CMS Experiences from Computing Software and Analysis Challenges

Ian Fisk
For CMS
Computing in High Energy Physics
September 4, 2007

What is a CSA?

Computing Software and Analysis Challenges

- Are opportunities to exercise aspects of the computing model and the software development program with analysis activities and users
  - Dedicated tests of components do not show interference problems
- CSA06 was intended to exercise the computing model at greater than 25% of the target for 2008
- The Challenge planned for 2007 should exceed 50% of final scale

Not all elements of the experiment activities are explored in each challenge

- Increasingly complete with a variety of elements added to the activity list in 2007

We attempt to align the challenge activities with other activities that involve the physics user communities

- The desire is to try to stress the system with real users
- We have a number of automated tools to generate load and stress the system, but the users make a much more realistic test

Basic Scaling Items Checked in CSA07

Service	2008 Goal	CSA07 Goal	CSA06 Goal	Status 2006
Tier-0 Reco Rate	150Hz - 300Hz	100Hz	50Hz	Achieved
Network Transfers between T0-T1	600MB/s	300MB/s	150MB/s	Achieved (6/7 cont.)
Network Transfers between T1-T2	50-500 MB/s	20-200 MB/s	10-100 MB/s	Achieved (15 sites)
Network Transfers T1-T1	100MB/s	50MB/s	NA	Not Attempted
Job Submission to Tier-1s	50k jobs/d	25k jobs/d	12k jobs/d	3k jobs/d
Job Submissions to Tier-2s	150k jobs/d	75k jobs/d	48k jobs/d	Achieved
MC Simulation	1.5 10^9 events/year	50M per month	NA	Not Attempted

Successful Demonstration of Services

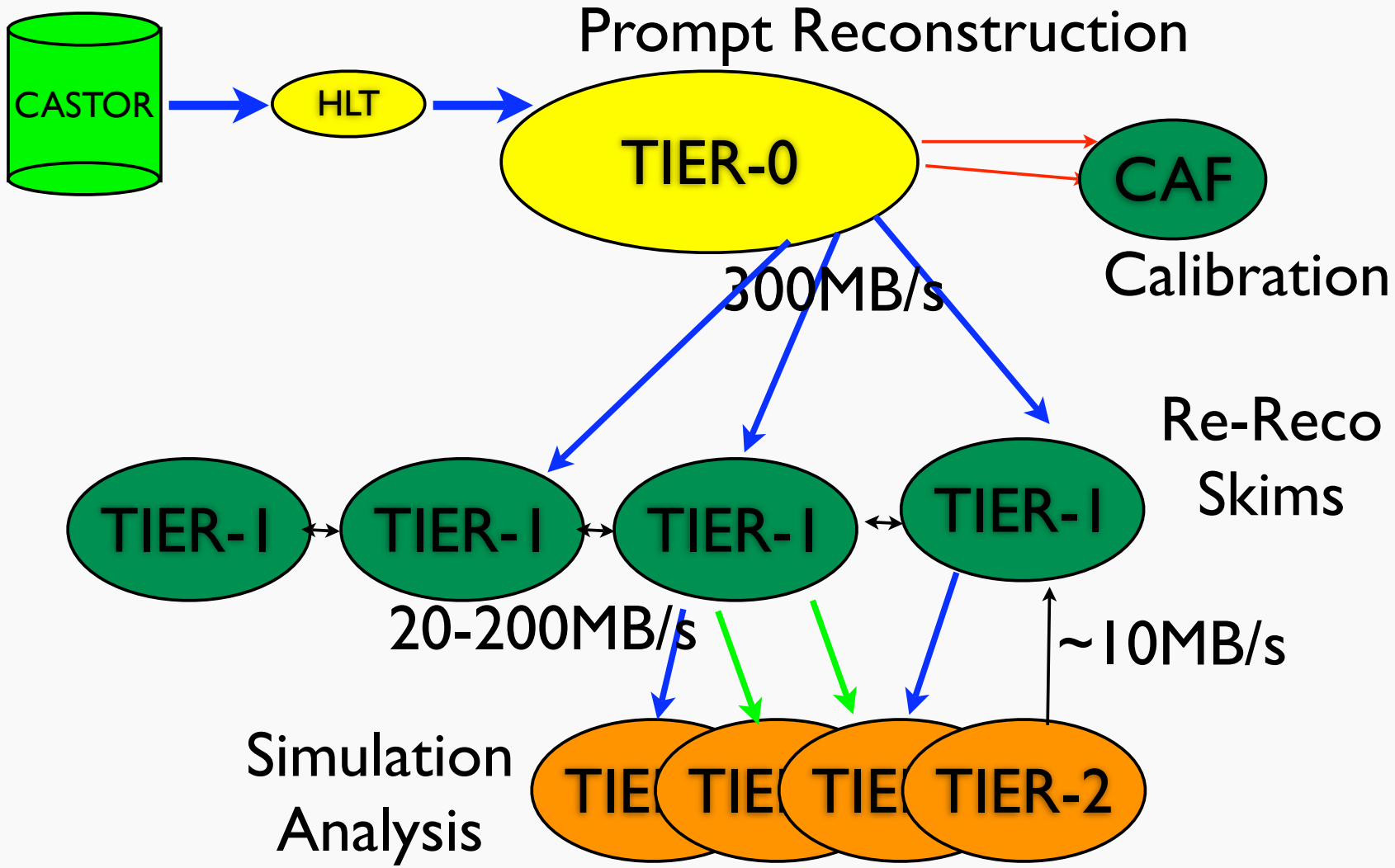
CMS is successfully using both the OSG and EGEE infrastructure to complete the computing software and analysis service challenge

- All Tier-1 centers contributing
- CMS had 19 EGEE Tier-2 sites and 8 OSG Tier-2 sites for CSA06
  - Roughly what is anticipated by the fraction of the collaboration
  - Will grow to XX EGEE Tier-2s and 9 OSG Tier-2s for 2007 challenge

Infrastructure for data selection, event simulation, user analysis submission, and load testing is designed to submit to both grids

- The 120M simulated events produced over the summer for the challenge were balanced over sites from both grids
- The load generators for simulated analysis access to Tier-2 sites perform direct Condor-G submission to OSG resources and use the gLite RB in bulk submission for EGEE resources
- The user analysis submission infrastructure can use the LCG-RB to submit to either grid transparently

CSA07 Workflows



Tier-0 Reconstruction

During the first four weeks of the challenge in 2006 over 200M were reconstructed at the Tier-0 center

- Averaged better than 50Hz for the first three weeks and increased to more than 100Hz for the final week
- The Tier-0 work flow beat the target rate each day with 100% up time
- New CMS Software framework has good performance and memory requirement properties
- Reconstruction, Calibration, and Analysis objects are produced and registered in the data management system
- Calibration information was accessed by the LCG-3D Frontier distribution system

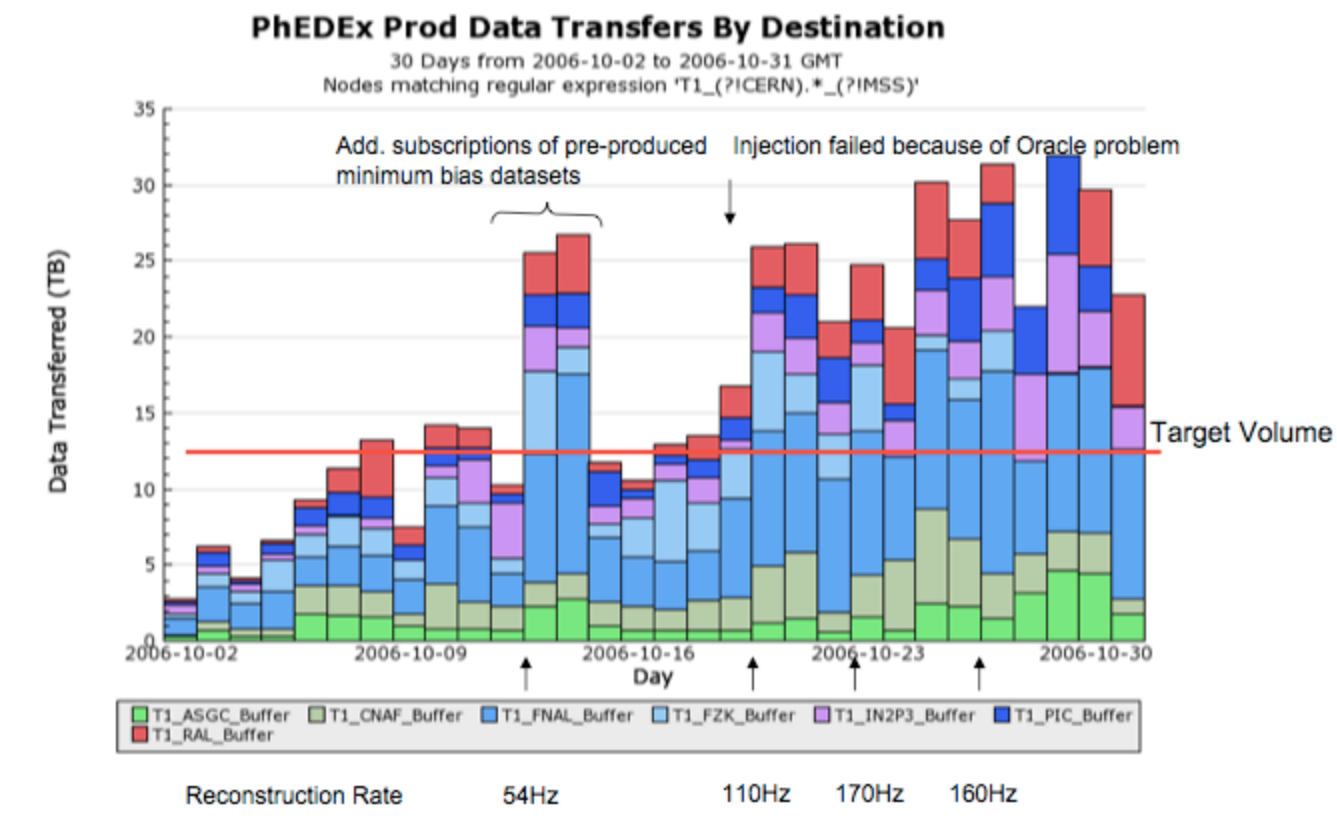
We have a more complete application for CSA07

- Expect to start at double the rate
- Included Higher Level Trigger Information
- Data divided into streams similar to actual detector data

Tier-0 to Tier-1 Transfer Performance

The Tier-0 to Tier-1 Transfer performance during event reconstruction

- More than doubled the initial challenge goals
- Will start at the higher rate from the beginning of the 2007 challenge
  - Hope to again beat the targets

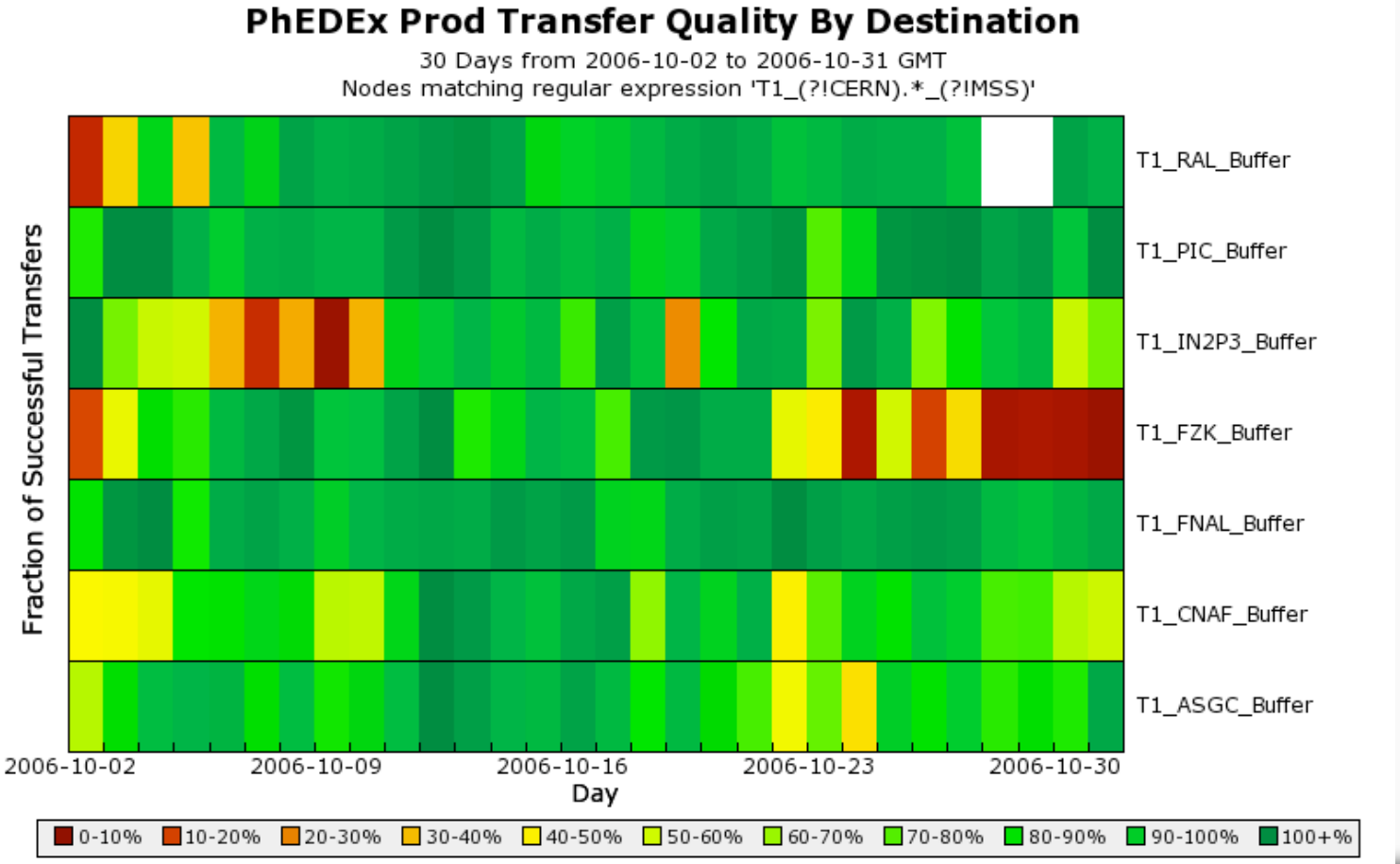


Site	Rate (MB/s)
ASGC	26
CNAF	37
FNAL	105
FZK	26
IN2P3	32
PIC	10
RAL	26

Tier-1 Transfer Quality

Stable performance from Tier-1s

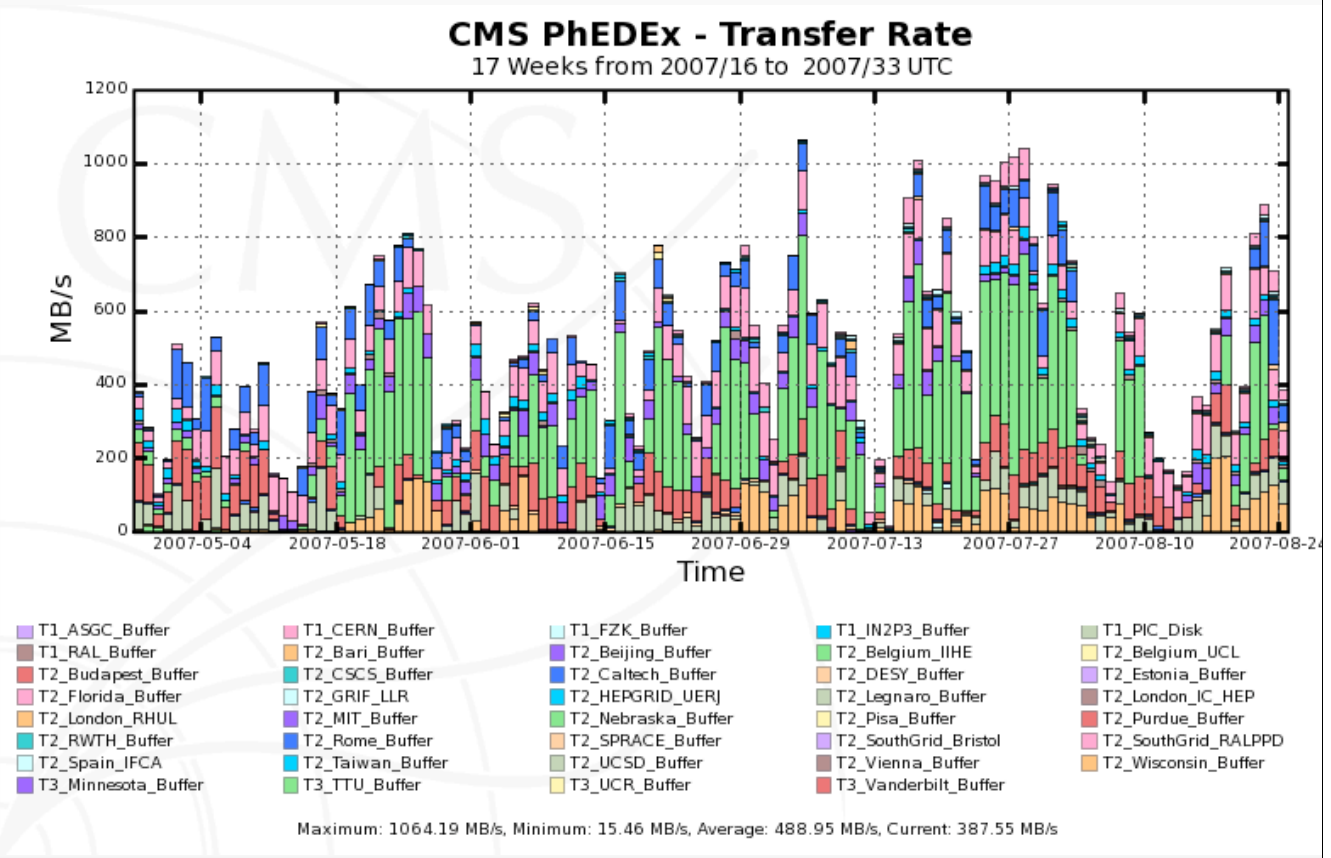
- A couple of sites have had specific problems or downtimes
  - This compares very favorably to the results from early 2006



Tier-1 to Tier-2 Transfers

In the CMS Computing Model the analysis objects (AOD) are stored at all Tier-1 computing centers

- The reconstructed and raw data may be at only one
  - A Tier-2 center can connect to any Tier-1
- Mesh of Tier-1 to Tier-2 centers is nearly 300 links
- Only a small fraction of these were tested in CSA06
- CMS has a Distributed Data Transfer



Transfer Functionality Untested in CSA06

Add Tier-1 to Tier-1 Transfers

- Tier-1 to Tier-1 transfers were tested opportunistically and only for short times
  - Roughly half the permutations were tested at all
- CMS Computing Model calls for the ability for predictable bursts of data transfers between Tier-1 centers

Incorporate all Tier-1 to Tier-2 transfers

- As daunting as it sounds all permutations need to succeed
  - Otherwise the Tier-2 will be limited for analysis

Include Tier-2 to Tier-1 Transfers

- Rate is modest, but must demonstrate the Tier-1 sites can simultaneously receive data from CERN and Tier-2s, while sending data to Tier-2s

Tier-1 to Tier-1 Transfers

Tier-1 to Tier-1 transfers in the CMS Computing model are intended to synchronize a re-reconstructed copy of the AODs

- The more data a Tier-1 site hosts the less needs to be imported and the more needs to be exported. Metric is 3 consecutive days
- Goal is to hit 50MB/s on average in and out of a nominal Tier-1 center

Import

Export	ASGC	CNAF	FNAL	FZK	IN2P3	PIC	RAL
ASGC	0	4	4	4	4	4	4
CNAF	6	0	6	6	6	6	6
FNAL	18	18	0	18	18	18	18
FZK	4	4	4	0	4	4	4
IN2P3	6	6	6	6	0	6	6
PIC	2	2	2	2	2	0	2
RAL	4	4	4	4	4	4	0

Site Resources

For the Tier-0 Reco Rate

- Need ~2500 CPUs at CERN for the challenge
  - Increase the completeness of the reconstruction algorithm
    - CSA06 Reconstruction rate was ~3s for minbias. Expected to be many times that when the experiment starts

Tier-1 capacity should be ~600 CPUs at a nominal Tier-1

- Nominally 150TB of mass storage
  - This is much smaller than the computing model calls for even for a 50% challenge, but the disk storage is scaled down because it's only a 1 month challenge and not a full year
- Tape ingest rate at all Tier-1s needs to be verified
  - ~half the total capacity was actually demonstrated in CSA06

Tier-2 capacity would be ~150CPUs at a nominal Tier-2

- Disk storage 20-30TB
- This is for a nominal Tier-2 and there is a wide variation

Processing in CSA

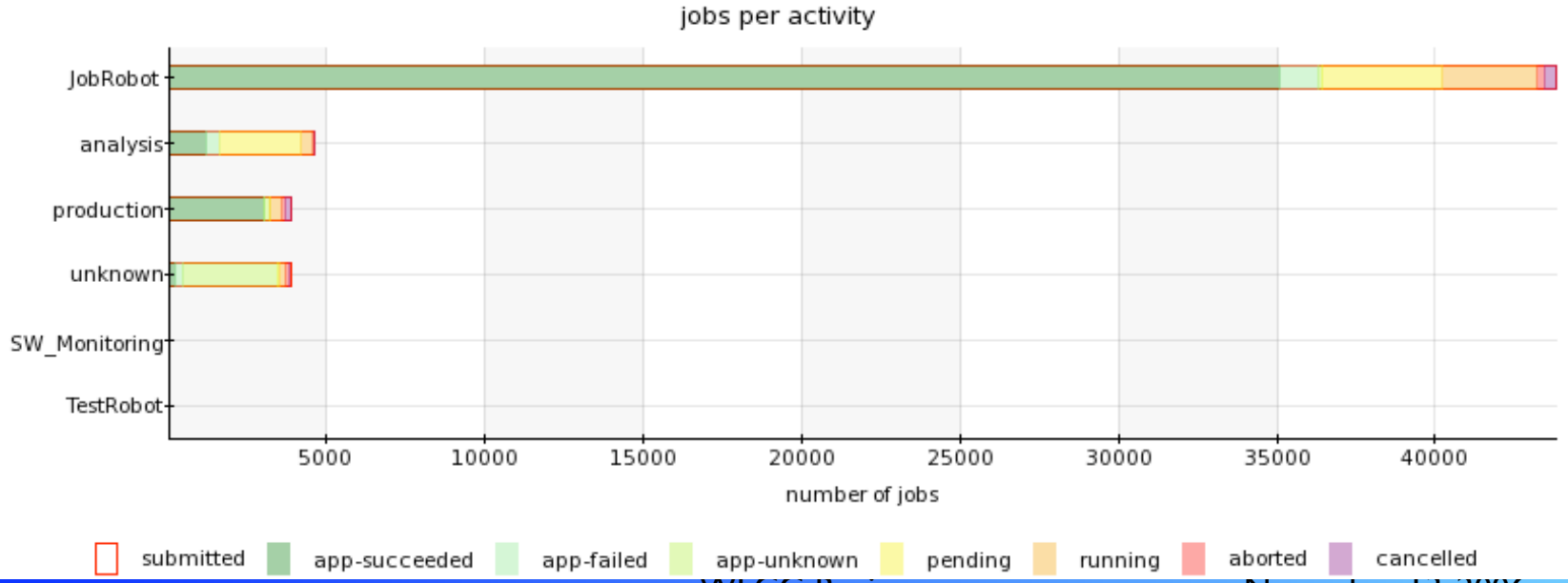
Aside from the Tier-0 prompt reconstruction, analysis and calibration object generation, there are four processing activities planned for CSA07

- Skimming
  - Selection of data at Tier-1 centers based on physics filters and transferred to Tier-2 centers for further analysis
- Analysis
  - Submission os analysis code to Tier-2 centers demonstrating local data access and the performance and the Tier-2 Storage Elements
  - Includes the contribution of the load generating job robot
- Re-Reconstruction
  - Processing of data at Tier-1 centers with access to new calibration constants through Frontier
- Simulation
  - Use the Tier-2 centers at 50% of the target rate of simulated event production

Job Submission

CMS was able to meet the metric of more than 50k job submissions per day during the last week of CSA06. Goal is 100k jobs in CSA07

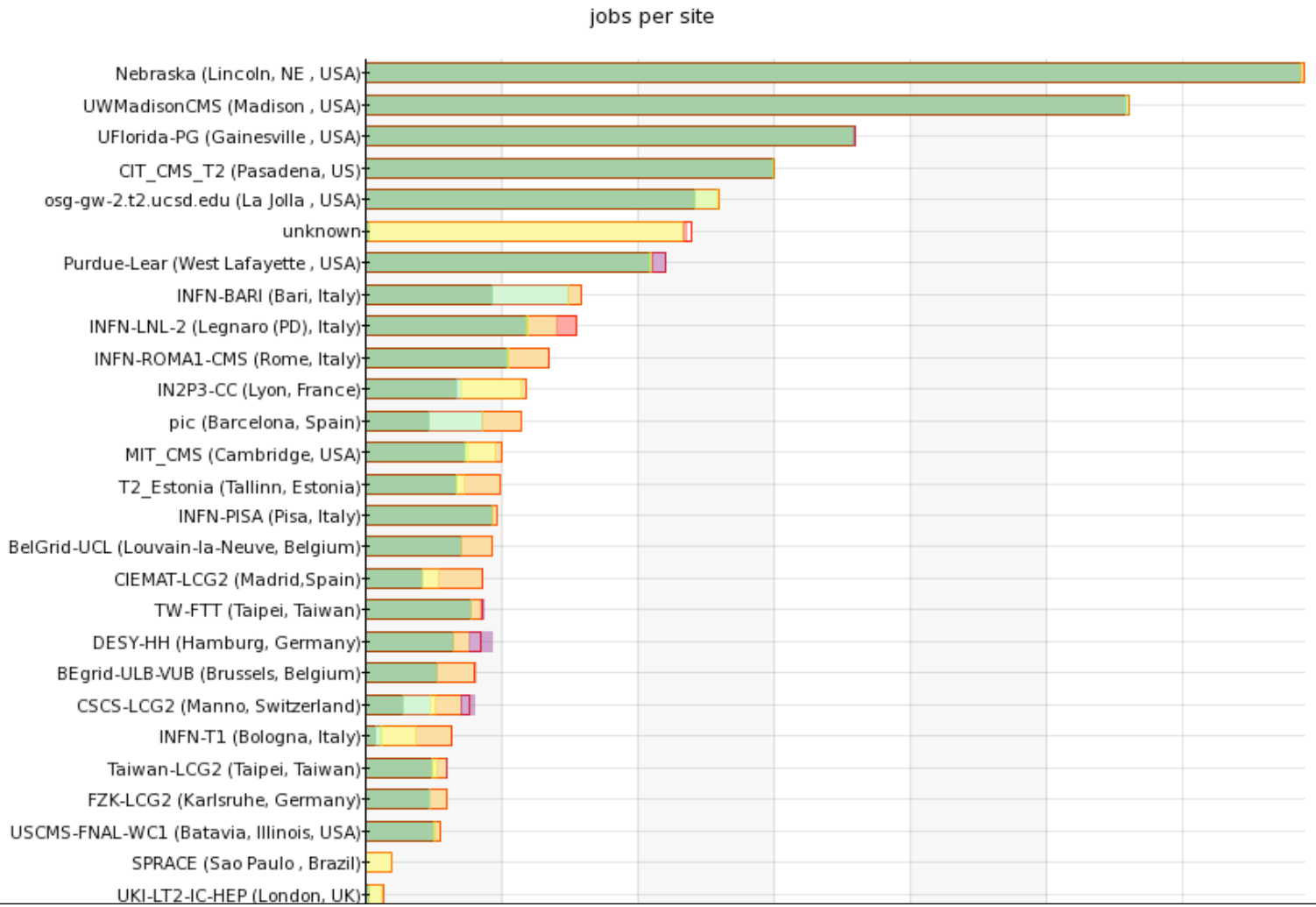
- User analysis was submitted by ~20 individuals
  - Hoping to increase this substantially in CSA07
- Skimming and Re-reconstruction has been executed at all Tier-1 centers
  - Successful validation of Frontier infrastructure
- The job robot introduces a load on sites that is very similar to user specified analysis applications



CMS Dashboard

The CMS dashboard is a successful collaboration of ARDA and CMS

- The goal was a 90% successful completion and currently is CMS is exceeding this



Schedule

We expect to begin producing a large sample of events with HLT information divided into Primary datasets in early September

About 20 days from the start we should complete sample preparation

- Begin reconstruction and distribution by the middle of September

Simulation at the Tier-2s will continue from the beginning

About a week after the beginning we expect to start the skimming at the Tier-1 sites

- Data movement and analysis access

By two weeks we expect to begin reprocessing at Tier-1 sites

Summary

CSA06 was a successful challenge

- Many elements of the CMS Computing Model were validated at the 25% scale and above
- Also gave some clear directions for areas we needed additional effort in debugging and commissioning

CSA07 will start shortly after CHEP

- New functionality will be added
- A 50% scale will be attempted
  - We will almost certainly find problems and bottlenecks at this new scale and functionality

It will be a busy fall