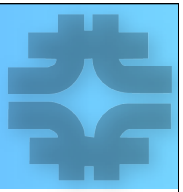


CMS Computing Model and Requirements

Ian Fisk

October 23, 2006

Introduction

CMS has had a distributed computing model from early in on. Motivated by a variety of factors

- The large quantity of data and computing required encouraged distributed resources from a facility infrastructure point of view
- Ability to leverage resources at labs and university
  - Hardware, expertise, infrastructure
- Benefits of providing local control of some resources
- Ability to secure local funding sources

~20% of the resources are located at CERN, 40% at T1s, and 40% T2s

Can only be successful with sufficient networking between facilities

- Availability of high performance networks has made the distributed model feasible

Also relies on the development and success of Grid services and interfaces

- Efficient distributed computing services

The Model

The CMS Computing Model baseline is described in the Computing Technical Design Report (CTDR)

- http://cmsdoc.cern.ch/cms/cpt/tdr/index.html

Document describes

- Explanation of computing capacity
- Interconnectivity
- Baseline services and activity descriptions

The CTDR was published in the summer of 2005, many of the concepts were also described in a computing model document that was released toward the end of 2004. Some basic concepts

- The complete experiment dataset is divided into streams
- The streams are assigned to Tier-1 centers for storage and serving
- Data is stored in files that should be accessible independently
- Processing requests go to centers the data is known to be

Input Parameters For Model

Event Sizes

- Current estimate of raw data event size is 1.5MB (1-2MB)
- Size of Reconstructed Event is 0.25MB
- Analysis Object data is 0.05MB per event

CMS best estimate is about 150Hz for the DAQ target Event rate

- ~ 250MB/s
- CMS is looking at first year scenarios with larger trigger rates

During normal CMS running we expect to log about 2PB of data per year of raw data

- About 30%-50% of that comes directly to FNAL for archiving and serving
  - 30% of raw, a larger fraction of reconstructed, and a full AOD copy
- During the first several years of the experiment the analysis will have to access more raw data
- Leads to larger data sets for analysis and larger selected datasets

CMS Computing Model

The CMS computing model is not the MONARC model circa 1998

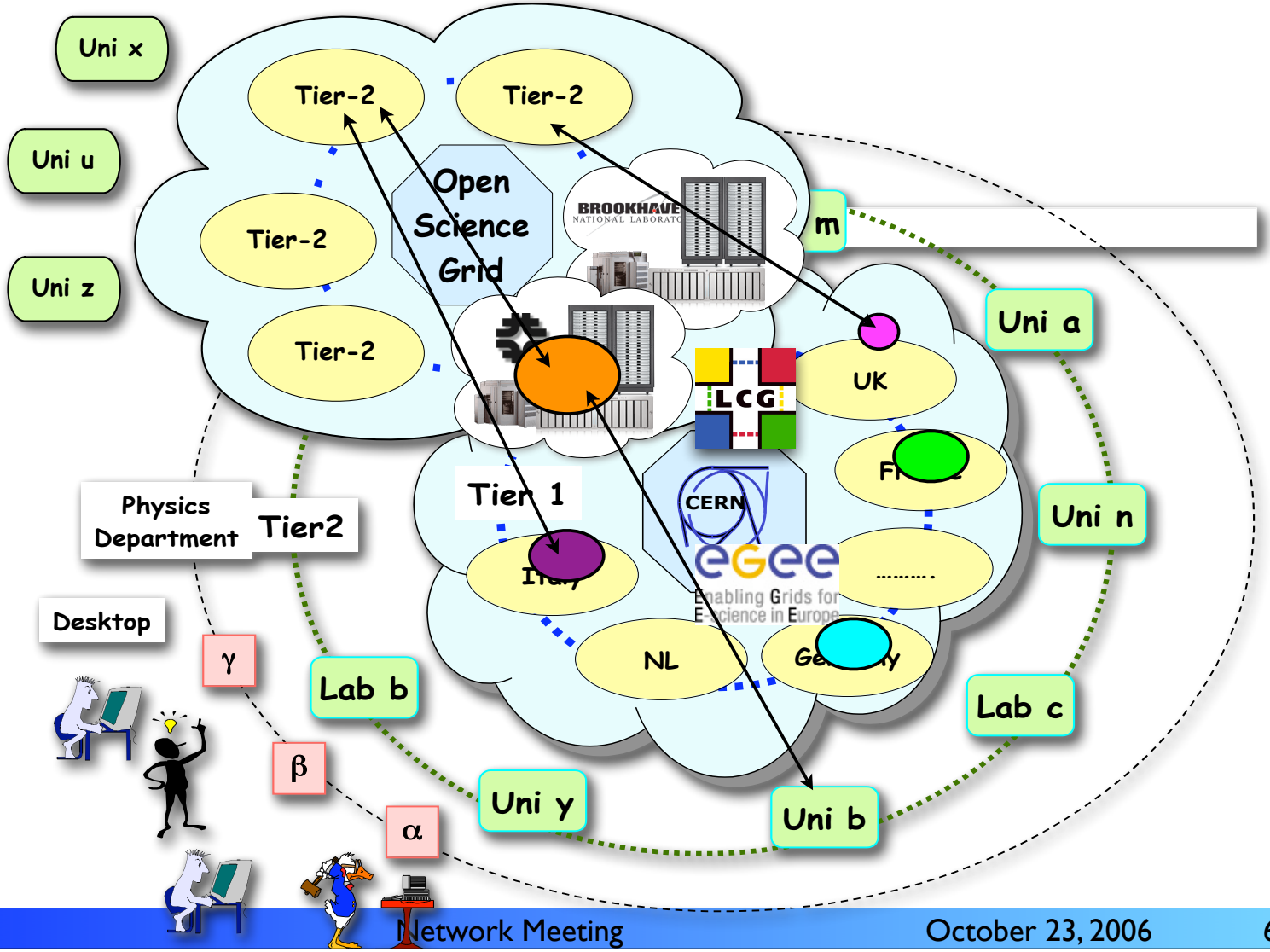
- The strict hierarchies of access do not exist
  - Tier-2 centers have to be able to connect to any Tier-1 center
  - Tier-1 centers communicate with each other

The CMS model is also not a pure grid computing cloud model

- Activities running at each tier are predictable and prescribed
  - Opportunistic computing is reserved for a very limited set of functionality
- The data location drives the activities at a site

Each colored oval represents a unique sample of data

- To analyze it each Tier-2 site needs to be able to receive it



Data Driven Baseline

Data placement drives activity at the Tier-0 and Tier-1 centers in the CMS baseline model.

- Data is partitioned by the experiment as a whole
- Tier-0 and Tier-1 are resources for the whole experiment
- Leads to very structured usage of Tier-0 and Tier-1
  - Tier-0 and Tier-1 centers are CMS experiment resources and activities are nearly entirely specified
    - Primary reconstruction, Re-reconstruction, Data and Simulation Archiving, Data and Simulation Serving, and Data Skimming

Tier-2 Centers are the place where more flexible, user driven activities can occur

- Portion of resources are controlled by the local community
- More chaotic analysis activities
- Very significant computing resources in need of good access to data

Roles and Responsibilities

Tier-0

- Primary reconstruction
- Partial Reprocessing
- First archive copy of the raw data

Tier-1s

- Share of raw data for custodial storage
- Data Reprocessing
- Data Selection
- Data Serving to Tier-2 centers for analysis
- Archive Simulation From Tier-2

Tier-2s

- Monte Carlo Production
- Analysis

Computing Center Specifications

Tier-0 Center

		Running Year 2007	2008	2009	2010	
Conditions		Pilot	2E33+HI	2E33+HI	E34+HI	
Tier-0	CPU	2.3	4.6	6.9	11.5	MSi2k
	Disk	0.1	0.4	0.4	0.6	PB
	Tape	1.1	4.9	9	12	PB
	WAN	3	5	8	12	Gb/s

Tier-1 Centers

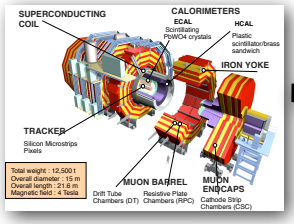
- 1/6
- US-CMS is roughly twice as large

A Tier-1	CPU	1.3	2.5	3.5	6.8	MSi2k
	Disk	0.3	1.2	1.7	2.6	PB
	Tape	0.6	2.8	4.9	7.0	PB
	WAN	3.6	7.2	10.7	16.1	Gb/s

Tier-2 Centers

US Tier-2	2008	
CPU	1	MSi2k
Disk	.200	PB
WAN	1-10	Gb/s

Data Flows



Tier-0

Tier-0 to Tier-1 Flow
Predictable
High Priority

Tier-1 to Tier-1
Burst with Rereco
Load Balancing

Tier-1

Tier-1

Tier-1

Tier-1 to Tier-2
Burst with User
Needs

Tier-2

Tier-2 to Tier-1
Predictable
Simulation Data

Tier-2 centers may have relationships with Tier-1 centers for management, support, and operations

- Data access may come from a variety of Tier-1 centers

Networking Estimates

The network requirements for Tier-0 to Tier-1 transfers are driven by the trigger rate and the event size

- Estimates are ~2.5Gb for a nominal Tier-1 center
  - The Tier-1 event share with a factor of 2 recovery factor and a factor of 2 provisioning factor

The Tier-1 to Tier-1 transfers are driven by the desire to synchronize re-reconstruction samples within a short period of time

- To replicate the newly created reconstructed and AOD between Tier-1 centers in a week is 1Gb/s, before the safety and provisioning factors

The Tier-1 to Tier-2 transfers are less predictable

- Driven by user activities.
- CMS model estimates this at 50-500MB/s (Includes safety factors)

Tier-2 to Tier-1 transfers are predictable in rate and low in volume

- Averaged over the entire it's ~1TB per day. Can go to any T1

Tier-2 Centers

Tier-2 computing centers represent the bulk of the analysis computing resources for the experiments

- In the early years of the experiment serious analysis may require frequent access back to the raw data samples
  - Making selections and moving the data to Tier-2s for detailed analysis
- Since each Tier-1 center only serves a portion of the raw data, the connections from a Tier-2 can go to any Tier-1
- The US Tier-2s will need to access European Tier-1s for data samples
- FNAL is the single largest Tier-1 center and European and Asian T2s will need to transfer data from it.

Data transfers have bursts

- The data requirements are driven by how frequently the Tier-2 cache needs to be updated and how long users are willing to wait for a transfer to be completed.

Transfers Across the Atlantic

At the current rate of tape pledges, roughly 50% of the CMS Data ends up at FNAL for arching and serving

70% of the CMS Tier-2 centers are located outside the US

The US estimates for Tier-2 connectivity is 2.5Gb/s

There is a lot of Tier-2 connectivity from Europe to the US-CMS Tier-1

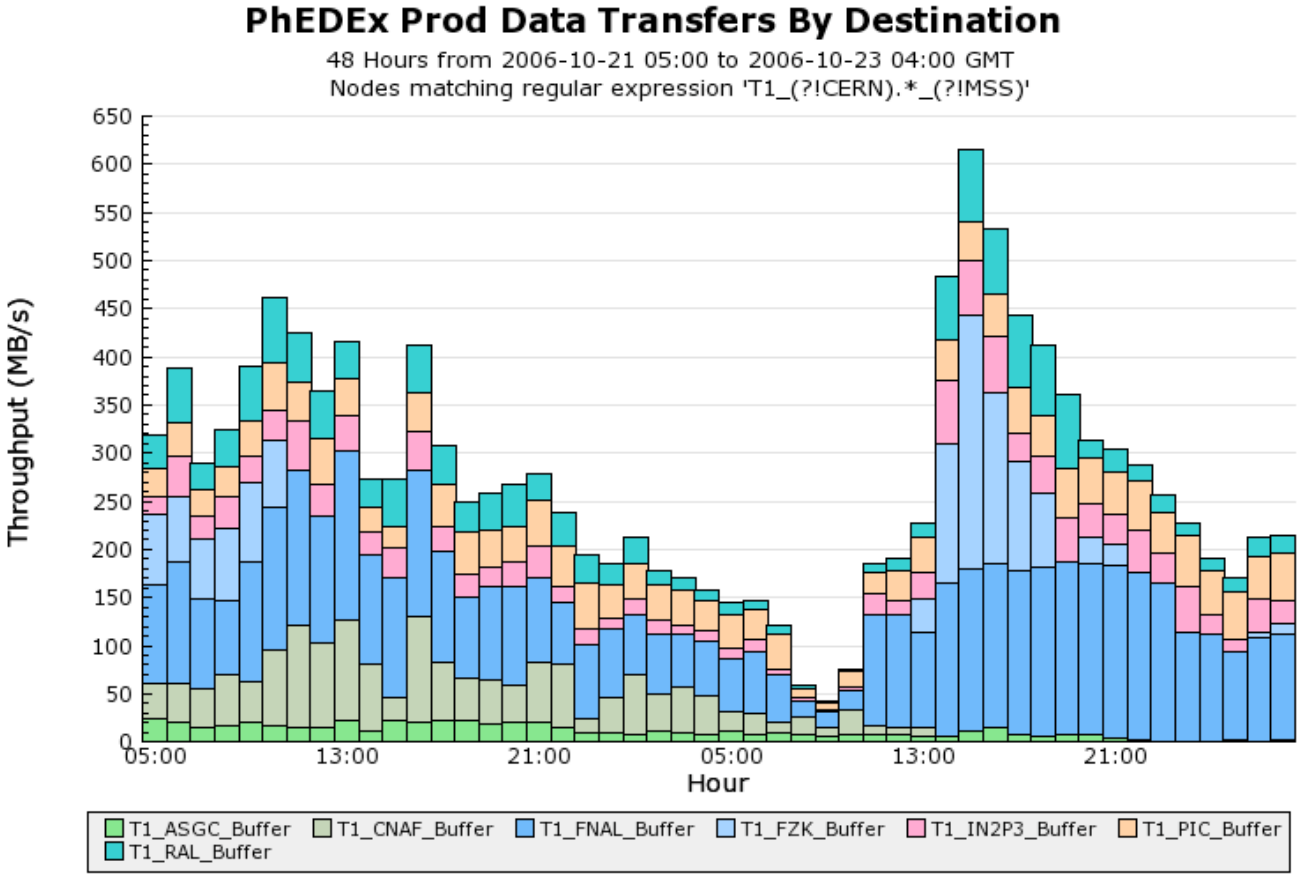
The US Tier-2s are large and there is 50% of the data located outside the US.

Utilizing the Networks (T0-T1)

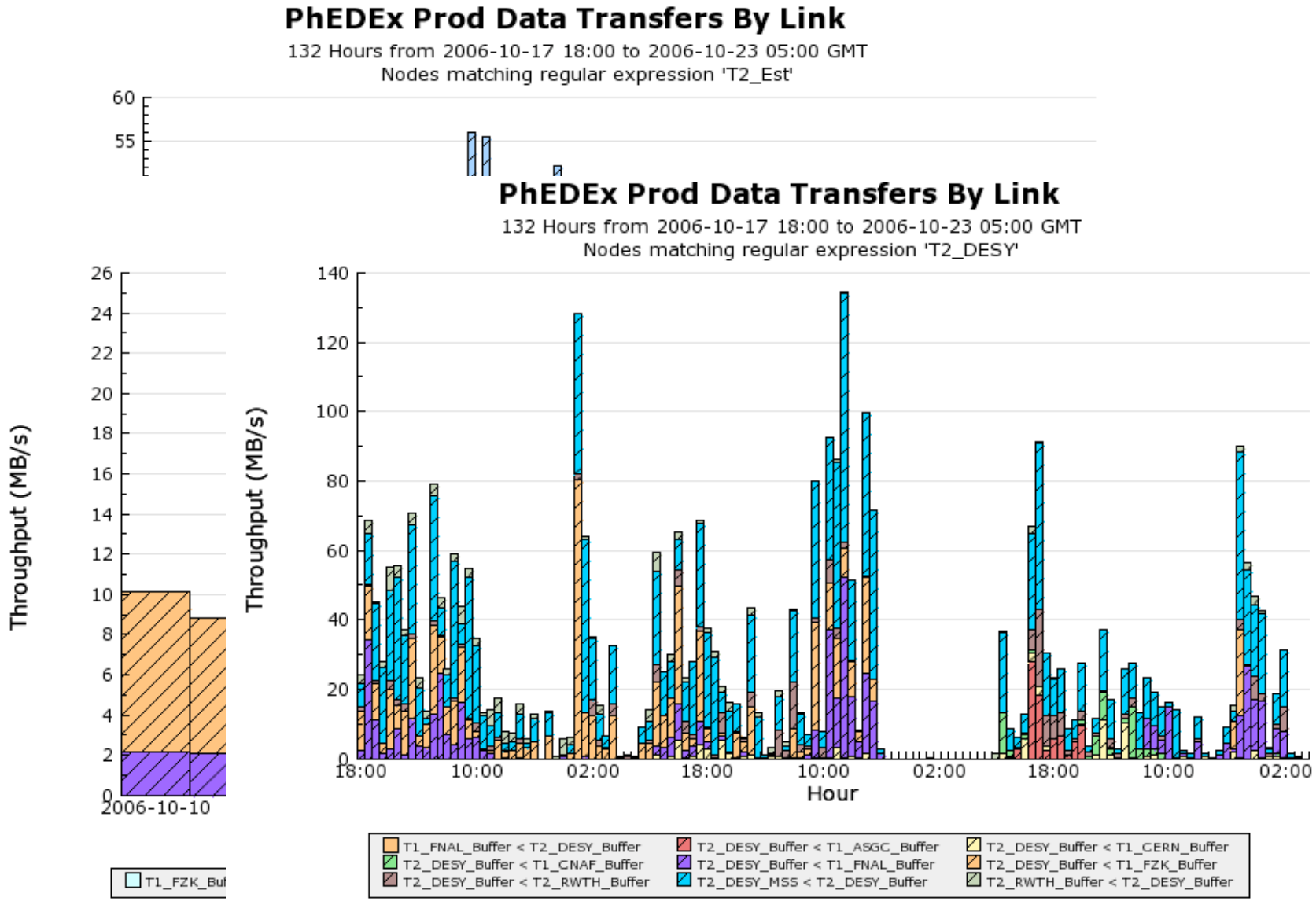
CMS is currently in the middle of the 2006 data challenge (CSA06)

- This is a 25% computing challenge. The sustained goal rate of Tier-0 to Tier-1 transfers is 150MB/s
  - During the challenge we went beyond the initial goals

- In the challenge we want to show stable operations with good efficiency
- We also want to demonstrate we can recover from interruptions



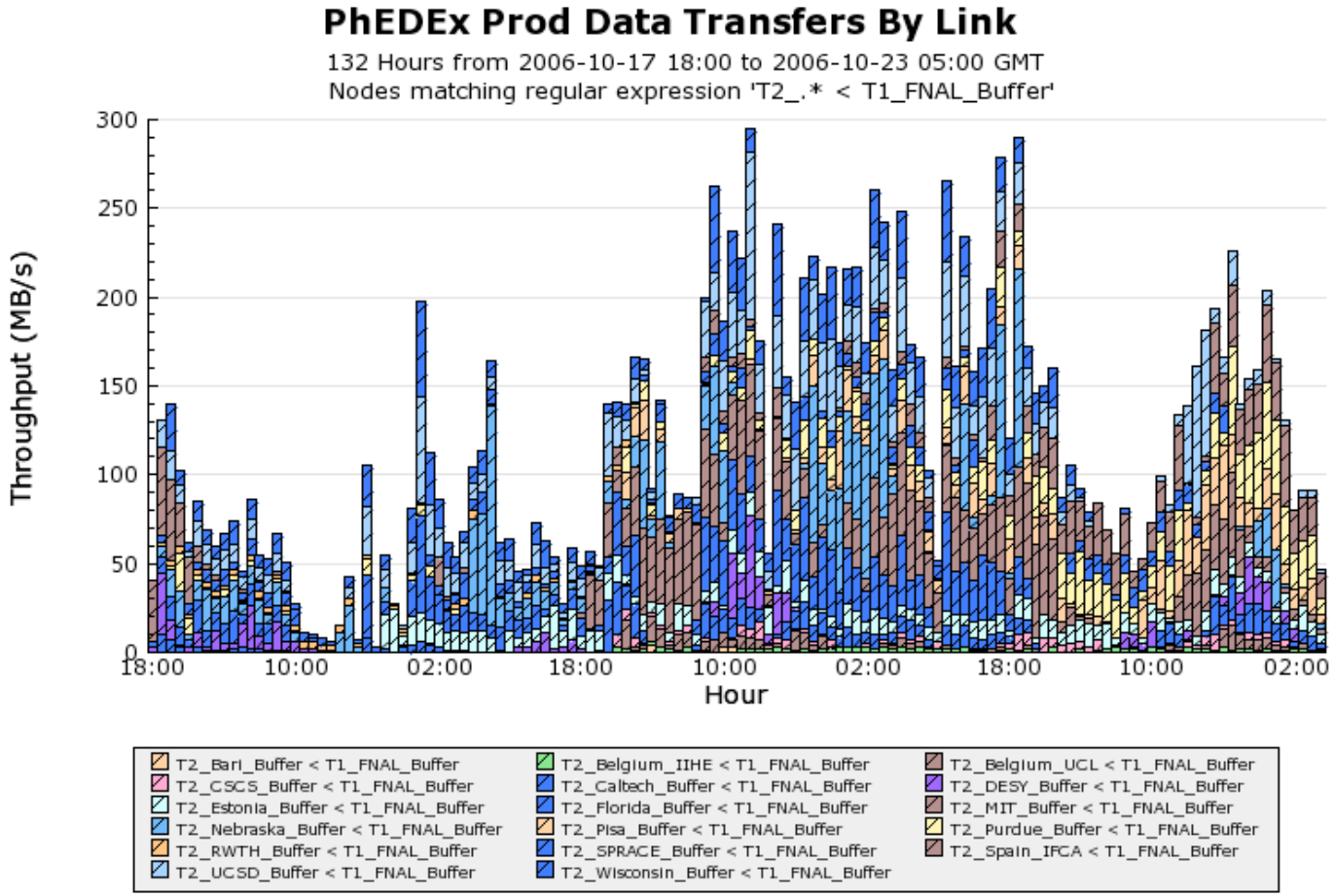
Utilizing Networks (T1-T2)



T1 Transfers from FNAL

In the last 5 days FNAL has transferred to 16 Tier-2 centers

- On three continents



Outlook

CMS has chosen a globally distributed computing model

- Majority of computing resources are located away from the host lab

CMS has chosen a model that drives activity at the computing tiers based on data distribution

- Maintains realistic expectations on Grid services and facilities
- Room for future growth of services and flexibilities

The model relies on reasonable networking to succeed

- Larger available networks provide for flexibility of site activities by enabling fast transitions

CMS is demonstrating aspects of the model at a 25% scale in CSA06