Region Report from US-CMS

Ian Fisk

WLCG Collaboration Meeting

January 24, 2007

The US-CMS Tier-1 Center at FNAL

- FNAL is a dedicated Tier-1 Facility for CMS
  - Meeting the obligations of the U.S. to CMS Computing
    - Supporting the local community
  - The only Tier-1 center for CMS in the Americas
- By head count US-CMS is about 30% of the CMS collaboration
  - FNAL is about two nominal Tier-1 centers by the computing model numbers
    - The single largest Tier-1 center for CMS

FNAL Tier-1 2008	CPU	4.3MSI2k	1000 dual CPU nodes
	Disk	2PB	200 Servers (1600MB/s IO)
	Network	15Gb/s	CERN to FNAL
	People	30FTE	Includes Developers and Ops

- FNAL recently completed the second year of the procurement ramp in preparation for the start of the experiment

Associated Tier-2s

US-CMS selected 7 production Tier-2 sites in late 2004

- Program began in June of 2005
  - 3 Prototype facilities have existed since 2002

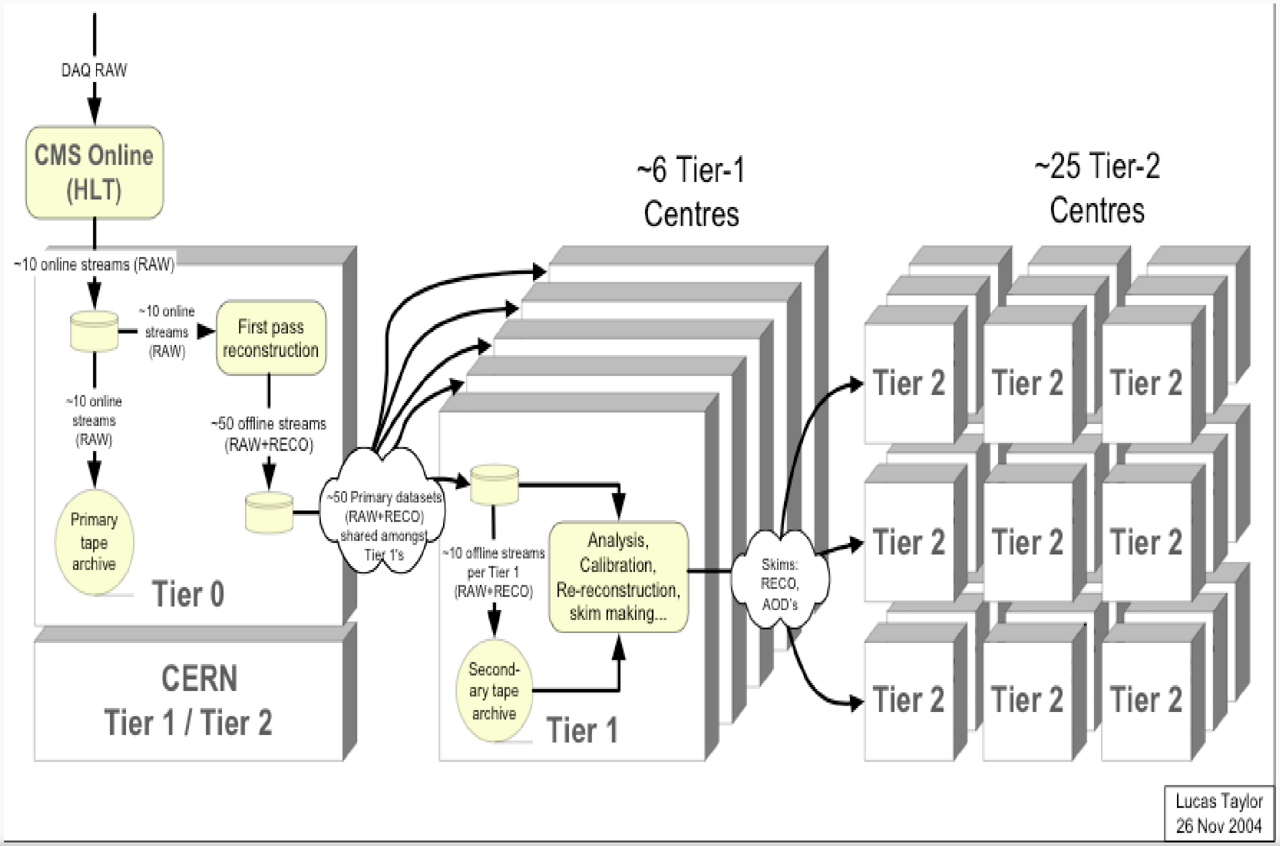
Tier-2 Planning

US-CMS Tier-2 2008	CPU	1MSI2k	100-200 dual CPU nodes
	Disk	200TB	200MB/s IO (Server numbers vary)
	Network	2.5-10 Gb/s	CERN to FNAL
	People	2FTE	Primarily Operations

The CMS Computing Model

CMS has proposed a computing model where the site activities and functionality is largely predictable

- ➡ Activities are driven by data location. Data is hosted from the Tier-1s
- ➡ Hosted data can be served to any Tier-2 center



Responsibilities of the Tier-1 and Tier-2

Tier-1 Centers serve as an extension of the experiment on-line computing

- ➡ Share of raw data for custodial storage
  - ● Second copy of the raw data is distributed to Tier-1 centers
- ➡ Data Reprocessing
  - ● CMS anticipates 2 reprocessing runs per year

They are entrusted with serving the data entrusted to them to Tier-2s

- ➡ Selecting and Skimming data for User Analysis and Calibration Tasks
- ➡ Data Serving to Tier-2 centers for analysis

Tier-2s are the only resource for simulation primary resource for analysis

- ➡ Expected to support about 40 physicists for analysis
- ➡ Total simulated event production is roughly the same

Facility Services

Grid Interfaces:

- US-CMS Supports both LCG-3 and the OSG-0.4 releases
  - Two doors into the same physical hardware at Tier-1
    - Cluster utilization at Tier-1 is roughly half grid submission and half local jobs

Processing:

- All Tier-1 resources were switched to a Condor based system in 2005
  - Cluster is scaling reasonably well. Priority scheduling allows reasonable allocation of resources.
    - Currently 1800 batch slots. Some initial issues when the farm was doubled in 2006

Storage:

- dCache/Enstore deployed for Mass storage
  - The dCache system has performed well under heavy load
    - Over 200TB delivered to applications in a single day

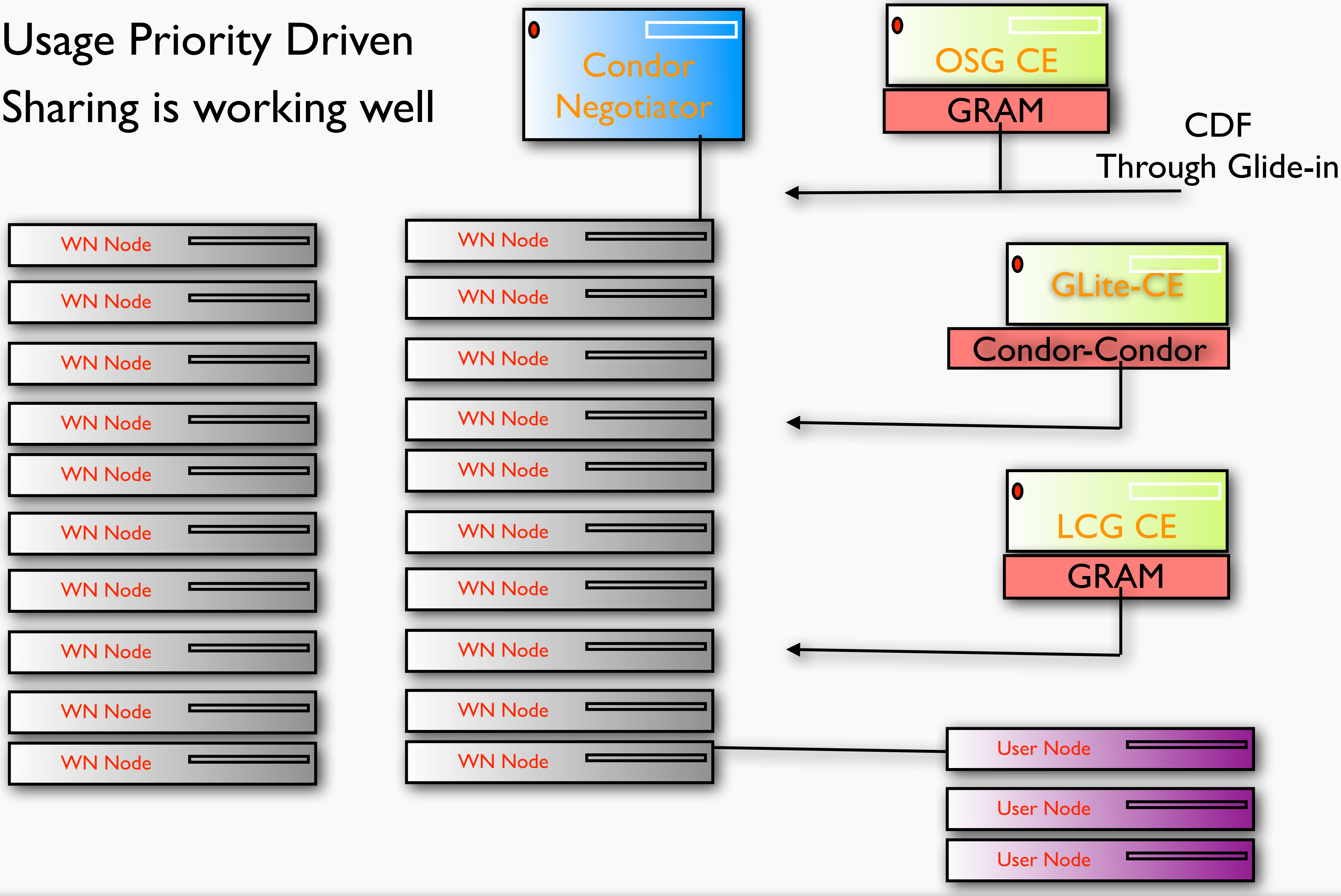
Networking:

- Current we have access to 10Gb networking at all but 2 sites

Grid Deployment

LCG and OSG have individual gatekeepers

- Usage Priority Driven
- Sharing is working well



Processing Resources

FNAL is currently at ~700 dual CPU nodes (1800 Batch slots)

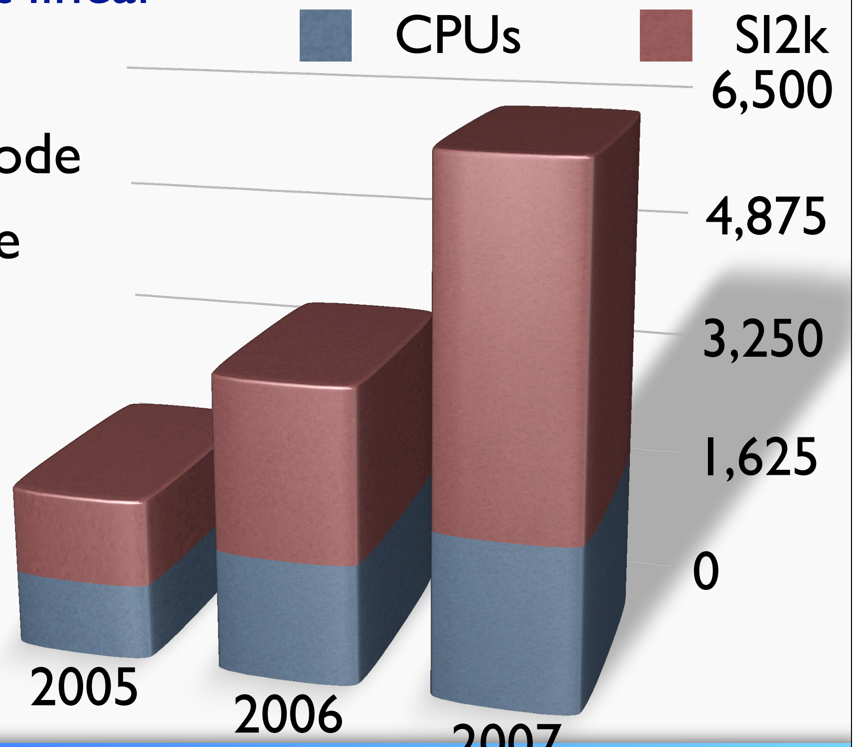
- Slowest are 2.4GHz Xeons and the Fastest are dual core Opteron 270s
- Facility is ~2000kSI2k (50% of the expected capacity in 2008)

The operational ramp to the start of the experiment is manageable

- Experience at FNAL configuring and running farms this size

The increase in number of nodes is almost linear

- Dual cores helped with ramp
  - We had an 18 month doubling per node
- Not clear that quad-cores make sense this year
  - Cost per node is high
  - Initial indications for application scaling to 4 processes looks OK



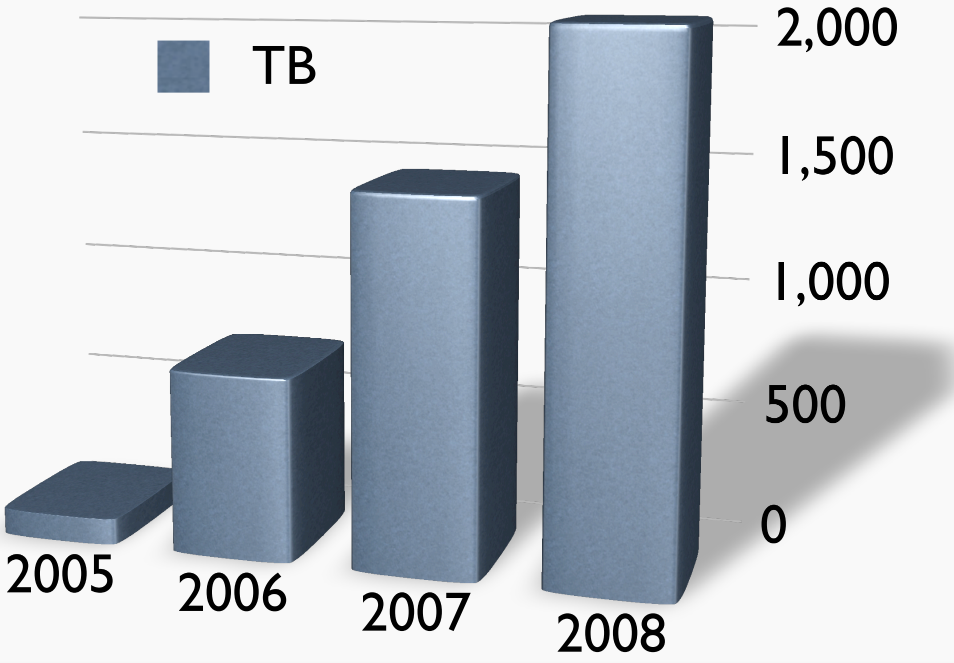
Storage Resources

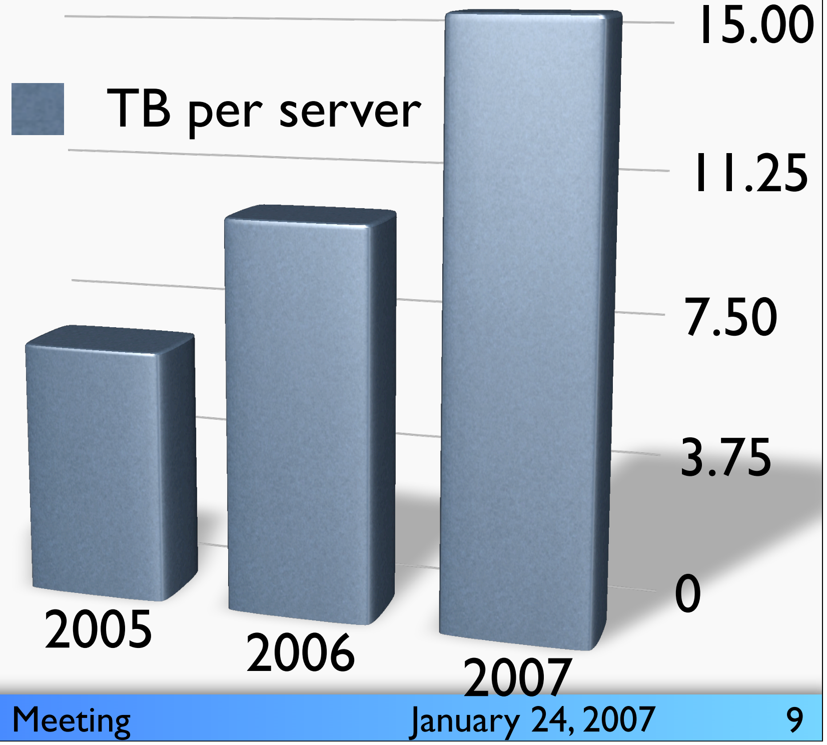
Currently the FNAL Tier-1 has ~700TB of dCache storage

- Roughly 33% of the capacity expected in 2008,

Reasonably steep operations ramp in disk storage before the experiment start

- Procuring, deploying and commissioning at a large scale





Associated Tier-2s

Tier-2s also at nearly 50% capacity. Facilities dedicated to CMS

- Sum of Tier-2 capacity is similar to the total Tier-1, as indicated in the model
- Tier-2 networking is in good shape

Site	CPU (kSI2K)	Disk (TB)	WAN (Gb/s)
Caltech	586	60	10
Florida	519	104	10
MIT	92	54	1
Nebraska	650	70	0.6
Purdue	743	184	10
UCSD	318	98	10
Wisconsin	547	110	10
TOTAL	3455	680	

Facility Planning

The progressive growth of the Tier-1 and Tier-2 centers has been very helpful

- ➡ Provided CMS with critical resources for preparation
- ➡ Gained operations experience
- ➡ Identified and solved scaling limitations with grid and facility services

Original goal had been to arrive at the nominal Tier-1 and Tier-2 sizes by the end of 2007

- ➡ The Tier-1 Facility is 4.3MSI2K, 2.0PB of disk, 4.7PB of tape
  - Additional analysis resources for the local community

The original plan had high energy running as a pilot in 2007 with low luminosity running in 2008 for 10^6 seconds

- ➡ The accelerator schedule released this year calls for a few weeks of colliding beams at 900GeV in 2007
- ➡ Low luminosity running for half as long in 2008, starting in the summer

Modified Facility Plan

In order to maintain a reasonable commissioning ramp, while trying to be the most economical

- We will deploy a 60% capacity facility by September of 2007
- Reach the nominal facility capacity by September of 2008

In 2007 we will have

- 3.9MSI2k for the Tier-1
- 1.4PB of disk

We believe this capacity will be more than sufficient to meet the data requirements of 2007

- Allows the continued ramp of facility services in preparation for 08/09
- 2006 facility cost projections were very accurate
  - Farther out projections in 2007 had more uncertainty
    - 6 month shift will help ensure the facility has some flexibility

Similar shift in US-Tier-2 Centers

Infrastructure

The US-CMS T1 relies heavily on infrastructure provided by FNAL

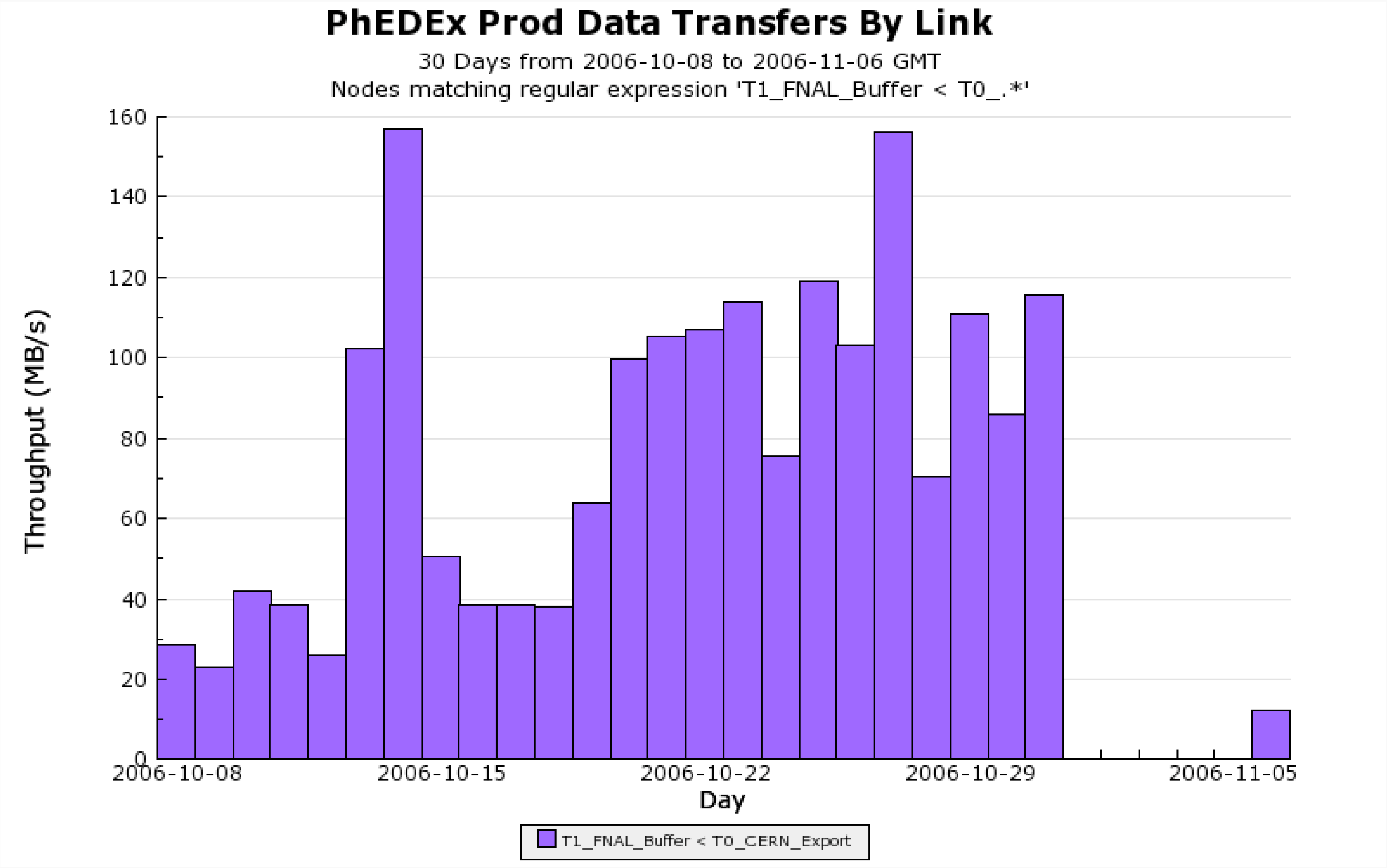
- ➨ We have been well provisioned with space and adequate power for the hardware resources
  - Tape robots are a central shared resource of the lab.   Support and expertise are excellent
  - Facility infrastructure is a common operation.   It's a constant struggle to provide the various stake holders with adequate power and cooling

Networking Provided by ESNET and US-LHCNet

- ➨ Networking situation for the Tier-1 has improved with the creation of the Chicago Metropolitan Area Network
  - Having access to the FNAL provided research link to StarLight was critical to preparation activities
  - MAN Should improve reliability and provide access to a production 10Gb/s infrastructure
    - Rely on US-LHCNet for trans-Atlantic networking to CERN

Experience From CSA06

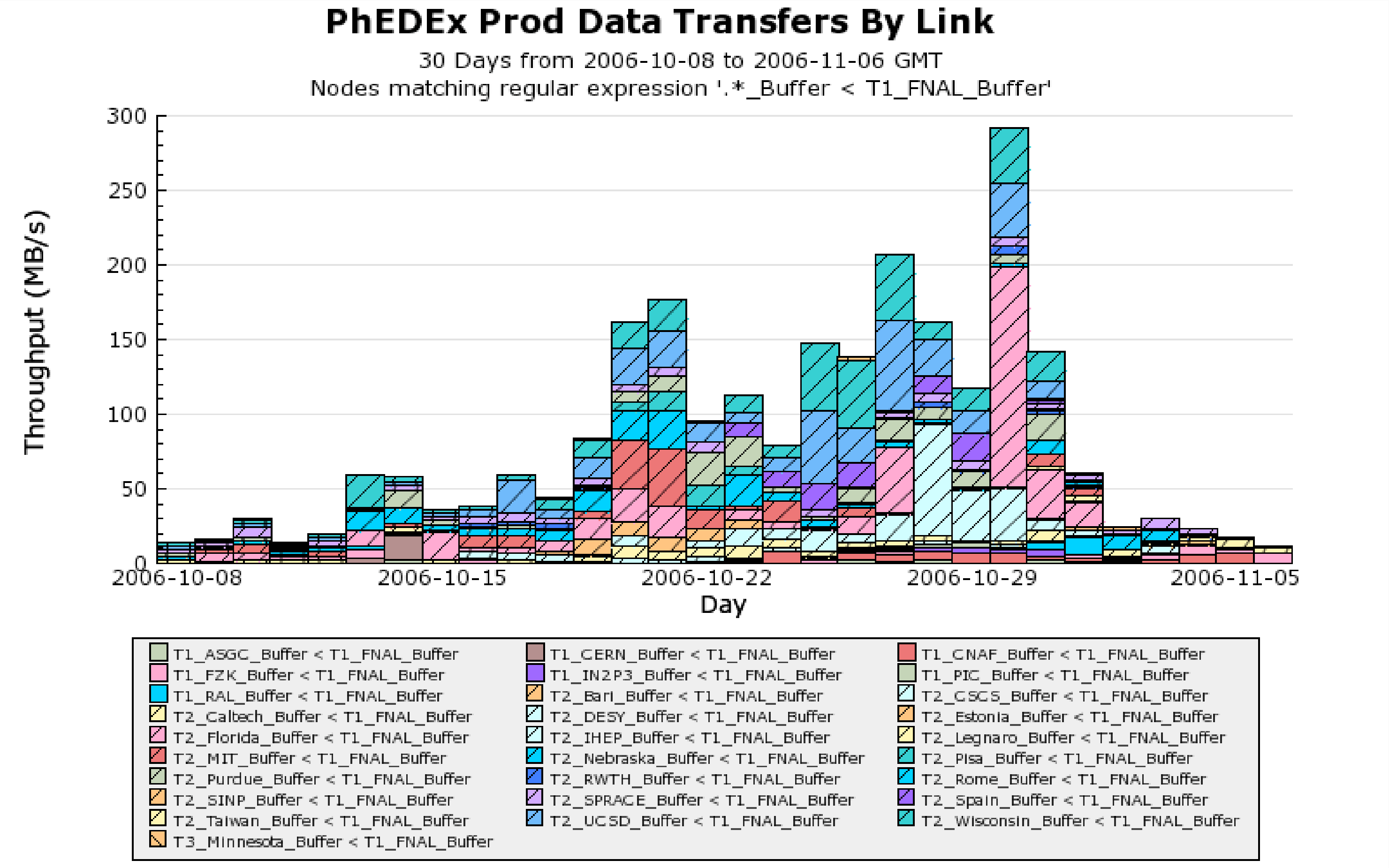
Ingest Rates from CERN



Experience From CSA06

Export Rates from FNAL

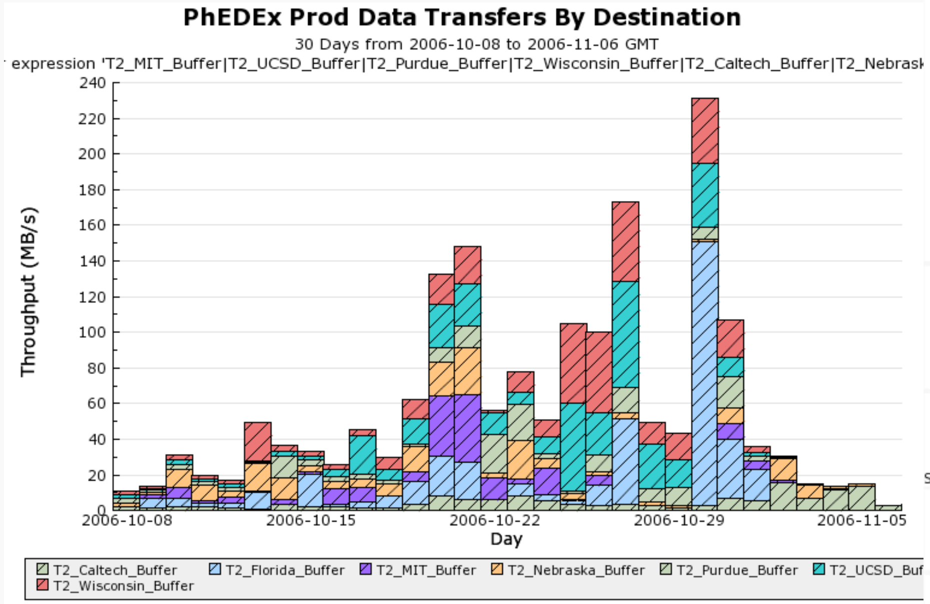
- Transfers to all Tier-1s and 21 Tiers

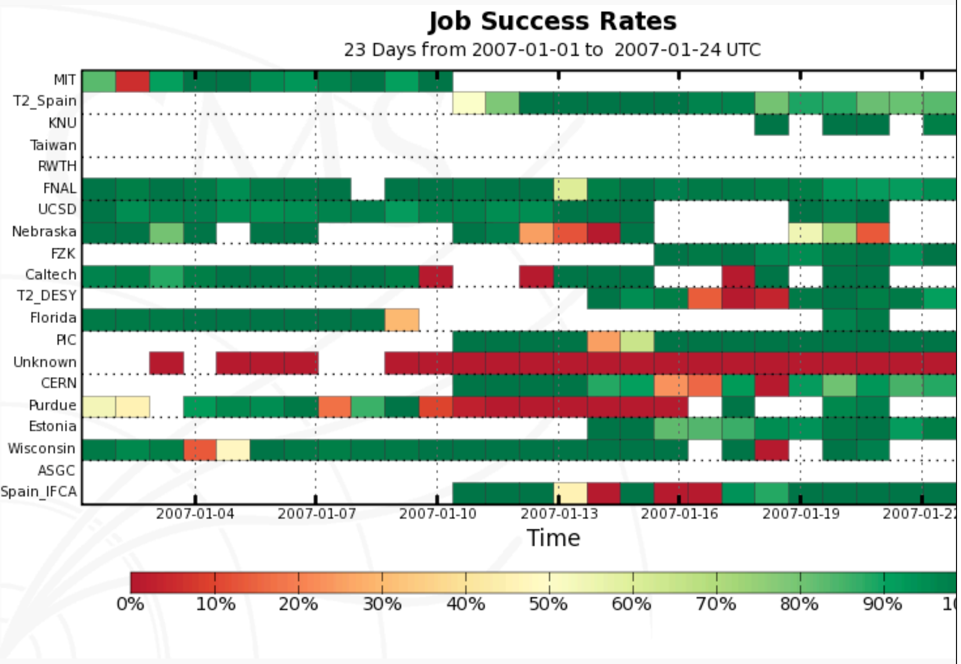


Tier-2 Experience

All US-CMS Tier-2 centers participated in the challenge

- ➡ US Tier-2 sites were among the highest burst transfer rates
  - UW had sustained 300MB/s, UCSD sustained at over 200MB/s, UFL greater than 150MB/s
    - Target maximum burst rate for CSA06 was 100MB/s

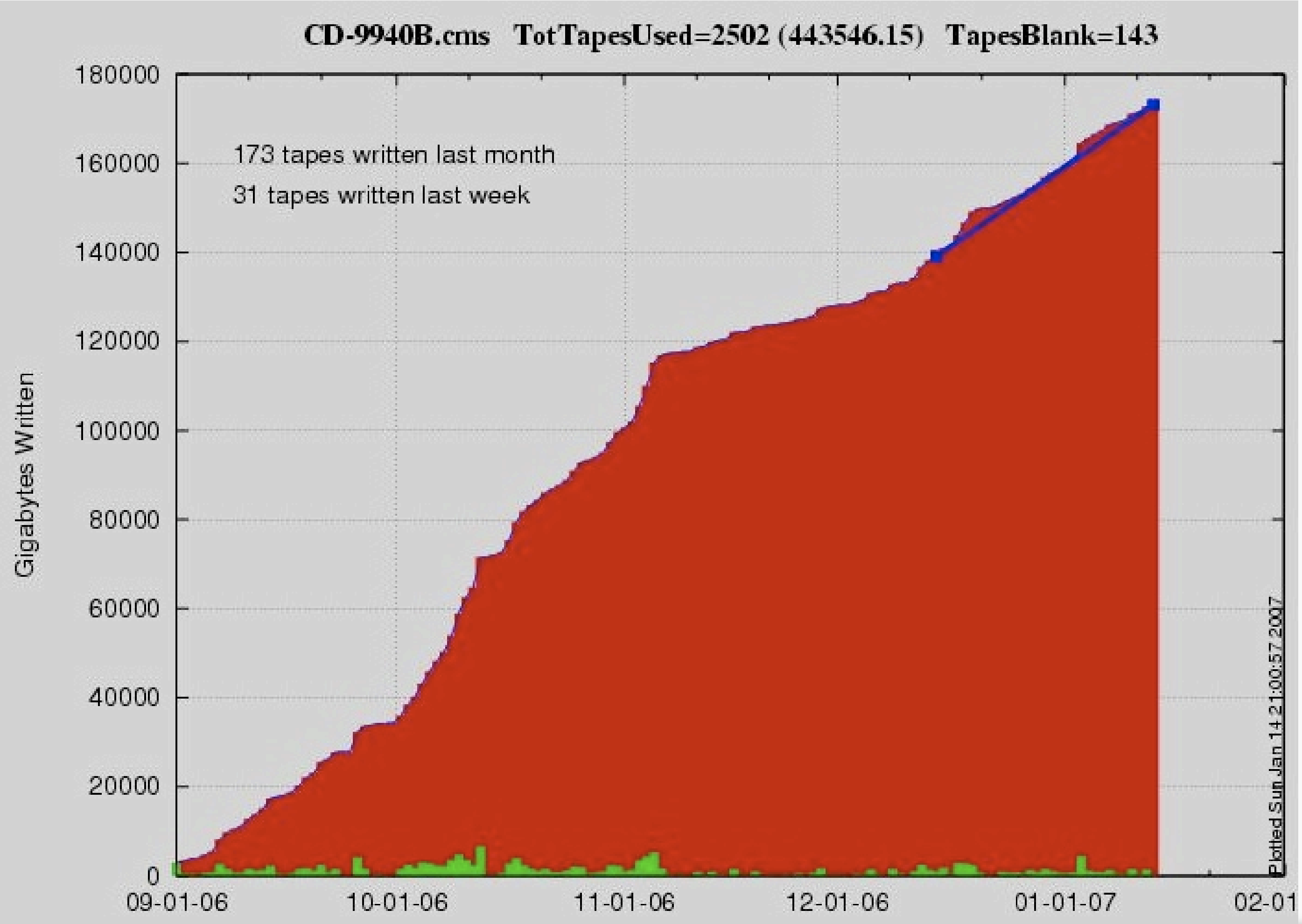




Tape Rates

Goal in CSA06 was for each Tier-1 to write the data to tape if possible

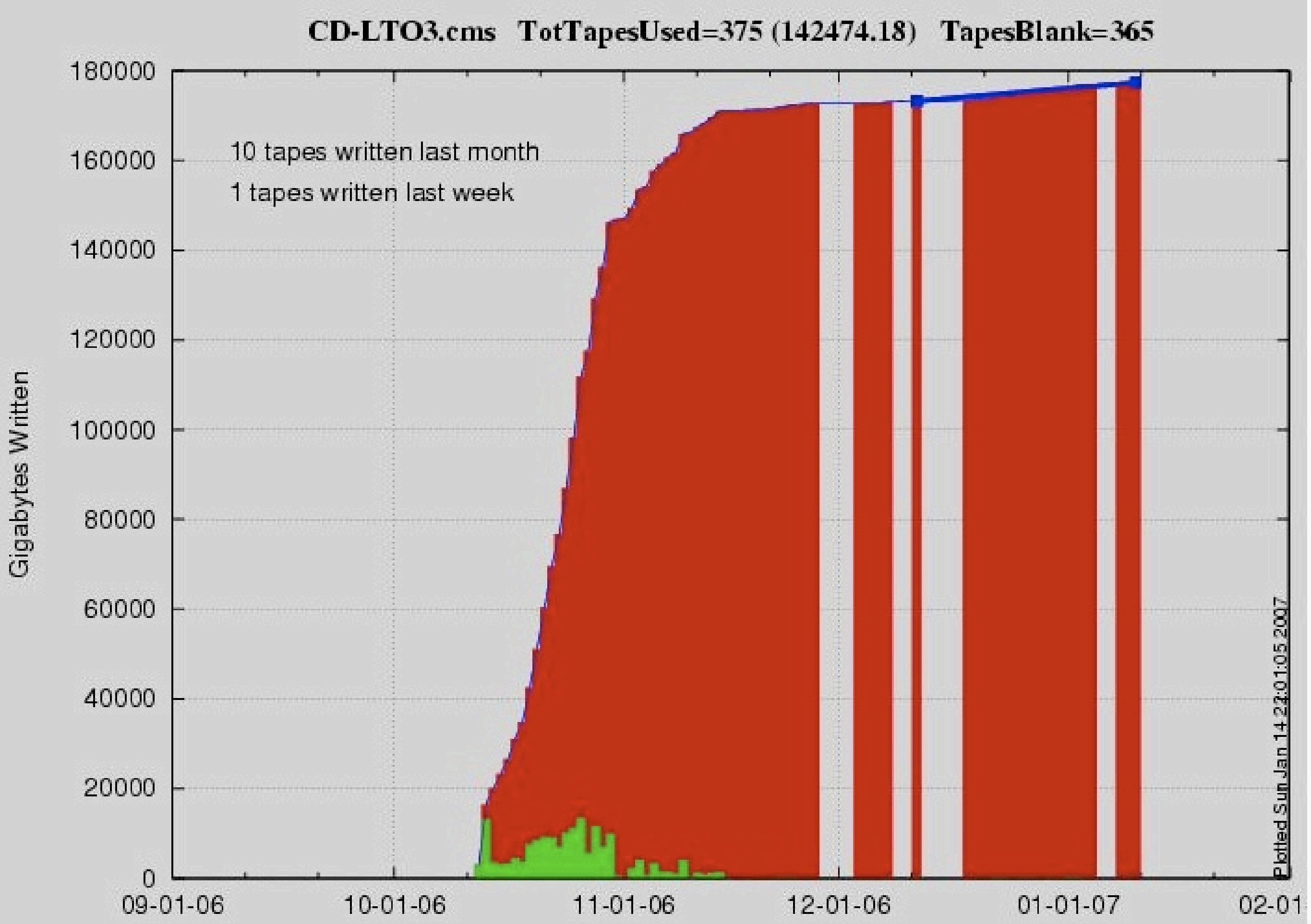
- Using the old tape robots we averaged about 1.2TB per day to tape



Tape Rates

We ran low on 9940B tapes during the challenge and switched to the newly commissioned robot stocked with LTO-3 tapes and drives

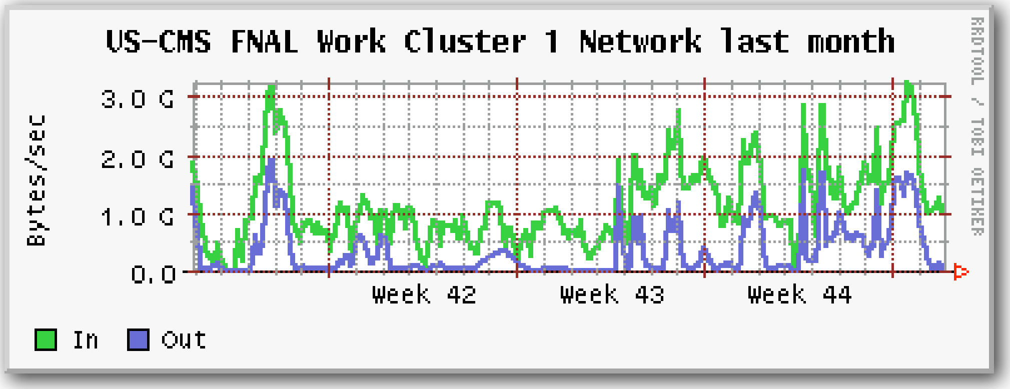
- Opening weekend was 200MB/s, 140TB written in 14 days



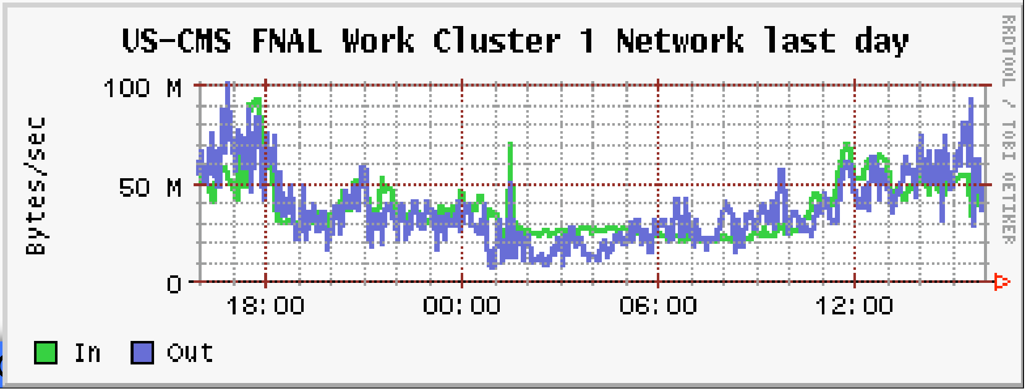
Stress Testing dCache

dCache Storage

- The previous CMS Application had a feature that caused the buffer to be inefficiently used in dCache. It has since been fixed.
- It was an excellent performance test of the system
  - Sustained periods of 2 and 3 gigabytes per second. More than 200TB served in a day



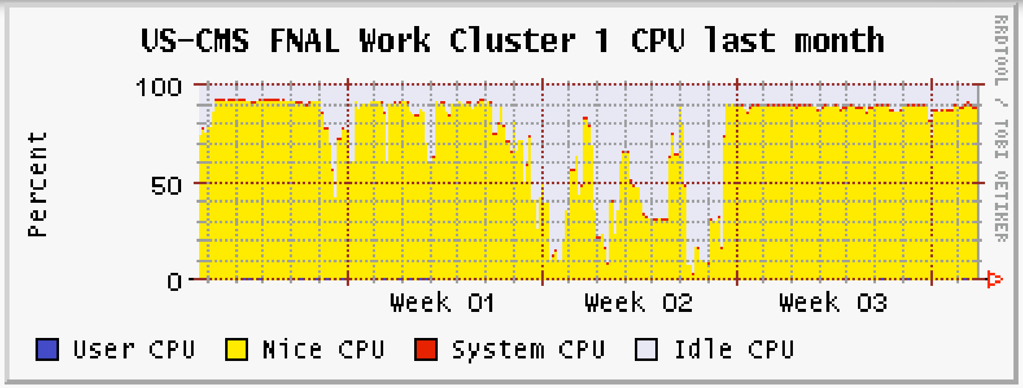
New application using client more efficiently



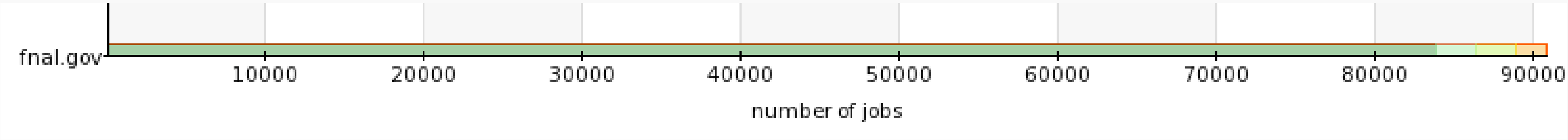
CPU Utilization

CPU utilization is looks OK.

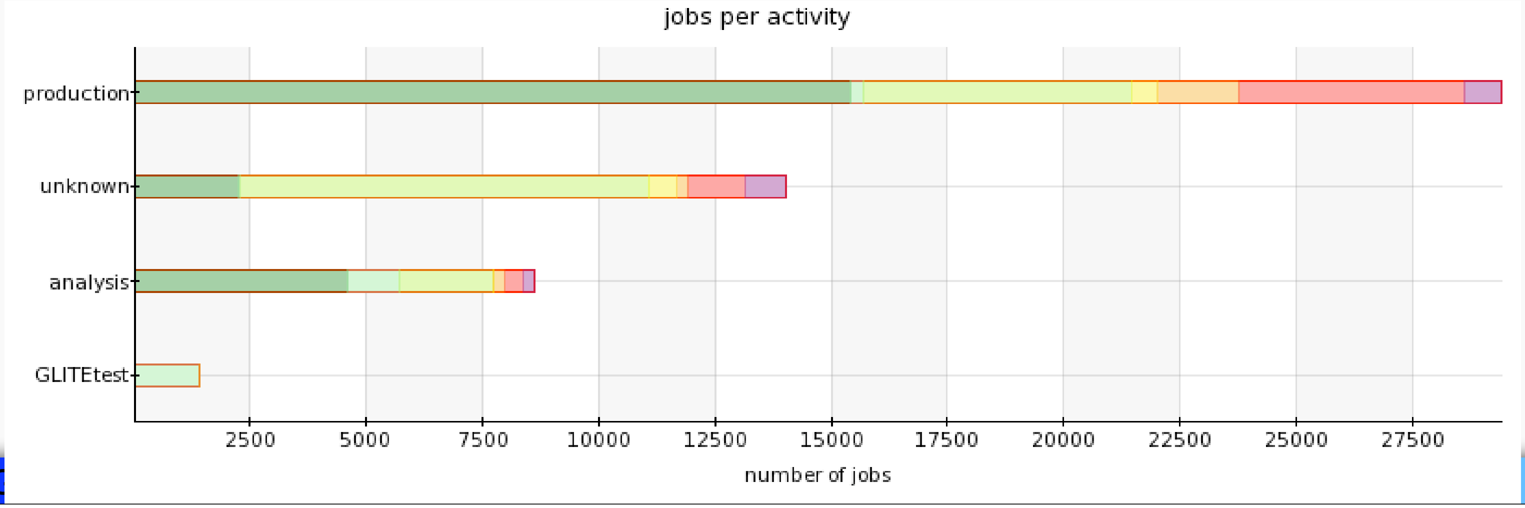
- 1800 batch slots used



OSG 0.4 Production



LCG-3 Production

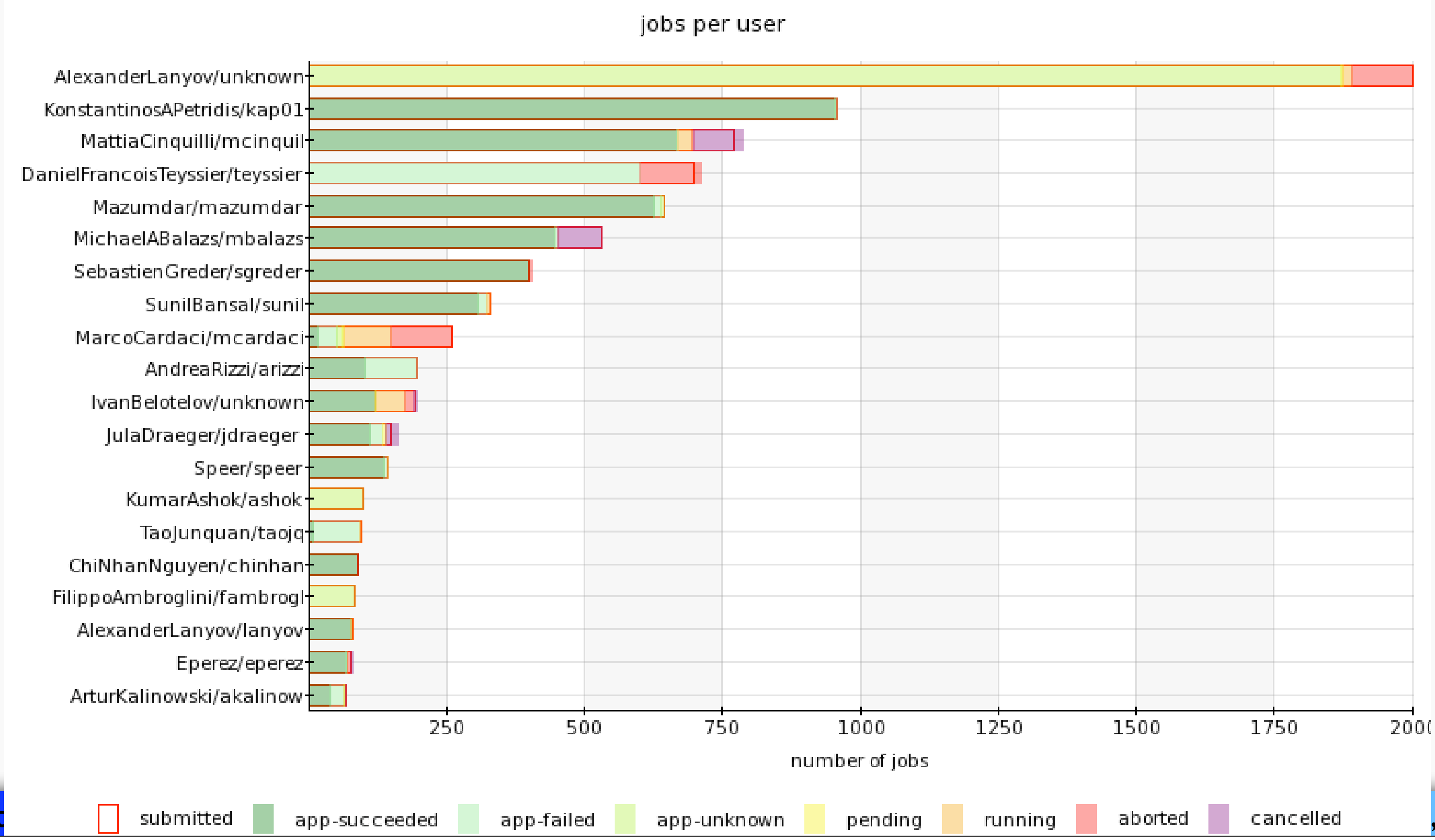


Growing User Load

There are over 500 individuals signed up for interactive access to the CMS farm at FNAL

- ➡ Somewhat ahead of our projected ramp for users.
- ➡ At any given time 10-15% of those are actively running jobs.

Grid Submissions are also increasing



Outlook

US-CMS Facilities are Growing

- Procurement ramp for the final two years is steep, but we expect to be ready with the appropriate sized center for high energy data

Grid, Storage and Processing Services are coming on line and becoming more reliable

- Operations experience still needs to improve and some of the services need to become more robust
- Facility processing is roughly 50% of final capacity
  - Bigger increase needed in disk storage
- Services are increasing in performance and capability, but development is needed

User access and subsequent support load is increasing

- Grid usage fluctuates by the long term trend is increasing
- Trying to increase the analysis use of Tier-2 centers