Experiment Status – Experience with SC4

Michael Ernst

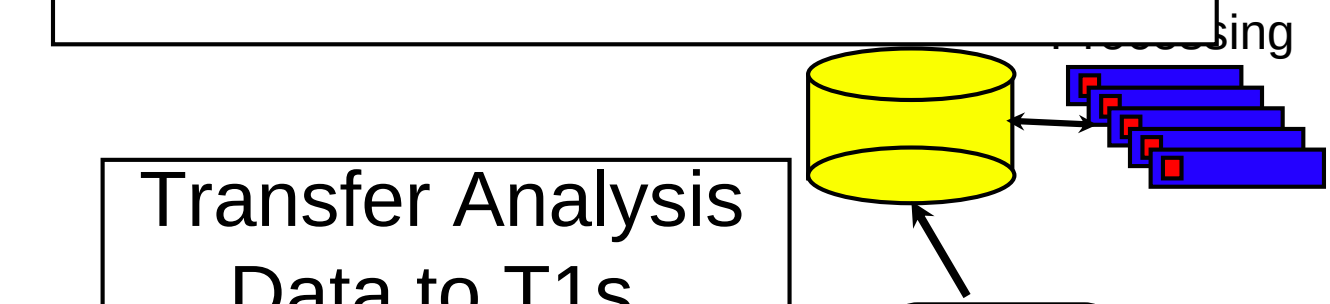
September 26, 2006

SC4 Workflows

Execute Analysis Jobs on Tier-1 Resources. Select Data

Transfer Analysis Data to T2s. Validate. Publish Existence Execute Analysis

Transfer Analysis Data to T1s. Validate. Publish Existence



Monitor Success, Rates, Failure Modes, Performance Predefined Metrics

Perform Simulation at T2s Transfer to T1 for Validation

Original Schedule Processing

Original schedule was to operate the first two weeks of June

- 25k jobs per day (50% analysis and 50% production)
  - Operate Job Robot on test simulation samples for analysis
  - Operate Prod_Agent for production job
- 90% success rate to complete jobs

We spent a lot of the first two weeks of June commissioning sites and did not start large scale operations until the middle of June

The original hope had been to be more dynamically moving data around and accessing it with analysis jobs

- In the end analysis was primarily on the same datasets
- Data was moved with MC simulation applications to CERN for CSA06

Site Preparation

CMS started with 5 tasks for the sites to complete to prepare for challenge activities

- Pass the grid site functional test (EGEE and OSG)
- Install CMS software
- Install and commission CMS Data Replication System (PhEDEx)
- Transfer a test sample and register it in the trivial file catalog namespace
- Accept an analysis application based on the CMS Remote Analysis Builder (CRAB) and successfully run it

All 7 Tier-1 sites eventually completed all the steps

24 of 26 Tier-2 sites that came forward completed the steps

- Took longer in both cases than we had budgeted
  - The steps lead to reasonably commissioned sites
- The commissioned sites certainly improved the speed of simulation ramp-up

Site representatives were very responsive

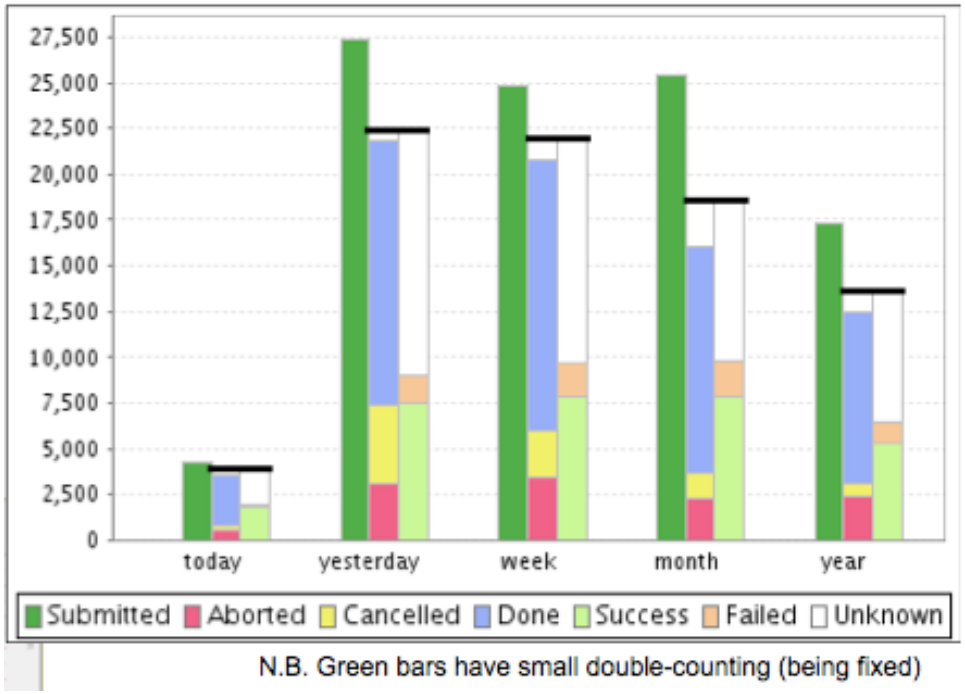
Meeting the Processing Metrics

CMS did pretty well at meeting the total number of job submissions

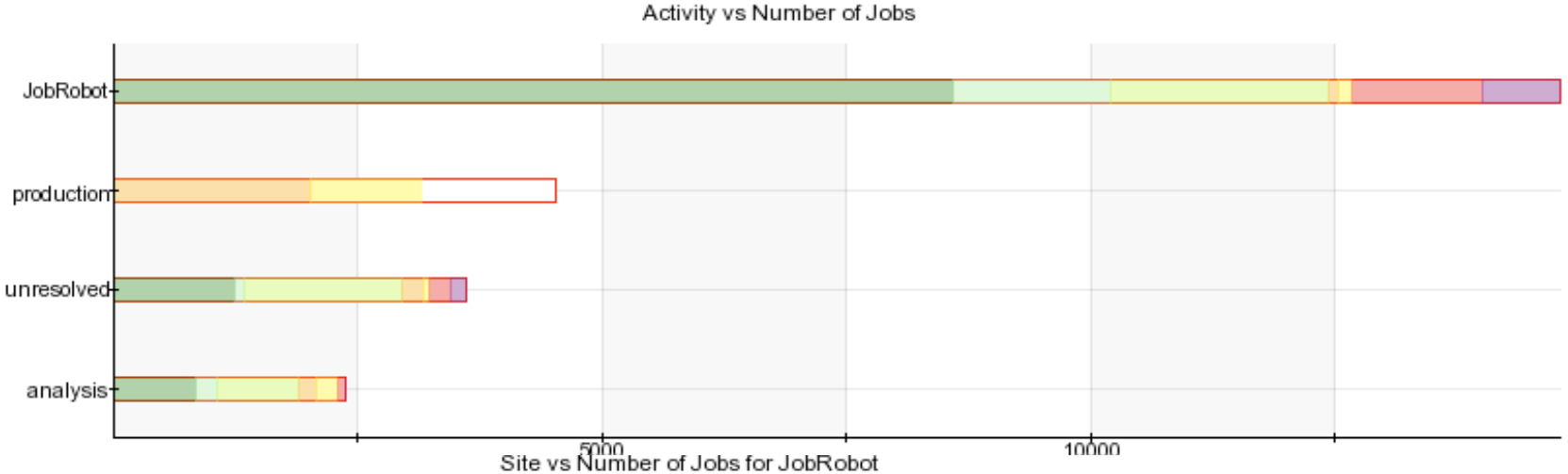
- 18k averaged over the last month and 22k averaged over the last week

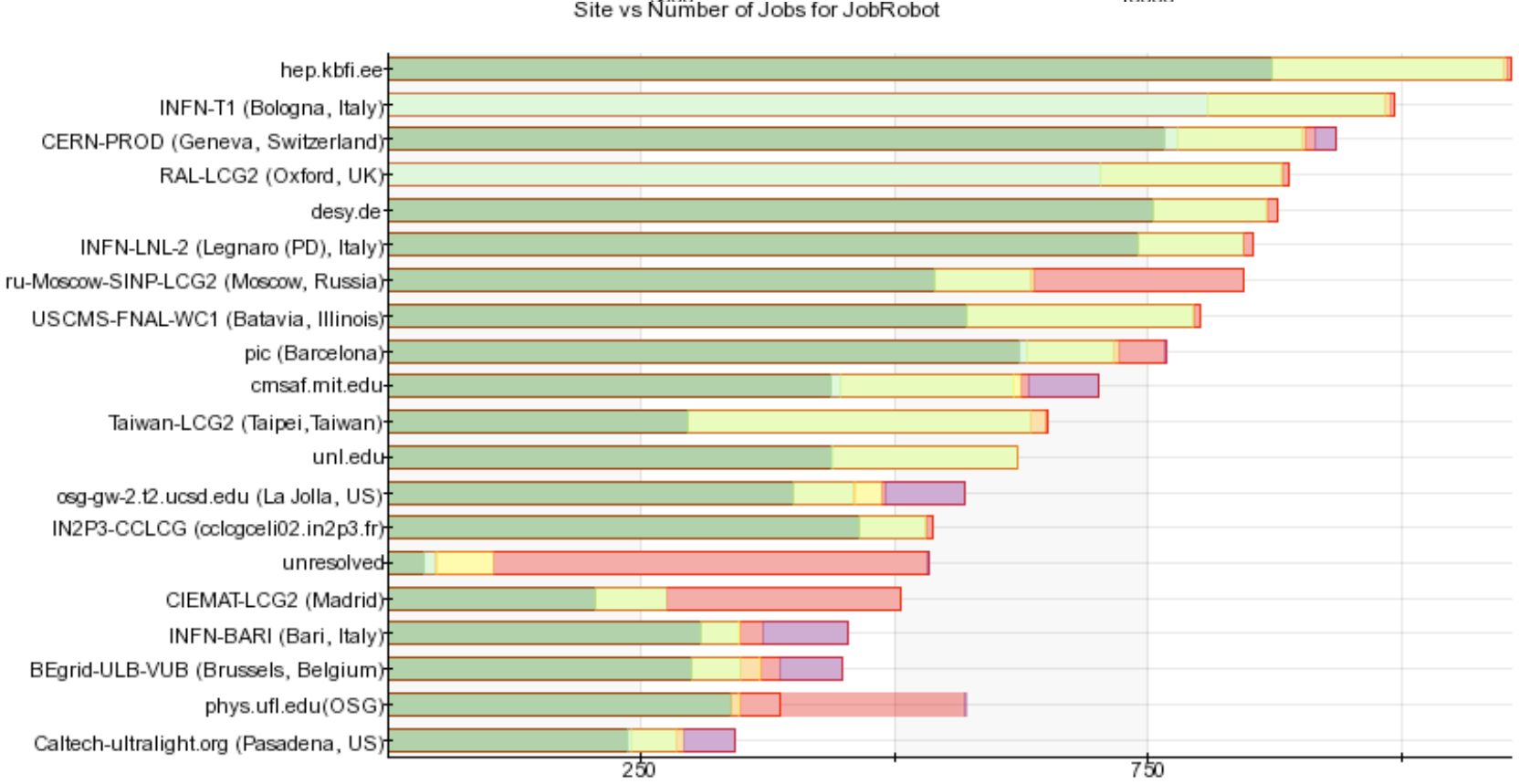
A few items of interest

- To reach the goal required care with the job submission
  - Special hardware was configured for the job robots
  - 4 RBs and re-worked robot
- Lots of attention from operators and sites
- We were at the limit of what the existing system can do



A typical day





Job Efficiency

We see issues where a few misconfigured worker nodes can significantly reduce the efficiency of a site

- Those nodes are preferentially available

We kill individual user jobs and need to resubmit them

- We can see 50% loss on sites with 1% badly configured nodes

This is an interesting use-case for pilot jobs

- Pilots determine the configuration and don't request work flows unless they are configured.
- We are finding even locally that we can have configuration issues that don't affect all VO's or only affect grid submissions
  - Diagnosis and debugging are challenging

Original Schedule Transfers

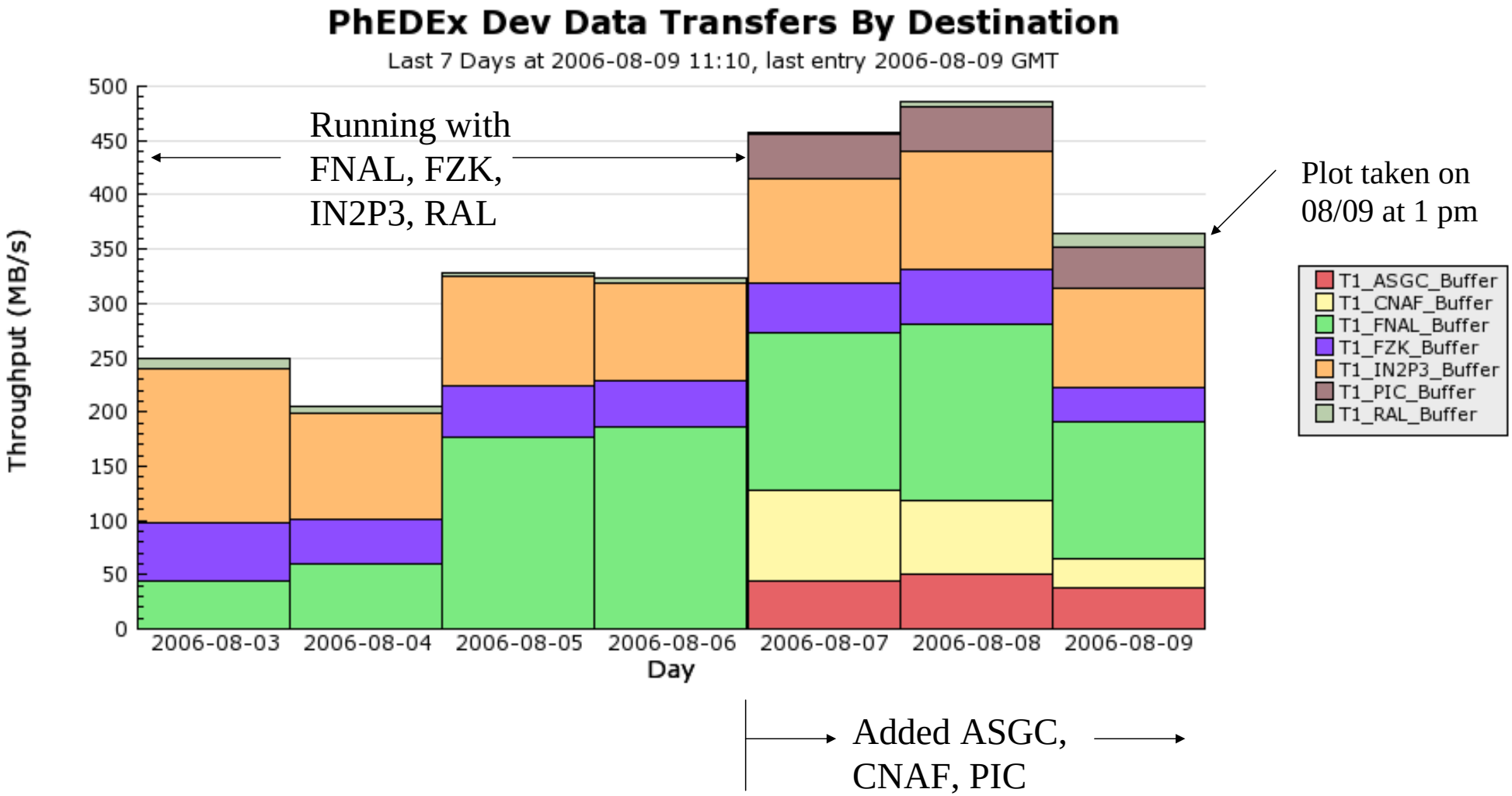
Scaling Tape Rates by pledge aiming for 150MB/s

- ASGC: 10MB/s to tape
- CNAF: 25MB/s to tape
- FNAL: 50MB/s to tape
- GridKa: 20MB/s to tape
- IN2P3: 25MB/s to tape
- PIC: 20MB/s to tape
- RAL: 10MB/s to tape

Networking provisioning should be at least twice this

- These goals are sufficiently modest that no center should struggle to sustain them

Summary of Tier-0 to Tier-1 (7 Days)

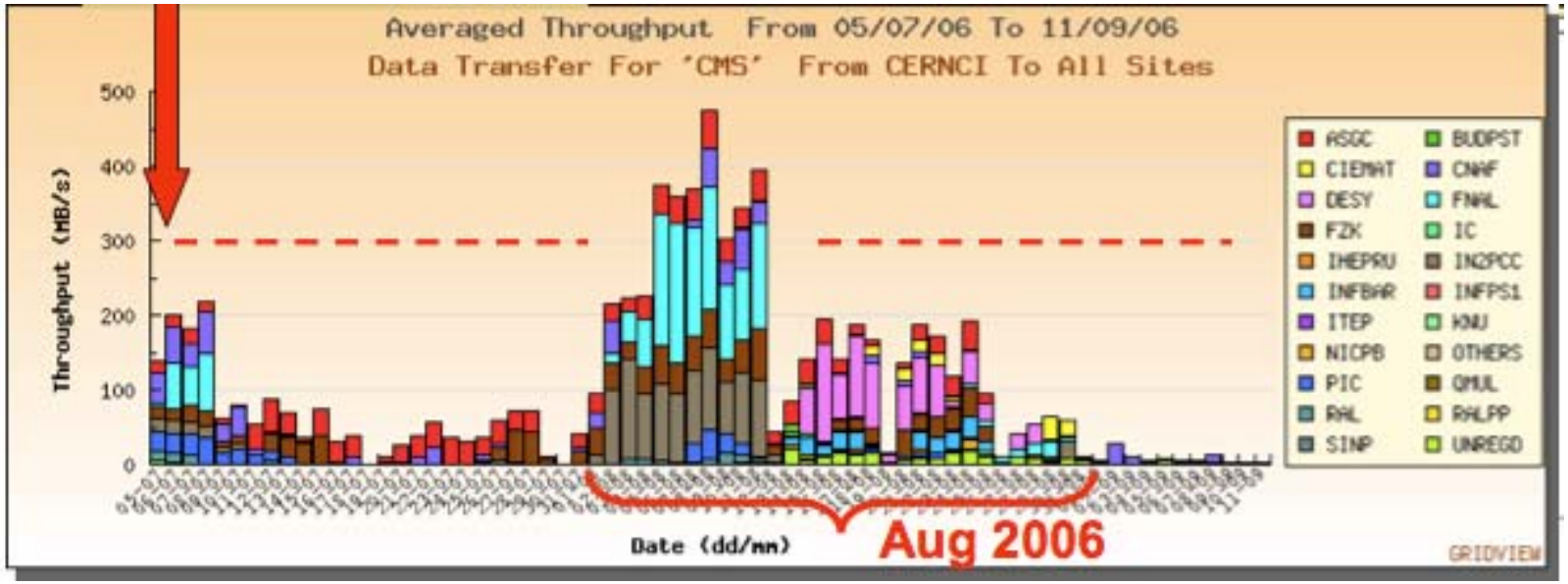


Tier-0 to Tier-1 Transfers

For the Tier-0 to Tier-1 transfers we were able to reach much higher rates than proposed but it required a very concentrated effort

- A throughput phase of LCG was conducted at Easter
  - The success of this test did not obviously translate into an easy turn on for CMS 2 months later
    - Good interaction with WLCG & CERN-IT and strong participation from CMS experts

300MB/s

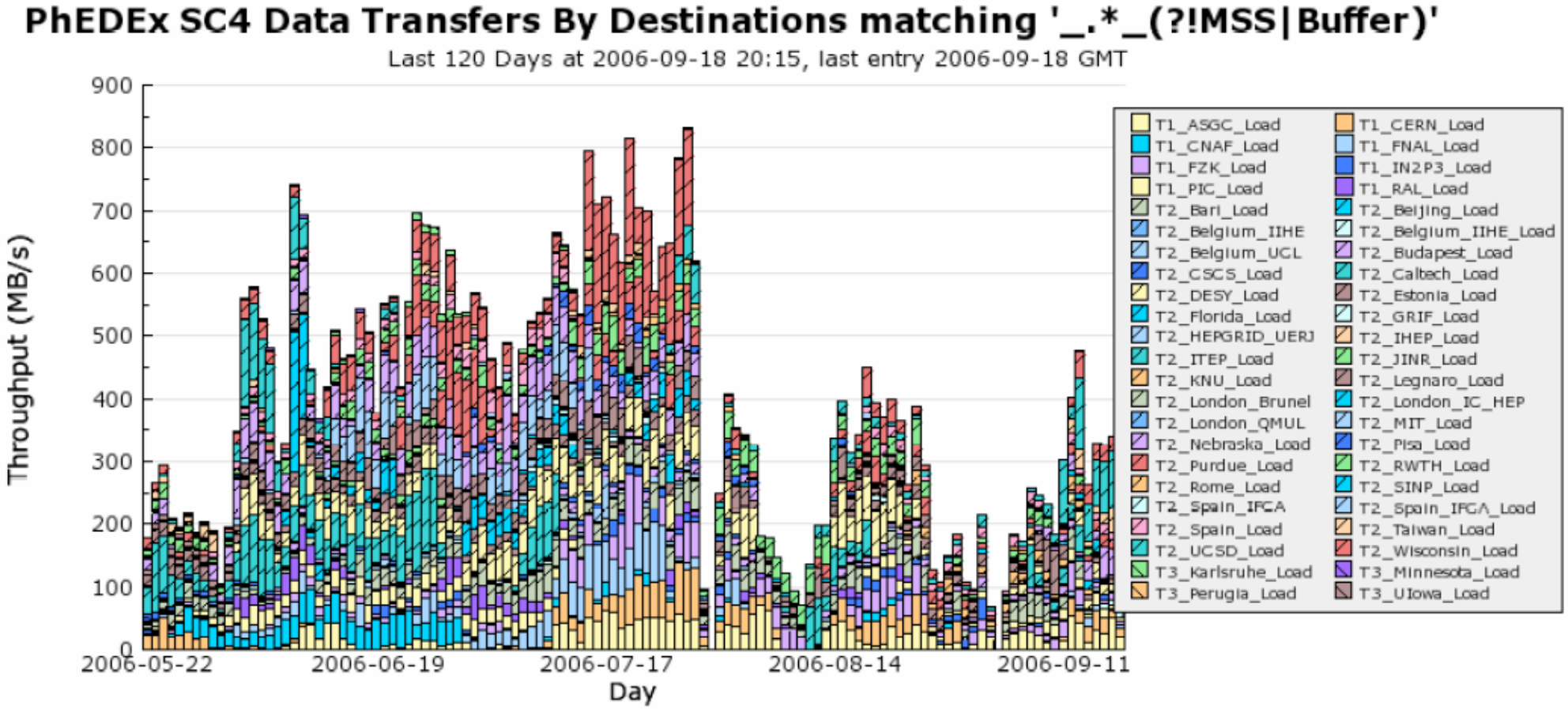


Tier-2 Transfers

Network Estimates for Tier-2 vary widely

- The computing model defines the expected minimum in 2008 at 1Gb/s
  - Naively taking 25% this would be 250Mb/s
- Given the number of Tier-2 centers already at 1Gb/s to 10Gb/s and the difficultly using reasonable scale networking end-to-end it makes sense to try much larger scale tests at some Tier-2 centers
  - Try to sustain ingest rate to Tier-1 centers from all Tier-2s
  - Drive Tier-1 to Tier-2 rates at 10MB/s to 100MB/s

SC4 Transfer Rate

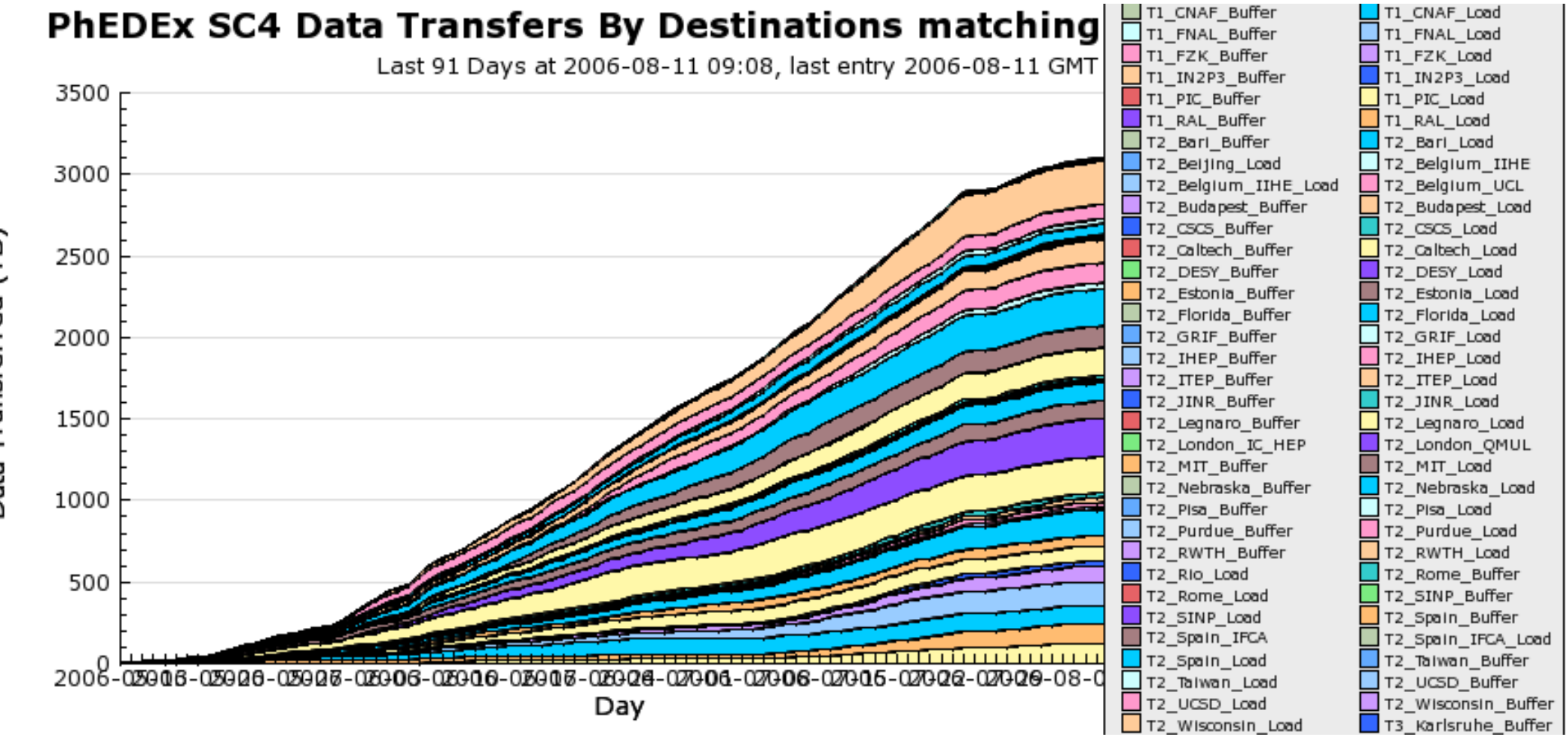


Daily average world-wide rates were regularly in excess of 500 MB/s

Transfer Destination

There are a lot of destinations that successfully received data at a reasonable rate

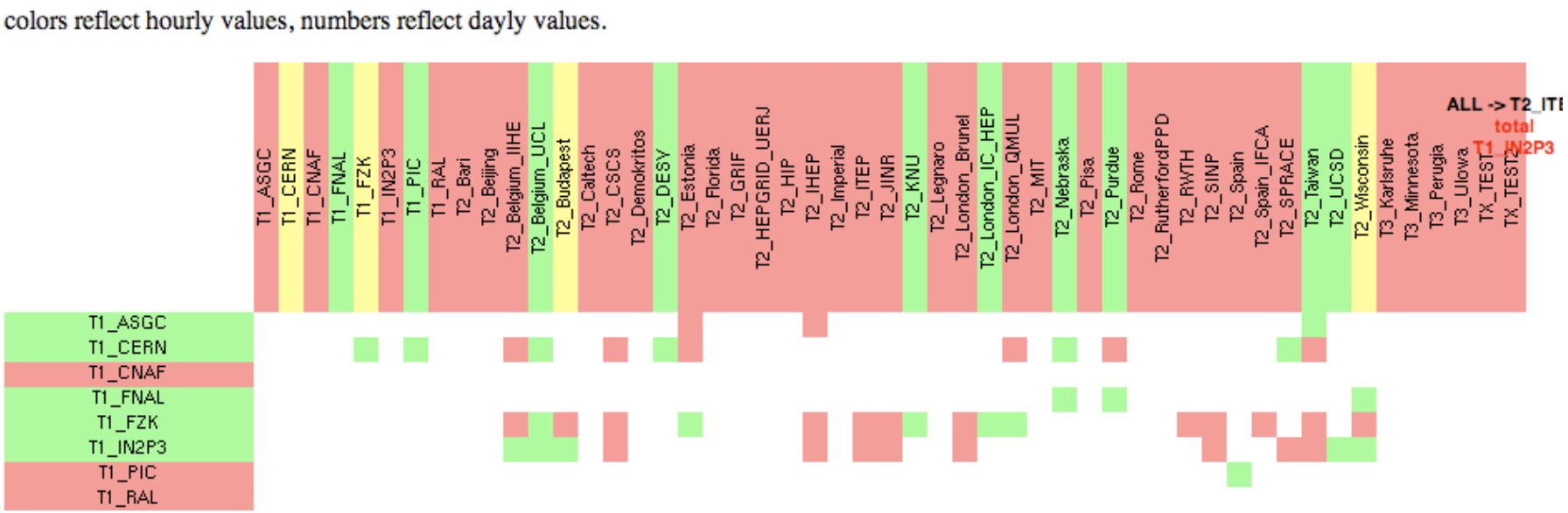
- The majority of the data came from one Tier-1 which was working well for the middle of the summer



Transfers between Tier-1s and Tier-2s

The transfers between Tier-1 and Tier-2 centers are an area we are potentially exposed for CSA06

- Trying to recover with preparation activities now
- We concentrated effort on success of Tier-0 to Tier-1 transfers and later on file uploads from production sites
  - There are a lot of elements of the matrix below and there are a lot of non-functional links at the moment
    - The Load test and this dashboard have been extremely useful tools



Summary on Data Transfers

Most of the technical metrics and demonstration of specific activities we had for SC4 were met

- Major progress made during last four months
  - Much of the infrastructure we use has changed
  - Significant progress in central and world-wide operations
    - Now better equipped to support multiple concurrent activities
  - Successfully addressed four top concerns CMS had in May
    - Data Transfer, Workload Management, Data Storage & Data Access at CERN, Operation of CMS Services
    - Taken coordinated immediate targeted actions
- Integration remains the top relevant issue
  - Operating multiple concurrent activities
  - Hiding boundaries of the computing components from users
  - Operation and support of a complex stack
    - From database server to middleware to networks to storage systems

Estimated Performance of Mass Storage

The expected mass storage performance

- At a nominal Tier-1 is 800MB/s to the worker nodes
- At a nominal Tier-2 is 200MB/s

We have been aiming for 1MB/s per batch slot

- 300MB/s at a nominal Tier-1
- 100MB/s at a nominal Tier-2

These are being exercised and documented in CSA06

- For Castor and dCache the performance has be good
- DPM is being exercised in SC4
- Common RFIO for DPM and Castor in preparation by WLCG & CERN-IT

We had only about 1/3 of the Tier-2 centers formally document they met these milestones

- In part because the current job robot jobs are not extremely data intensive

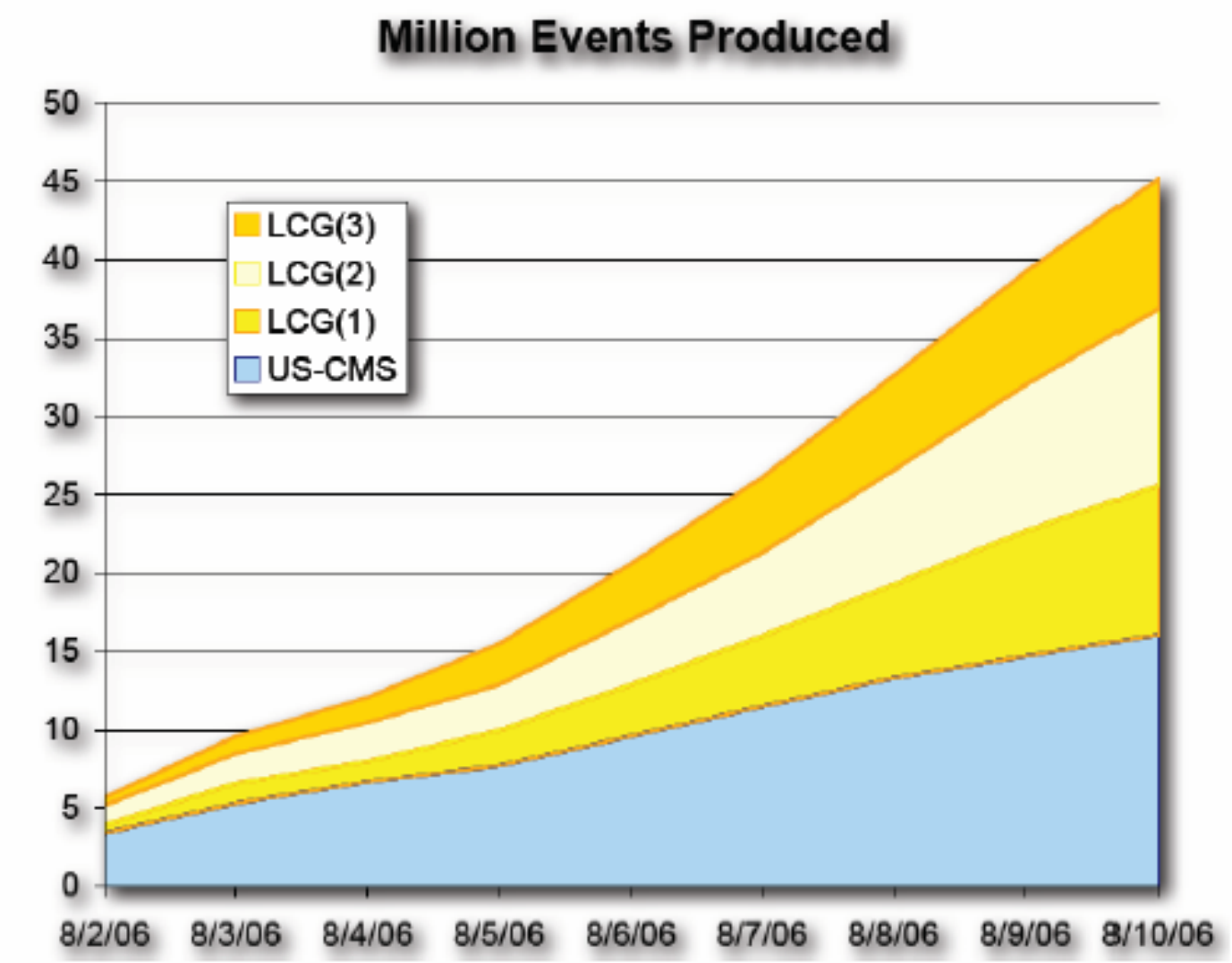
Production

From the standpoint of SC4

- The simulation jobs were able to run in parallel with the job robot analysis jobs
  - Basic implementations of experiment priorities were implemented at 1/3 of the sites
- The production teams did very well, the Prod_Agent functioned very well and the sites remained responsive and very efficient

The result was beyond expectations

Production Ramp-up (Minbias Sample)



General Comments

Most of the technical metrics and demonstration of specific activities we had for SC4 were met

- An elements of general concern
  - The ramp to full scale operations was long. Even in the end we did not execute every element simultaneously
    - This makes discovering interference effects difficult between components and the long ramp demonstrates the difficulty associated with some of the tasks

Next Steps

Experiment activities, operating in parallel will drive the activities

- This is good for CMS
- We need to ensure we are performing meaningful activities to increase the scale and functionality of the computing systems
  - CSA06 gets us into the fall
  - Concentrations after CSA06 will be guided by what we learn during the data challenge

Operational Goals

CMS needs to be at production scale services in 2008

- Assuming we cannot easily more than double the scale each year, we should be able to demonstrate 25% of the expected 2008 scale in this year and be able to reach 50% scale early in 2007

Service	2008 Goal	2006 Goal	%
Network Transfers between T0-T1	600MB/s	150MB/s	25%
Network Transfers between T1-T2	50-500 MB/s	10-100 MB/s	20%
Job Submission to Tier-1s	50k jobs/d	12k jobs/d	25%
Job Submissions to Tier-2s	150k jobs/d	40k jobs/d	25%
MC Simulation	1.5 10^9 events/year	25M per month	25%

- Network transfers between T0-T1 centers
  - 2008 scale is roughly 600MB/s
- Network transfers between T1-T2 centers
  - 2008 Peak rates from Tier-1 to Tier-2 of 50-500MB/s
- Selection Submissions and Transfers to Tier-1 centers
  - 2008 submission rate 50k jobs per day to integrated Tier-1 centers
- Analysis Submissions to Tier-2 Centers centers
  - 2008 Submission rate 150k jobs to integrated Tier-2 centers
- MC Production jobs at Tier-2 centers
  - 2008 rate is 1.5 x10^9 Events per year