evt Size (MB)	Evts/yr	Volume/yr (PB)
DAQ+RAW	Detector data + L1 trig. results	1-1.5	1.5×10^9	-
RAW	Detector data after online formatting + L1 trig results + HLT trig. Bits	1.5	3.3×10^9 2 copies+overlap	5.0
RECO	Reconstructed objects, tracks, jets… + all hits & clusters	0.25	8.3×10^9 2 copies + 3 reprocessing	2.1
AOD	Reconstructed objects, tracks, jets… + hits & clusters + small quantities of localized hit info	0.05	53×10^9 4 versions + 8 copies at Tier1	2.6
TAG	Run/evt number, high level physics objects used to index events	0.01	-	-
FEVT	Term used to denote RAW+RECO	-	-	-

The CSA06 Challenge: What is it?

- **A 50 million event exercise to test the workflow and dataflow associated with the data handling and data access model of CMS**
- **A 25% capacity test of what we will need in 2008**

Overall Goals of CSA06

- Demonstrate designed workflow and dataflow
- Demonstrate Computing-Software synchronization
  - Go smoothly through one or more CMS Software updates
- Demonstrate production-grade reconstruction software
  - Includes calibration; detector performance
- Demonstrate all cross-project actions
  - Determination and use of calibration/alignment constants including insertion and extraction and offline use of said constants via distributed constants database system
  - The HLT exercise: Split pre-challenge samples into multiple "tagged" streams and process these through complete DM system
- Provide services to a wide user community
  - Not just robotic GRID submissions
  - Support local and remote GRID users

Quantitative Goals and Metrics

- Site participation
  - Tier1s: Goal is all, but more than 5
  - Tier2s: Goal is 20 but more than 15
- Tier0 Processing farm
  - Number of weeks of sustained running: Goal is four weeks
  - Tier0 efficiency: Goal of 80% but more than 30%
- Data Management and Movement System
  - Data transfer rates from Tier0 to Tier1 to tape per site: Goal more than 50% of site specific capacity
  - Data transfer rates from Tier1 to Tier2: Goal is 20 but more than 5 MB/sec/site
- Physics Analysis Jobs on the GRID
  - Running jobs at Tier1 and Tier2 [2hr jobs/day]: Goal is 50K/day but more than 30K
  - GRID job efficiency: Goal 90% but more than 70%

Binary Goals and Exercises

- Are global systems operational?
  - Does the Data Mgmt system work? Focus on T0 –>T1 –> T2 transfers
  - Does Constants DB system work? Can we read DB offline & remotely can we insert new constants into system?
  - Can we run analysis & skim jobs via CRAB on the GRID?

- Exercises designed to test Workflows in CMS
  - Alignment Exercises
    - Get corrections from misaligned datasets (Tracker, & Muon Sys)
    - Inject corrections into DB and apply in Re-reconstruction
  - Calibration Exercises
    - ECAL, HCAL misalignment and re-reconstruction as above
    - ECAL, HCAL: ϕ symmetry calibration, HCAL jet corr. functions…
  - Analysis Exercise Demonstations
    - Extraction of signals: dimuons distributions, Higgs mass plots…
    - T-Tbar, SUSY/BSM, Standard model background and other studies

Defining theCSA06 Challenge

66M MC events simulated up to detector digitization with no-pileup

HLT tagged events were based on MC truth tables

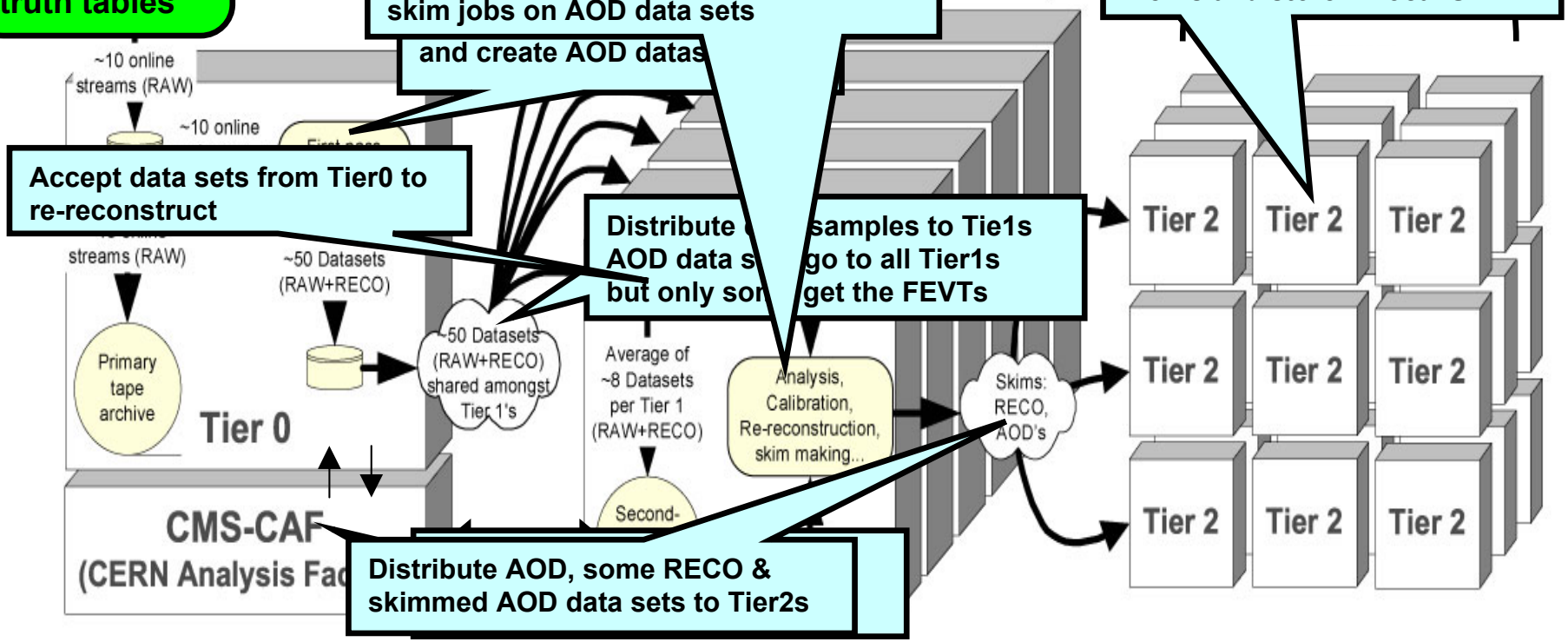
Tier1&Tier2s process GRID analysis jobs submitted either by job robots or by CMS physicists using CRAB (CMS' GRID job builder/submitter)

Run re-reconstruction with updated constants and CMS software. Also run skim jobs on AOD data sets and create AOD datas

Accept selected data sets from Tier1s and store in local SE

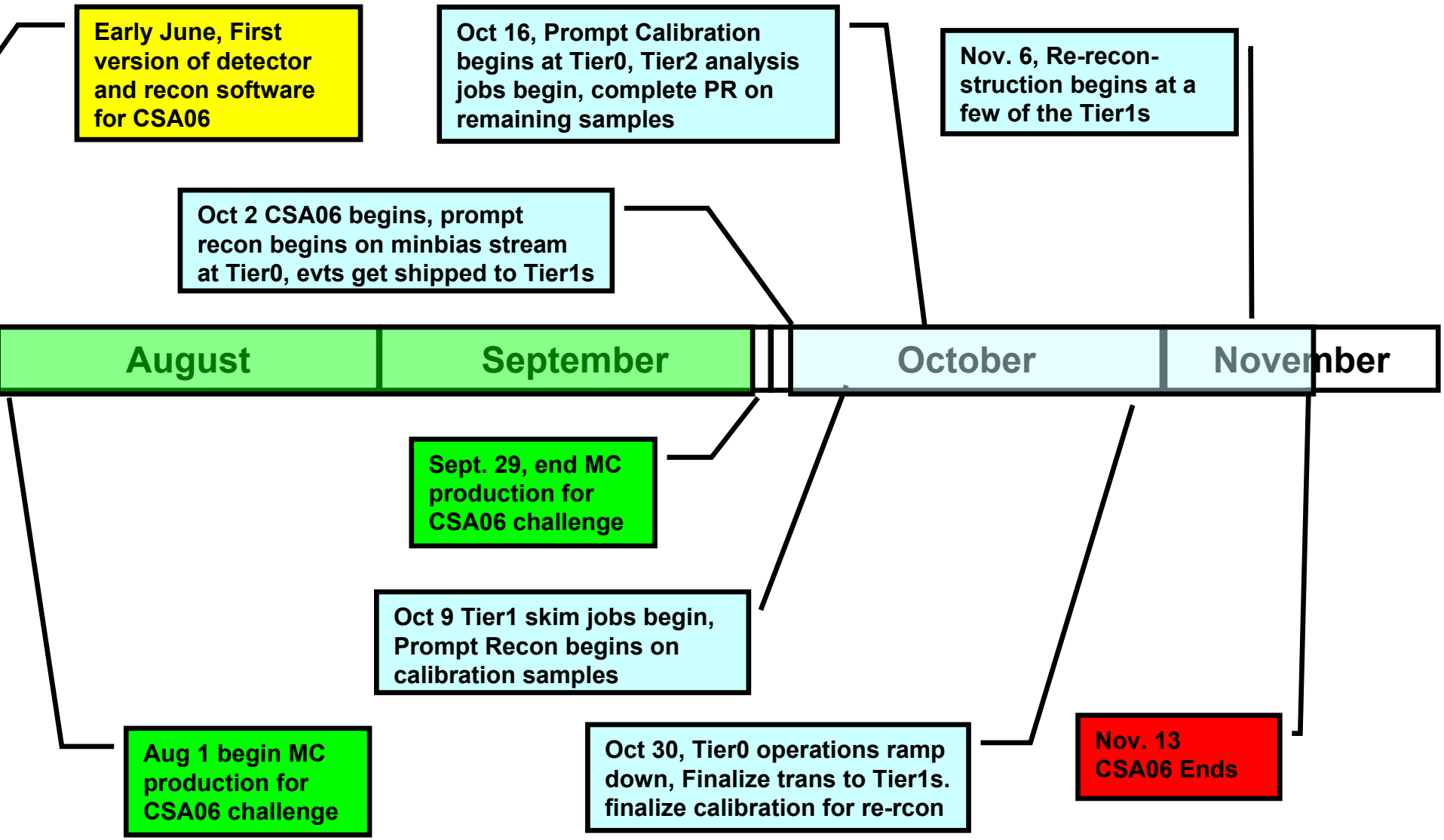
Accept data sets from Tier0 to re-reconstruct

Distribute samples to Tie1s AOD data s go to all Tier1s but only sor get the FEVTs



Distribute AOD, some RECO & skimmed AOD data sets to Tier2s

CSA06 Time Line



Status as of 10/31/06

CSA06 Status Report

"Many pieces of CSA06 already successful"

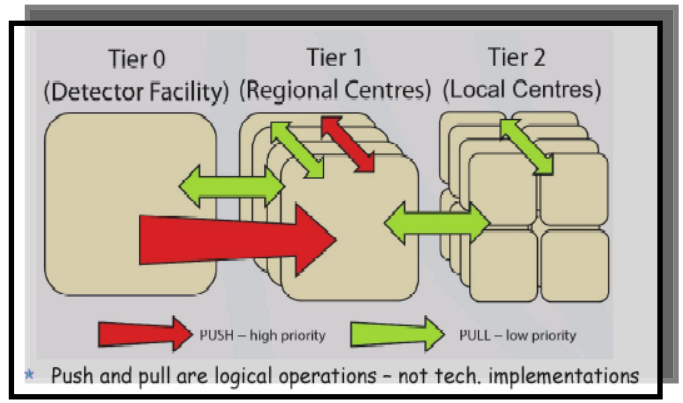
Michel Ernst @ RRB Meeting on 10/24/06

- Some Highlights
  - 66M events simulated with new CMSSW (Aug-Sept)
  - 207M events reconstructed at Tier0 with 100% uptime for 4 weeks
  - All Tier0 workflows completed successfully
    - <50 Hz> for first 3 weeks, increased to 100Hz the last 5 days
    - New CMS software stable and performs well
    - Calibration output produced & registered into DMS quickly & efficiently
  - ➢ More than 900TB of data transferred to Tier1 & Tier2s via the DMS
  - ➢ Skim & Analysis jobs underway & running very well
    - Average about 20,000 jobs/day across the GRID
    - Physics Analysis jobs success rates @Tier2s are excellent

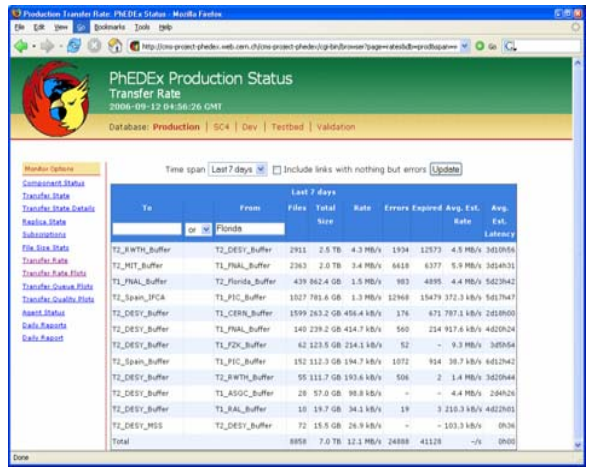
CMS Data Management System

A Distributed Data Management System

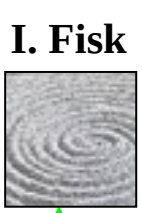
- Identification & location of data
- Worldwide movement of data replicas
  - Tier0 ⇔ Tier1 ⇔Tier2 ⇔Tier0…
    - Volume ~ PBs ≈ O(10 M) files
    - Transfer speeds ~ 5 Gb/s
- Data collections
  - Datasets: collections of data blocks
  - Blocks: collections of data files ~ 10TBs
  - Files: collections of events ~ GBs
- Components include
  - PhEDEx:
    - Hart of the system
    - Manages data at the block level
  - DLS: Data location Service
  - DBS: Data Bookkeeping System



PhEDEx gives user scontrols over movement of replicas via a user interface. It also provides monitoring of transfers and of volumes.



CMS Data Management System



I. Fisk

DLS

data block location (Stor.Elem)

DBS

data set data block data files

PhEDEx.cern

TMDB

data block files.root files.size chksums

FNAL

PhEDEx.local

srmdCache

0101001011101010

srmcp srm://<fnal> srm://<uftier2>

UFlorida Tier2

PhEDEx.local

download- agent

pool01

pnfs.local

dCache.local

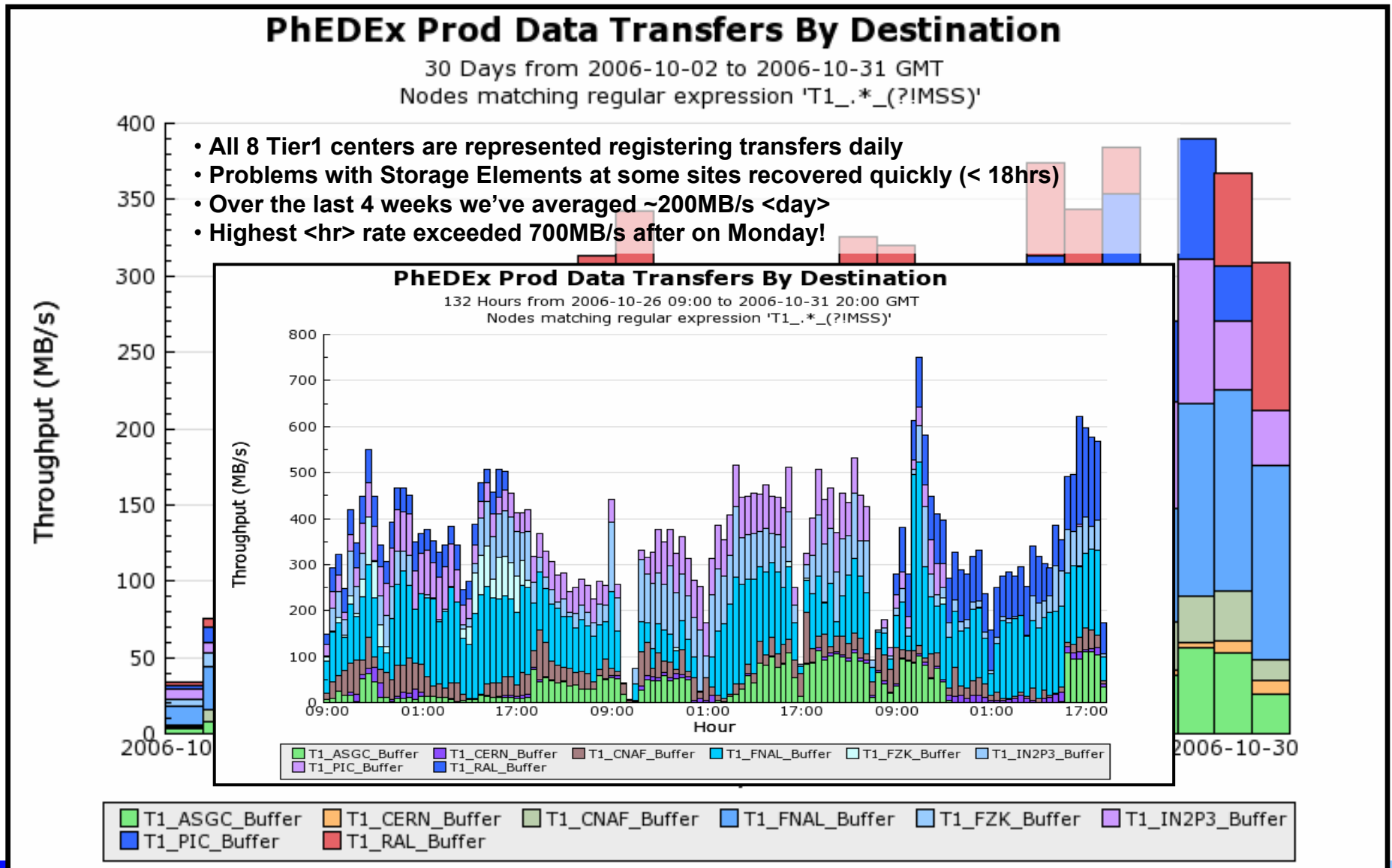
gridftp1.ocal

gridftp2.ocal

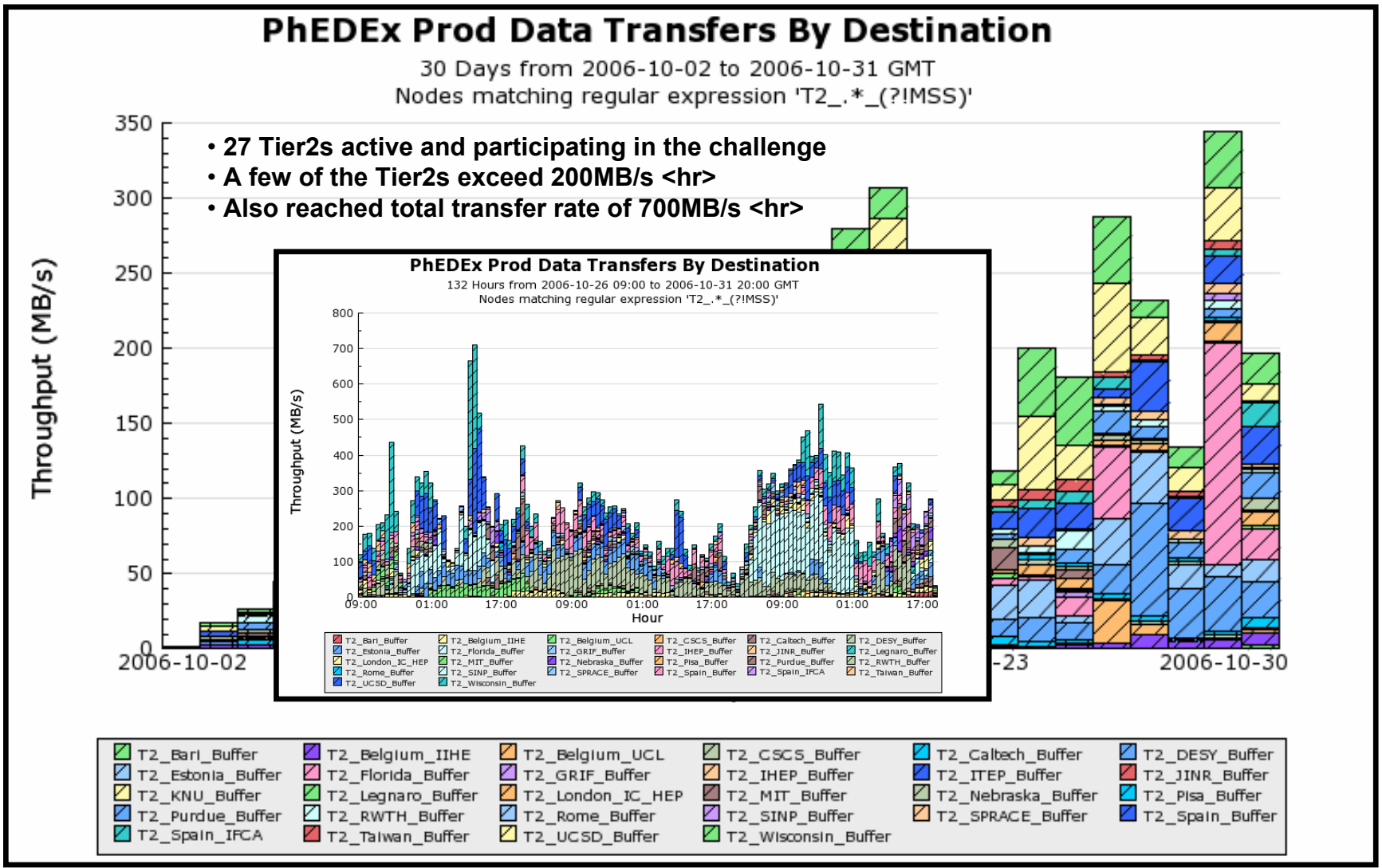
gridftp3.local

srm.local

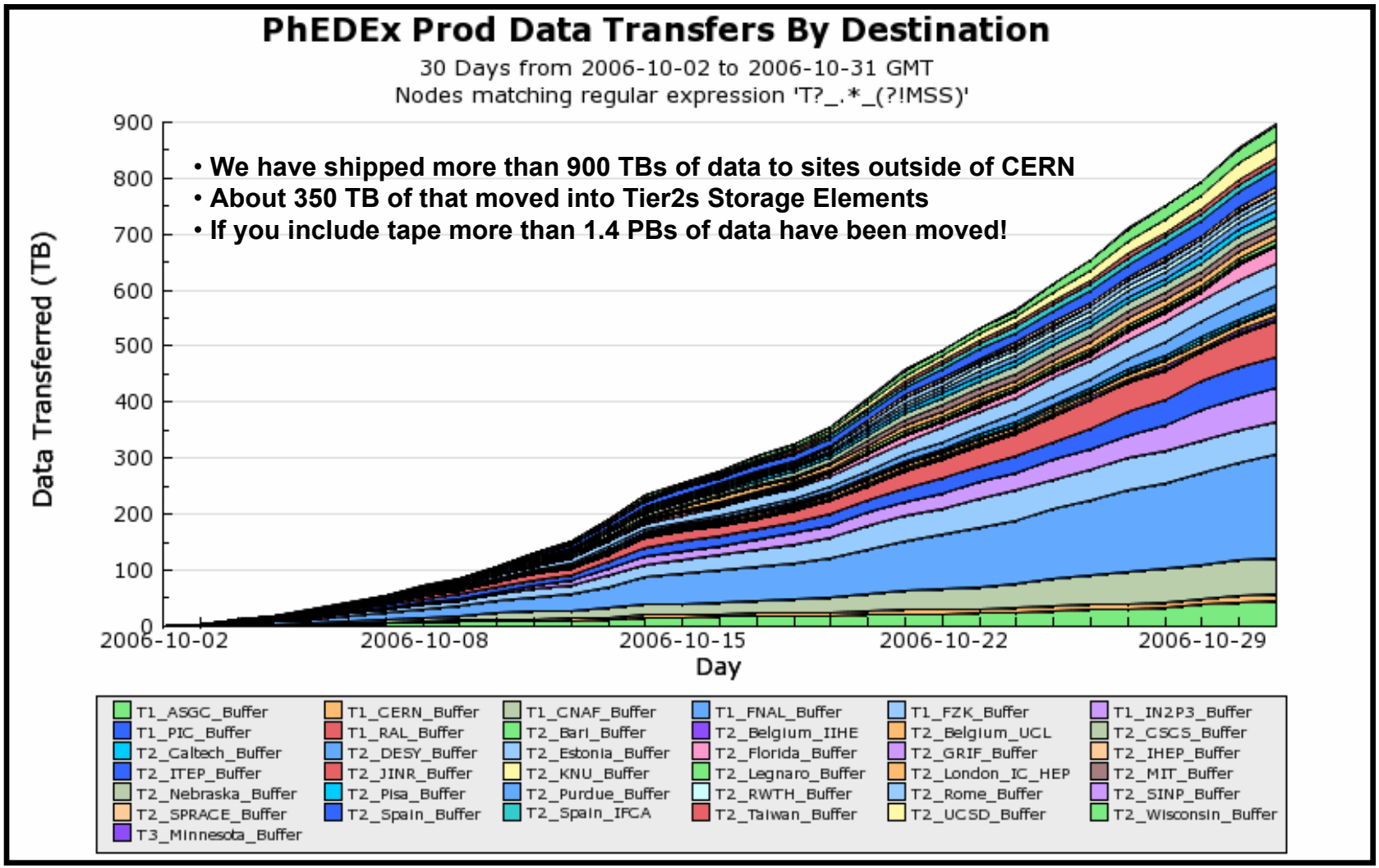
PhEDEx Tier1 Transfer Rate



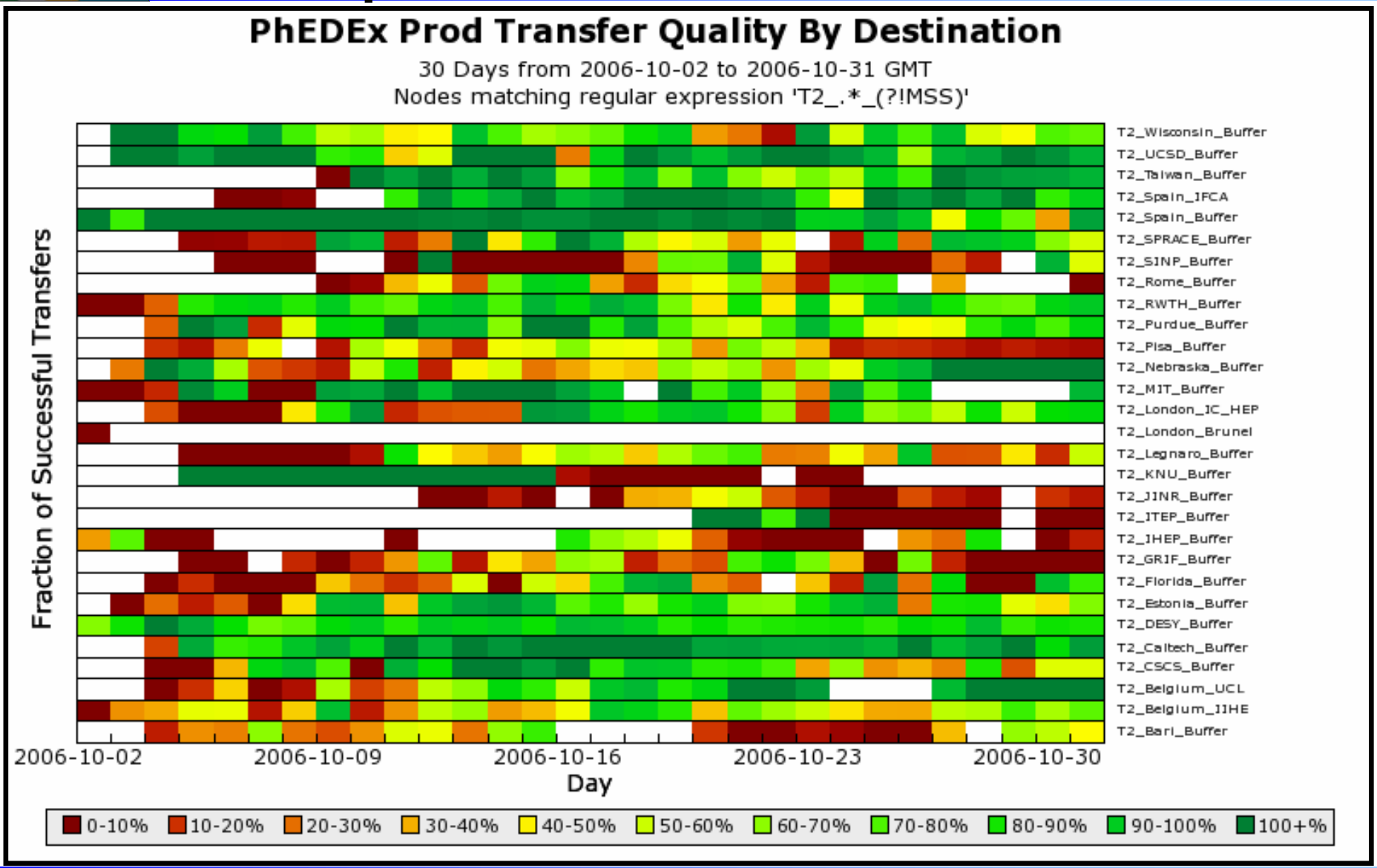
PhEDEx Tier2 Transfer Rate



Total Data Volumes Moved



Operational Performance

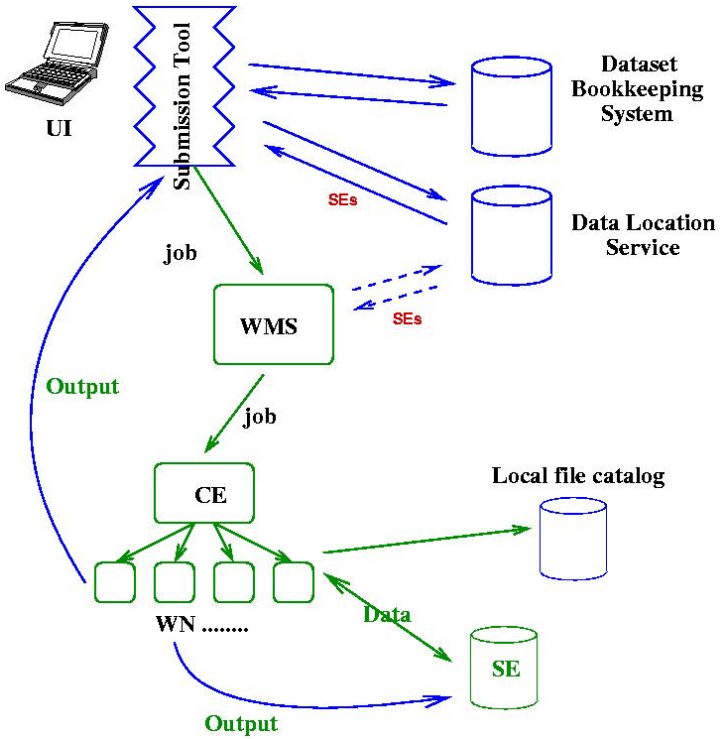


CRAB: CMS Remote Analysis Builder

An application to enable CMS analyses on the GRID

– User provides CRAB with:

- Dataset name, number of events
- Analysis code and job parameters: output file name & location etc…



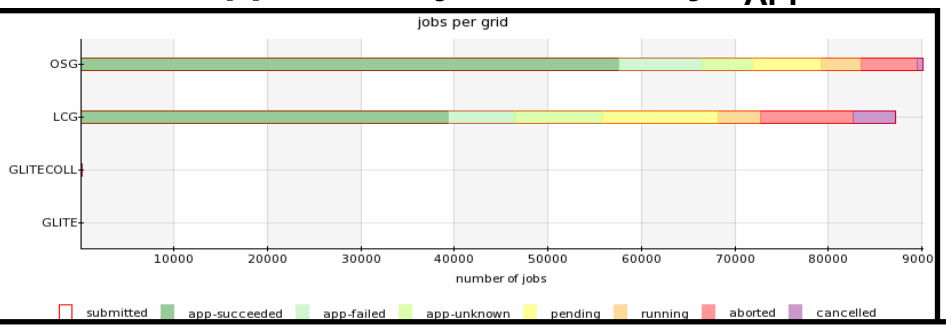
– CRAB provides users with:

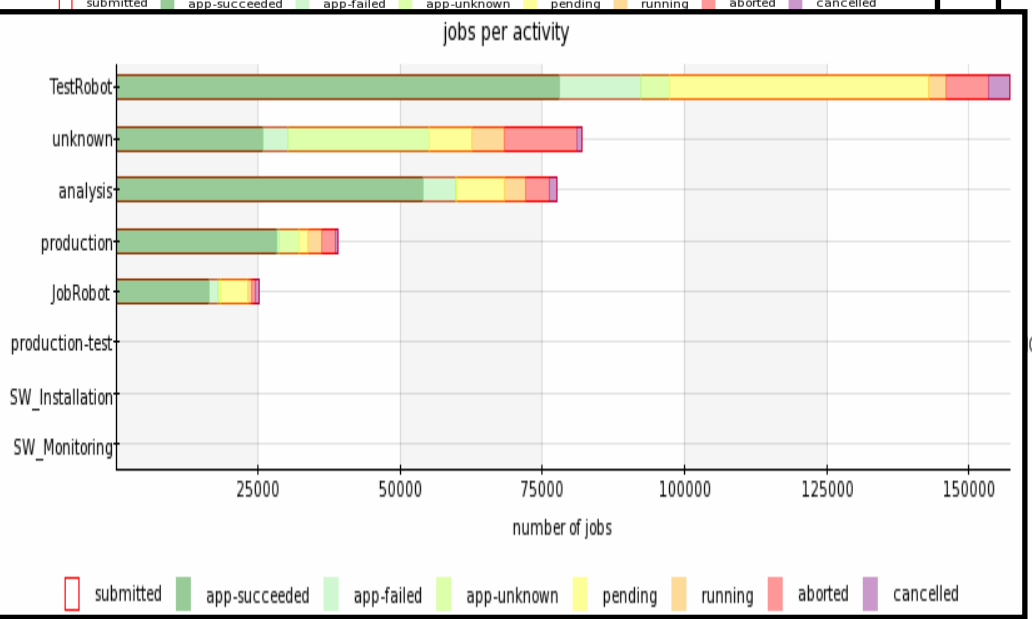
- Job creation and submission to GRID
- Data location and discovery
- Packaging of user code, bin, libs ,data…
- Job control, scripts, shell wrappers
- Job monitoring and output management

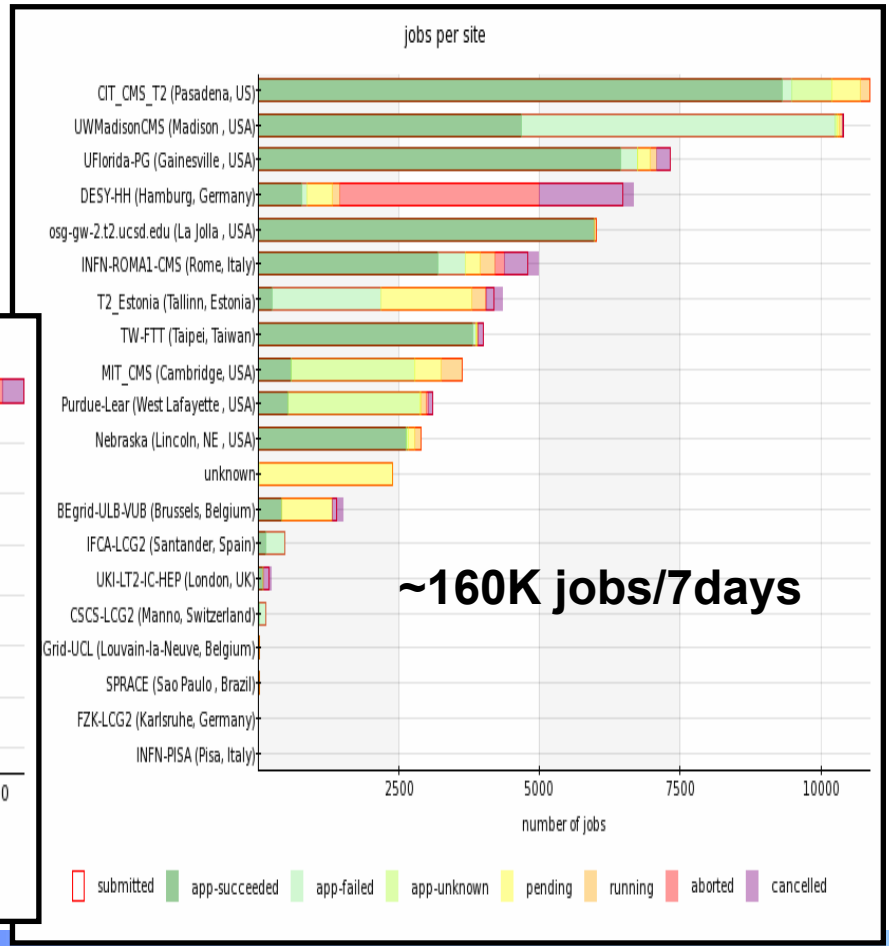
Preliminary Look @ CMS GRID jobs

10/24/2006 – 10/31/06

- **Overall GRID job efficiency ε_{GRID} = 85%** : **ε_{GRID} are jobs that get to site's worker node**
- **Overall application job efficiency ε_{APP} = 86%** : **ε_{APP} are jobs that return exit code 0**

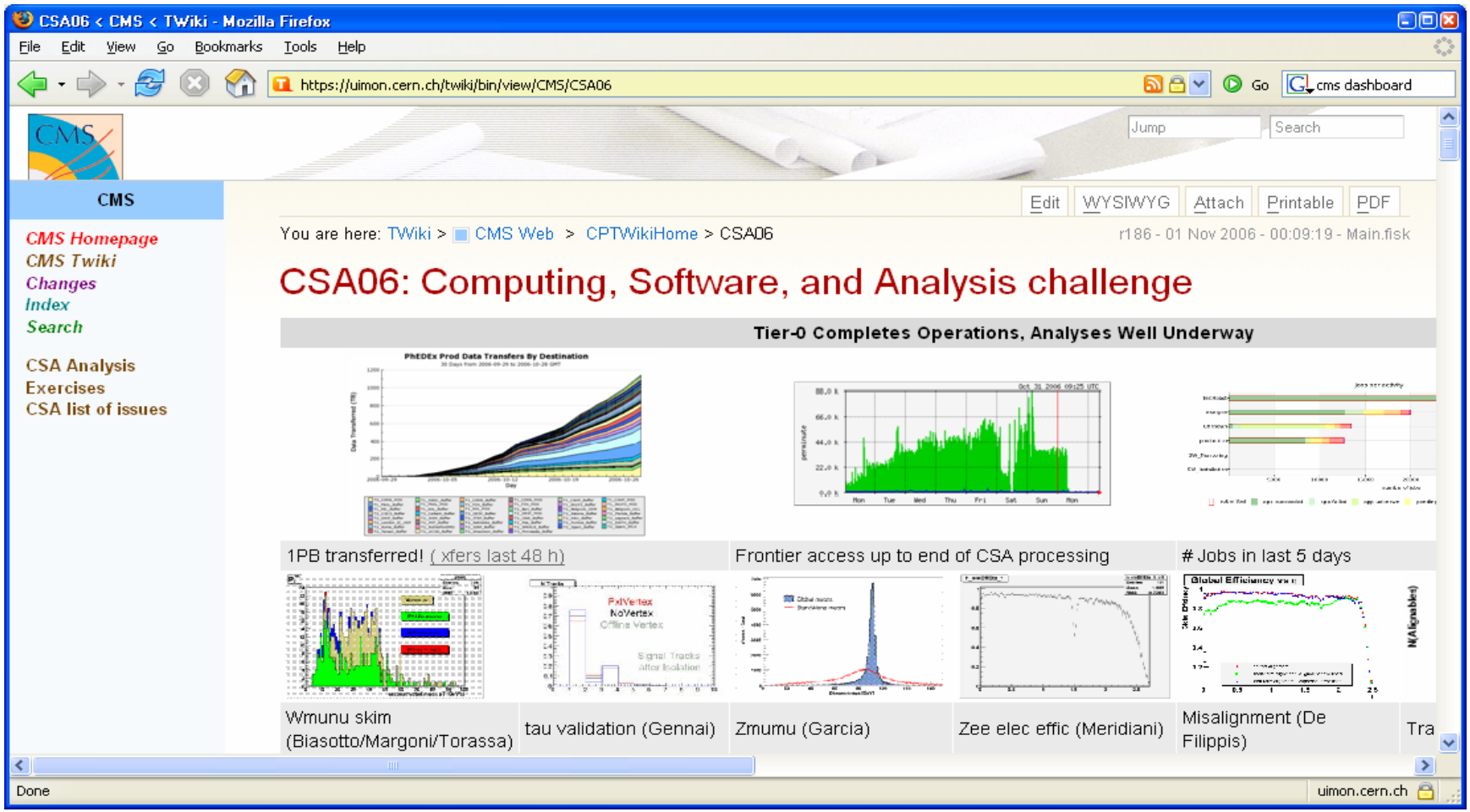






Status of Exercises

https://uimon.cern.ch/twiki/bin/view/CMS/CSA06



Summary

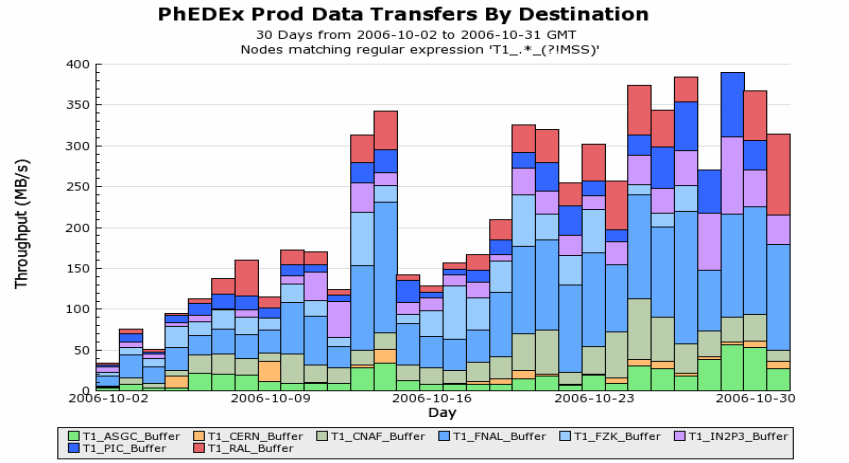
CSA06 already is quite a success story

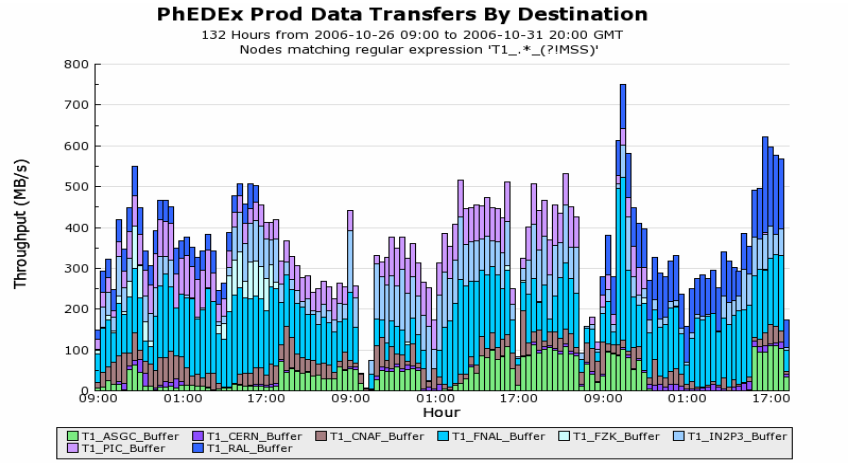
- We have met or exceeded most of our targets and goals
  - Greater than 50 Hz processing at Tier0 (100% uptime for 4 wks)
  - Impressive transfer rates to Tier1s & Tier2s
  - All workflows and dataflows successful
  - CMS Data Management system shown to perform very well in realistic environment
    - Lots of data moved around the world, automatically
    - Limited only by data availability at the Tier0
- CSA06 ends in two weeks
  - Tier0 processing now completed (200M+ events processed)
  - More than 1.4 PBs of data moved by CMS' DMS
  - Analysis jobs are now running at full bore

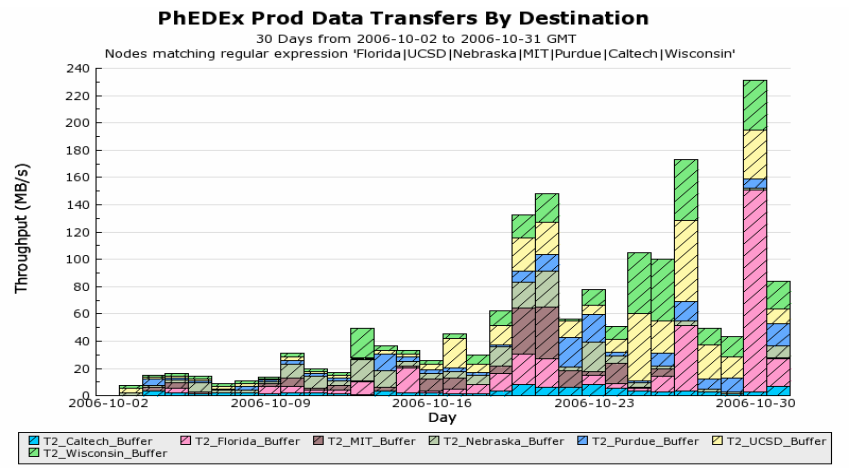
THE END

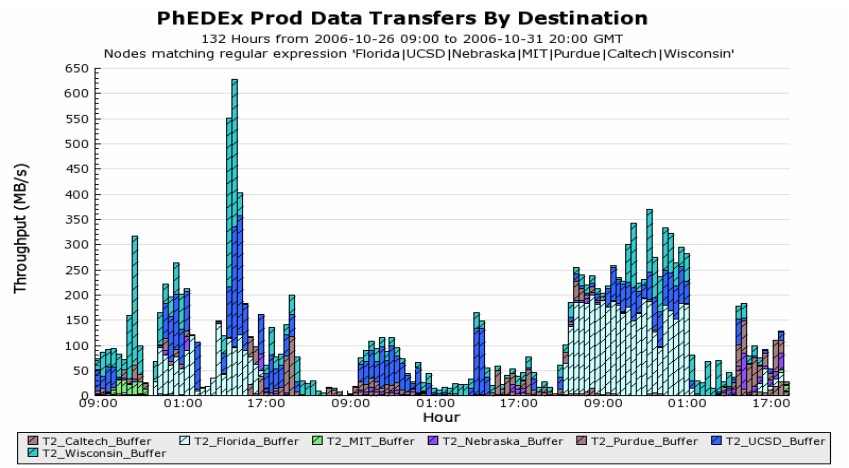
EXTRA SLIDES

PhEDEx transfer rate plots

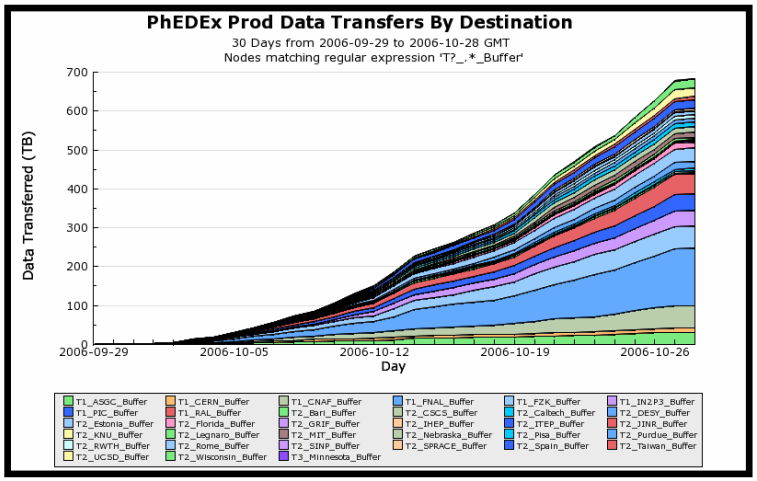


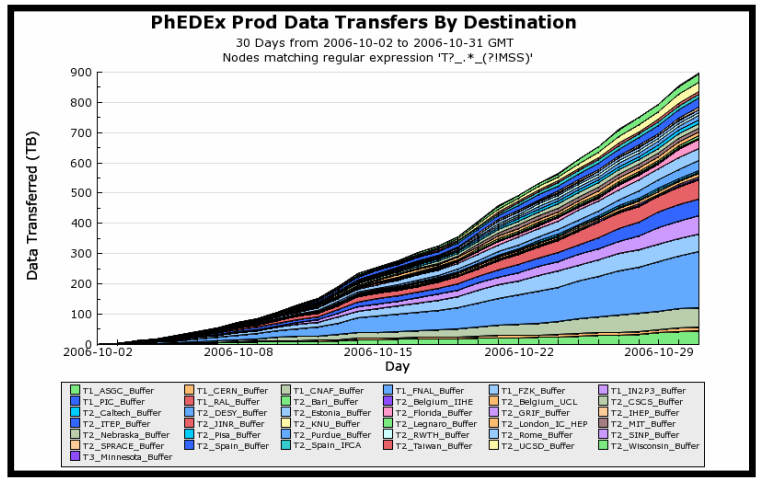


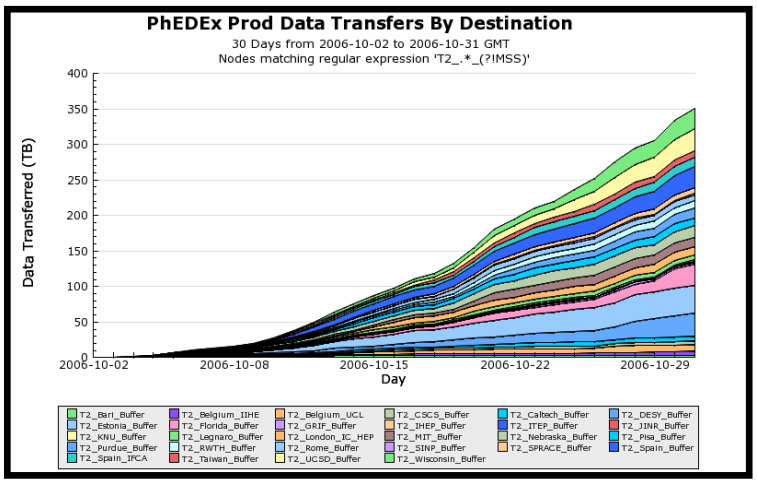


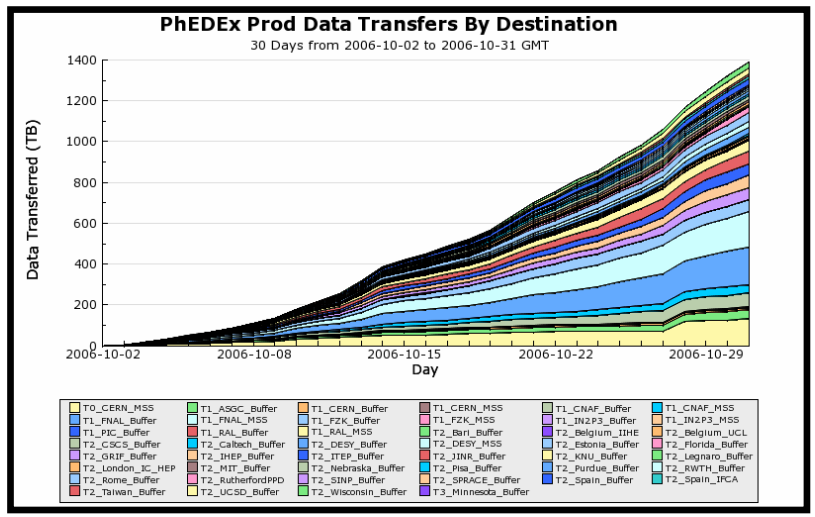


PhEDEx Data Volume









Goals and Metrics for CSA06

- Resources used for CSA06
  - About 1200 CPU at the Tier0
  - About 2000 to 2500 CPUs across all Tier1s
  - About 2000 to 2500 CPUs across all Tier2s
  - From 70 - 200TB disk + tape at participating Tier1s
  - From 5 - 25TB disk at participating Tier2s

MC Samples Created for CSA06

1. Minimum bias (40M)
2. T-Tbar (6M)
3. Z→μμ (2M)
4. W→eν (4M)

For calibration exercises

Total: ~66M events
Simulated up to detector digitization
No pile-up

5. Jet calibration soup (1M)
   dijet + Z+jet, various pt-hat ranges
6. Electroweak soup (5M)
   W –> l nu + Drell-Yan (m>15 GeV) + WW +H –> WW
7. Soft Muon Soup (2M)
   Inclusive muons in minbias + J/Psi production
8. Exotics Soup (1M)
   LM1 SUSY, Z' (700 GeV), and excited quark (2000 GeV)
9. HLT soup (5M)
   W (leptons) + Drell-Yan (leptons) + t-tbar (all modes) + dijets
   To be split into individual datasets for input to Tier0

Data Placement

Centre	Minbias	T-Tbar	Z->mumu	W->enu	Jet Soup	EWK Soup	Soft Muon	Exotics Soup	HLT Streams
ASGC	10% (8TB)					10TB			
CNAF	15% (12TB)		4TB	8TB	2TB	10TB	4TB		
FNAL	35%(28TB)	12TB		8TB	2TB	10TB		2TB	10TB
GridKa	15%(12TB)	12TB				10TB		2TB	
IN2P3	15% (12TB)				2TB	10TB		2TB	10TB
PIC	5% (4TB)		4TB			10TB	4TB		
RAL	5% (4TB)	12TB				10TB		2TB	10TB

HLT Filters

Name	Mnemonic	threshold	Bit position in 0_8_4	Bit Position 0_8_3	HLT Soup Efficiency
Single Gamma	p1g	80	0	6	0.9%
Double Gamma	p2g	30,20	1	2	3.0%
Single electron	p1e	26	2	4	32%
Double electron	p2e	12,12	3	0	3.3%
Single Muon	p1m	19	4	9	35%
Double Muon	p2m	7,7	5	7	3.2%
Single Tau	p1t	100	6	12	0
Double Tau	p2t	60,60	7	11	0
Single Jet	p1j	400	8	8	2.7%
DiJet[?]	p2j	350	9	10	2.2%
TriJet[?]	p3j	195	10	1	0.5%
Quad Jet	p4j	80	11	5	0.6%