



# Introduction to the Computing Model and the LPC CAF

Liz Sexton-Kennedy for Ian Fisk  
June 20, 2008



# 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

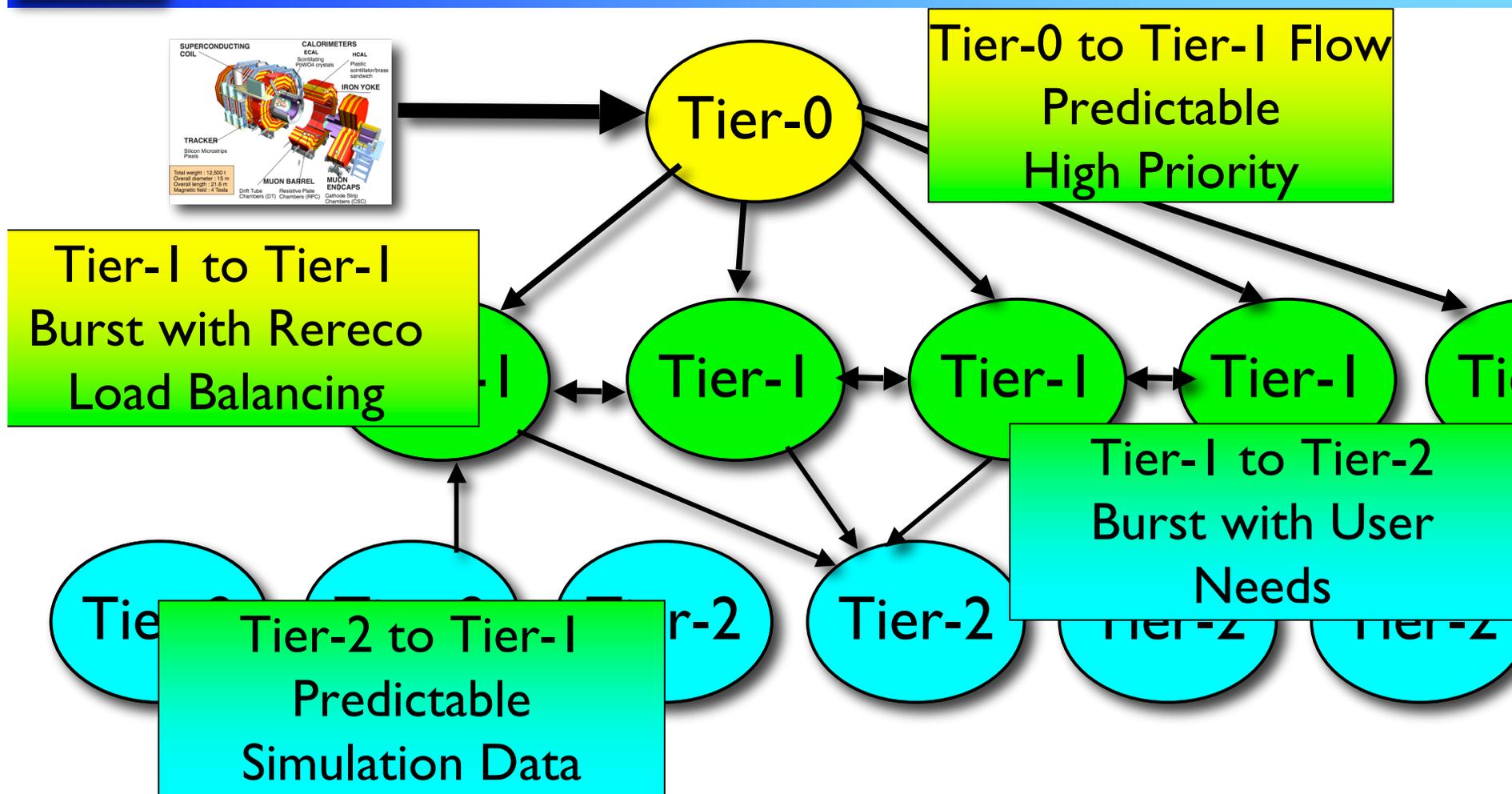
~20% of the resources are located at CERN, 40% at T1s, and 40% T2s

Relies on the grid services to provide consistent interfaces to clusters and tools to make submissions and data transfers easy for users.

Can only be successful with sufficient networking between facilities

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

# Data Flows

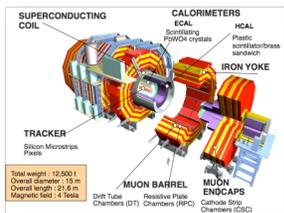


Tier-2 centers may have relationships with Tier-1 centers for management, support, and operations

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



# Workflows



Tier-0

Tier-0 only used for organized processing  
- No user access

Tier-1

Tier-1

Tier-1 centers perform reprocessing and skimming  
- Primarily organized, though may be specified by users  
- Grid Access

Tier-0

LPC CAF is entirely intended for analysis  
- Currently only log in access

LPC-CAF

Tier-2

Tier-2

Tier-2

Tier-2

Tier-2

Tier-2

Tier-2 centers are roughly half simulation and half analysis  
- User resources  
- Grid access, though some may offer log in



# 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



# Data Driven Baseline

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

- ➔ Data is partitioned by the experiment as a whole
- ➔ Tier-0 and Tier-I are resources for the whole experiment
- ➔ Leads to very structured usage of Tier-0 and Tier-I
  - Tier-0 and Tier-I 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



# 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
- ➔ Full mesh of Tier-1 to Tier-2 connectivity is needed
- ➔ Tier-2 centers are associated with 1 or more analysis groups, DPG + PAG

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.



# US Sites

The US has a Tier-I Center at FNAL

- ➔ FNAL is the largest Tier-I center in CMS
  - One of 7
  - The only Tier-I center in the Americas
    - The center will reach 6MSI2k, 2PB of disk, 4.7PB tape
    - WAN network is 20Gb/s

The US has 7 Tier-2 computing facilities

- ➔ Caltech, Florida, MIT, Nebraska, Purdue, UCSD, Wisconsin
  - All are on target to meet the CMS specifications for a fully functional Tier-2 facility
    - At least 1MSI2k of computing, 200TB of disk
    - All US sites now have 10Gb/s network links

In addition FNAL has an analysis farm called the LPC CAF

- ➔ Much like a Tier-2 in terms of analysis
  - 3MSI2k, 0.5PB of disk



# Data Management

CMS data is divided into

- ➔ Datasets - A group of events that can be accessed together
- ➔ Data Blocks - A group of files that for an object that is tracked by the data transfer system.
  - A dataset is a group of file blocks
- ➔ Logical file names - A group of events that can be accessed independently

The Dataset Discovery Page is at

- ➔ [https://cmsweb.cern.ch/dbs\\_discovery/](https://cmsweb.cern.ch/dbs_discovery/)
  - Wildcard searches
  - Select on software releases

Data is tracked by DBS

Data is also divided by tier

- ➔ RAW, RECO, AOD
  - Datasets are skimmed and reduced

The screenshot shows a web browser window titled "DBS data discovery page" with the URL "https://cmsweb.cern.ch/dbs\_discovery/". The page has a navigation bar with links for "Frontier CMS Server", "PhEDEx", "Lassi Plan", "CCS", "Lothar's Calendar", and "FNAL Data". Below the navigation bar is a "Dashboard" section with tabs for "Dashboard", "DBS Discovery", "ProdRequest", "PhEDEx", and "SiteDB". The main content area is titled "DBS discovery :: Navigator" and contains several search filters: "Physics groups" (Any), "Data tier" (Any), "Software releases" (Any), "Data types" (Any), and "Primary dataset/MC generators" (Any). There is also a checkbox for "composed tier, e.g. GEN-SIM:" and "Reset" and "Find" buttons.

The input form supports **wild-card** , **regular expressions** , **like** searches.

Auto-completion: **on** | **off**



# Data Movement

Data Movement in CMS is handled by a tool called PhEDEx

- ➔ PhEDEx replicates individual files and updates the data management system when complete blocks have been transferred
- Uses grid services and interfaces to replicate data
  - Handles retries, data integrity checks and prioritization

Data Movement to Tier-1 centers is a decision of central CMS

- ➔ Samples are entrusted to computing centers for archiving, reconstruction, and serving to Tier-2 computing facilities

Data Movement to Tier-2 centers is intended to be driven by needs or users and groups

- ➔ There will be open data subscriptions that grow as data is available
- ➔ There will be burst transfers that make dynamic use of the storage at the Tier-2s
- The majority of the storage at the Tier-2s is intended for serving experiment data



# Data Transfers

The PhEDEx page is available at

➔ <http://cmsdoc.cern.ch/cms/aprom/phedex/prod/Activity::RatePlots?view=global>

Any user can make a transfer request.

- ➔ Only a site data manager can approve a request
- ➔ Data managers make sure the site is not over subscribed

Data on Tier-2 sites is not intended to be kept forever

Production Requests - Create Request - CMS PhEDEx

Getting Started Latest Headlines Frontier CMS Server PhEDEx Lassi Plan CCS Lothar's Calendar FNAL Data Apple

PhEDEx - CMS Data Transfers

DB Instance: **Production** »  
Ian Fisk | [Sign out](#)  
Logged in via Certificate

[Info](#) [Activity](#) [Data](#) [Requests](#) [Components](#) [Reports](#)

[Overview](#) | [Create Request](#) | [View/Manage Requests](#)

### New Transfer Request

E-mail:

DBS:

Data Items:

/Primary/Processed/Tier  
or  
/Primary/Processed/Tier#Block  
(Use \* as wildcard)  
[More Help](#)

Destinations:

<input type="checkbox"/> T0_CERN_Export	<input type="checkbox"/> T2_Bari_Buffer	<input type="checkbox"/> T3_IN2P3_IPNL
<input type="checkbox"/> T0_CERN_MSS	<input type="checkbox"/> T2_Beijing_Buffer	<input type="checkbox"/> T3_IRES_Buffer
<input type="checkbox"/> T1_ASGC_MSS	<input type="checkbox"/> T2_Belgium_IHHE	<input type="checkbox"/> T3_Karlsruhe_Buffer
<input type="checkbox"/> T1_CERN_CAF	<input type="checkbox"/> T2_Belgium_UCL	<input type="checkbox"/> T3_Minnesota_Buffer
<input type="checkbox"/> T1_CERN_MSS	<input type="checkbox"/> T2_Budapest_Buffer	<input type="checkbox"/> T3_Napoli_Buffer
<input type="checkbox"/> T1_CNAF_MSS	<input type="checkbox"/> T2_CIFMAT_TMP	<input type="checkbox"/> T3_Perugia_Buffer

Transfer Type:  [What's this?](#)

Priority:  [What's this?](#)

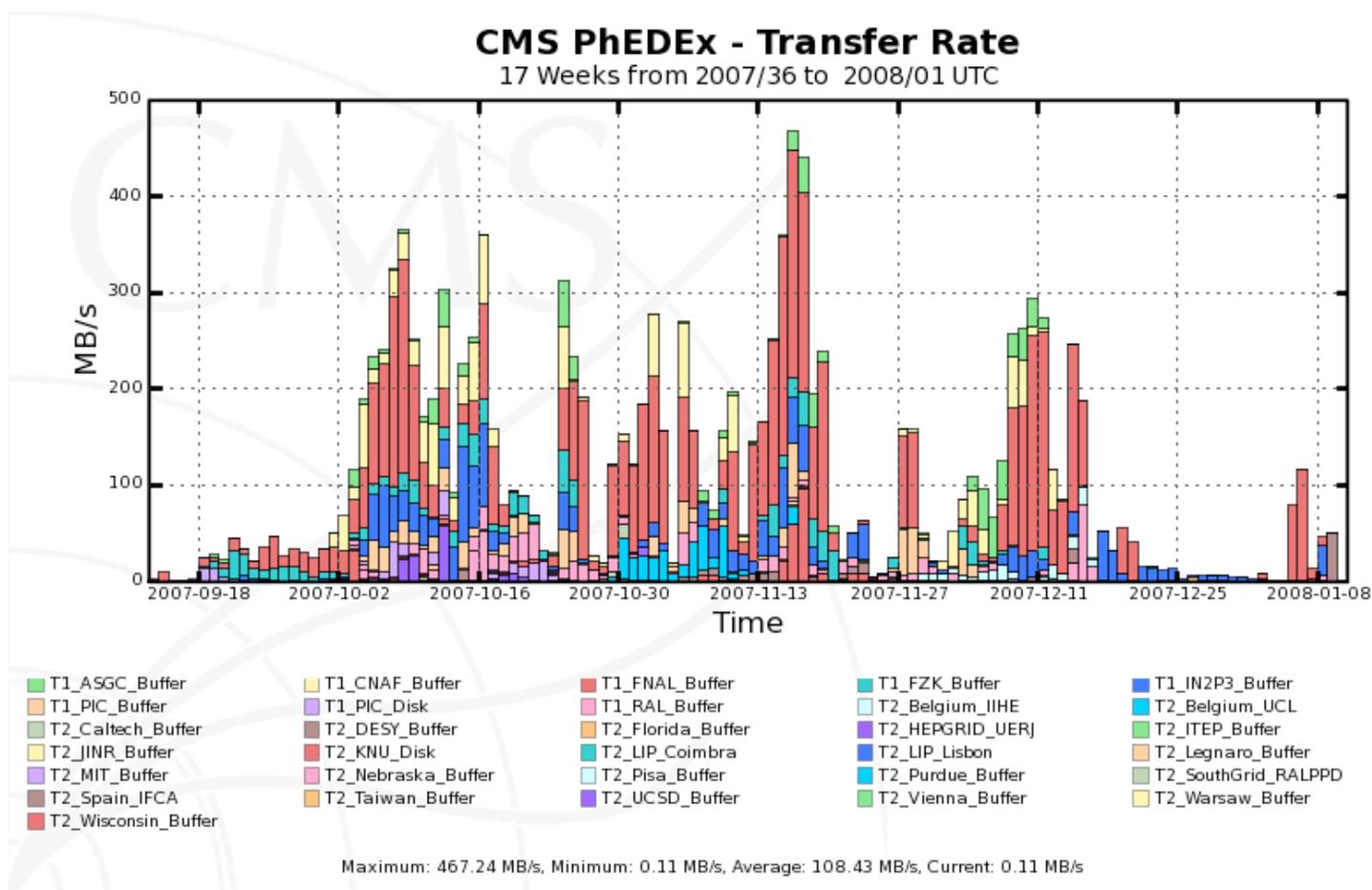
Comment:



# Data Transfers CERN to Tier-I

In the final configuration CMS expects to export peaks of 600MB/s from CERN to the Tier-I centers

## ➔ Data Transfers over the last 120 days from CERN

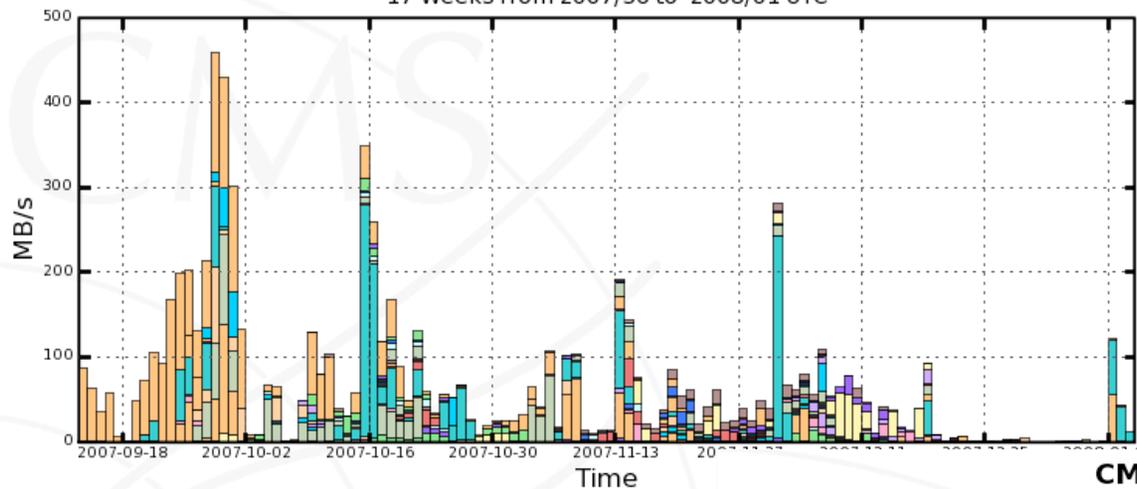




# Tier-1 to Tier-2 Transfers

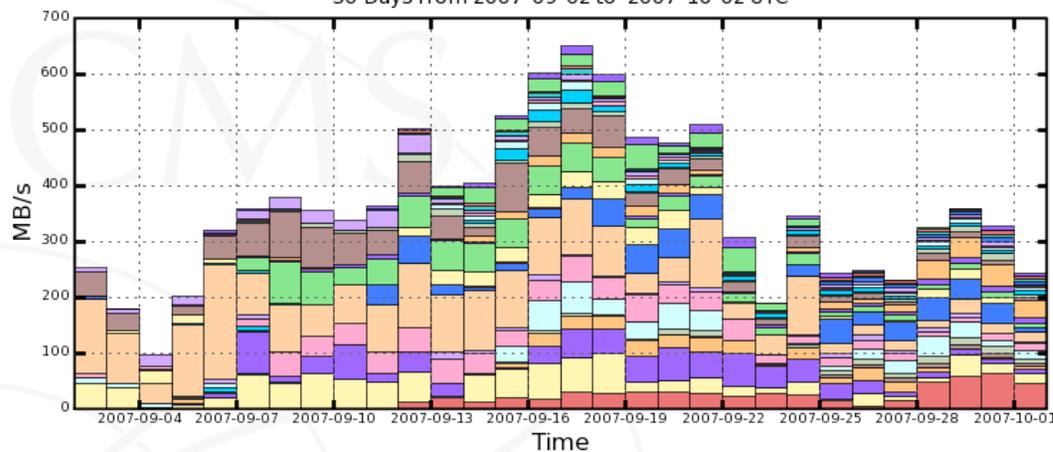
**CMS PhEDEx - Transfer Rate**

17 Weeks from 2007/36 to 2008/01 UTC



**CMS PhEDEx - Transfer Rate**

30 Days from 2007-09-02 to 2007-10-02 UTC



- T1\_ASGC\_Buffer
- T1\_PIC\_Buffer
- T2\_Belgium\_UCL
- T2\_Florida\_Buffer
- T2\_Nebraska\_Buffer
- T2\_Spain\_IFCA
- T3\_Minnesota\_Buffer
- T1\_CERN\_Buffer
- T1\_RAL\_Buffer
- T2\_CSCS\_Buffer
- T2\_HEPGRID\_UERJ
- T2\_Pisa\_Buffer
- T2\_Taiwan\_Buffer
- T3\_TTU\_Buffer
- T1\_CNAF\_Buffer
- T2\_Bari\_Buffer
- T2\_Caltech\_Buffer
- T2\_Legnano\_Buffer
- T2\_Purdue\_Buffer
- T2\_UCSD\_Buffer
- T3\_UCR\_Buffer

Maximum: 458.17 MB/s, Minimum: 0.02 MB/s, Average: 68.53

- T1\_ASGC\_Buffer
- T1\_PIC\_Disk
- T2\_Belgium\_UCL
- T2\_Florida\_Buffer
- T2\_Nebraska\_Buffer
- T2\_Spain\_CIEMAT
- T2\_Wisconsin\_Buffer
- T1\_CERN\_Buffer
- T1\_RAL\_Buffer
- T2\_CSCS\_Buffer
- T2\_HEPGRID\_UERJ
- T2\_Pisa\_Buffer
- T2\_Spain\_IFCA
- T3\_Vanderbilt\_Buffer
- T1\_CNAF\_Buffer
- T2\_Bari\_Buffer
- T2\_Caltech\_Buffer
- T2\_LIP\_Lisbon
- T2\_Purdue\_Buffer
- T2\_Taiwan\_Buffer
- T1\_FZK\_Buffer
- T2\_Beijing\_Buffer
- T2\_DESY\_Buffer
- T2\_Legnano\_Buffer
- T2\_RWTH\_Buffer
- T2\_UCSD\_Buffer
- T1\_IN2P3\_Buffer
- T2\_Belgium\_IHE
- T2\_Estonia\_Buffer
- T2\_MIT\_Buffer
- T2\_Rome\_Buffer
- T2\_Warsaw\_Buffer

Maximum: 649.25 MB/s, Minimum: 96.62 MB/s, Average: 360.07 MB/s, Current: 242.95 MB/s

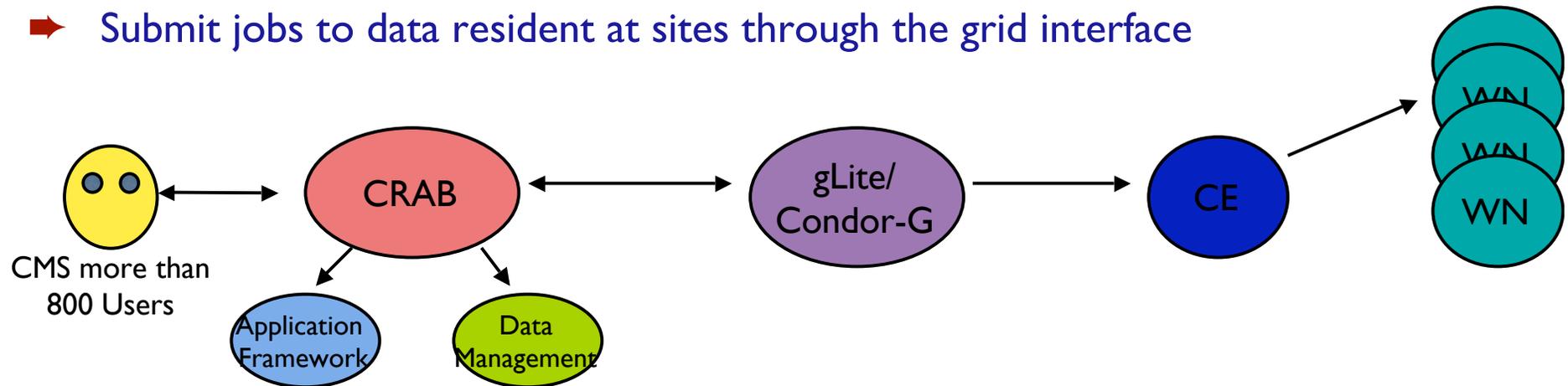
# Jobs to Data

While data can be moved to available resources, jobs can be submitted to remote analysis resources.

- ➔ CMS has developed services to provide transparency to the inherent structure of the computing system
- Expectation is the bulk of CMS analysis will be performed with CRAB
  - Nothing in the model prevents users from having interactive accounts at sites

CMS has the CMS Remote Analysis Builder (CRAB)

- ➔ Submit jobs to data resident at sites through the grid interface



Currently CMS is averaging

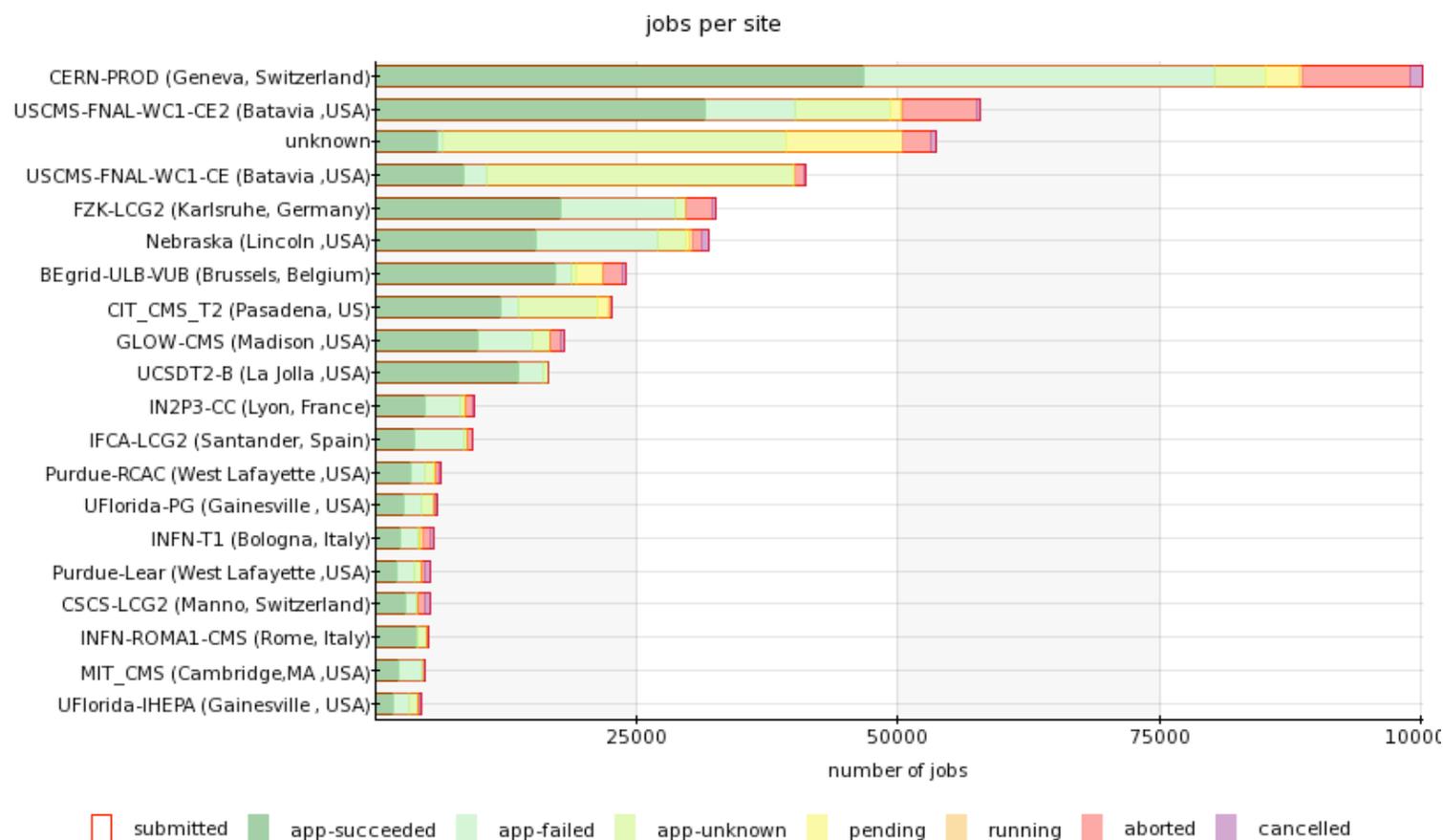
~15k-20k jobs per day

- ➔ Estimate in 08 is > 100k

# Analysis Submissions During Fall

Working on exporting more of the analysis submission away from CERN and the Tier-1 sites

➔ US Tier-2s are contributing well to the analysis resources

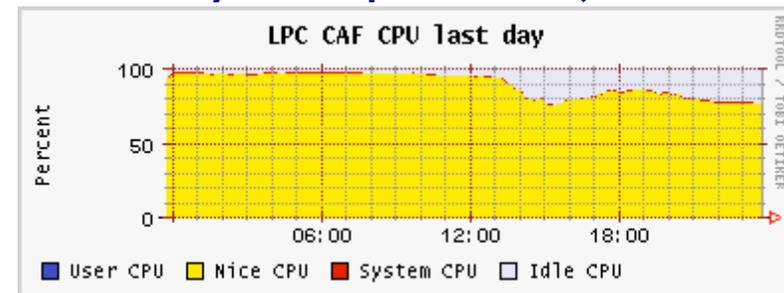




# LPC CAF

The LPC CAF has a lot in common with other Tier-2-like analysis centers

- ➔ The disk requirements were specified with the expectation that the analysis center had access to the data stored on the Tier-I disks
  - Data not expected to be served from FNAL has to be subscribed to the LPC dedicated storage and approved by a site data manager
- ➔ Unlike most of the Tier-2s the LPC CAF only accepts local job submissions
  - You need to log in and submit
  - Reasonably well utilized cluster



Accounts can be applied for using instructions on the web page

- ➔ <http://www.uscms.org/SoftwareComputing/UserComputing/GetAccount.html>



# Outlook

A lot of the basic functionality for managing data, transferring data and accessing data is in place

- Working on improving functionality, improving scaling and improving reliability
  - A lot of work to do as we wait for data
- Now is a good time to get practice working within the computing system
- CMS will be in the interesting position of commissioning the detector while we commission a distributed computing infrastructure