



# FNAL Experiences from February Tests

April 22, 2008



# February Tests

From the site standpoint CMS exercised a variety of tests during February

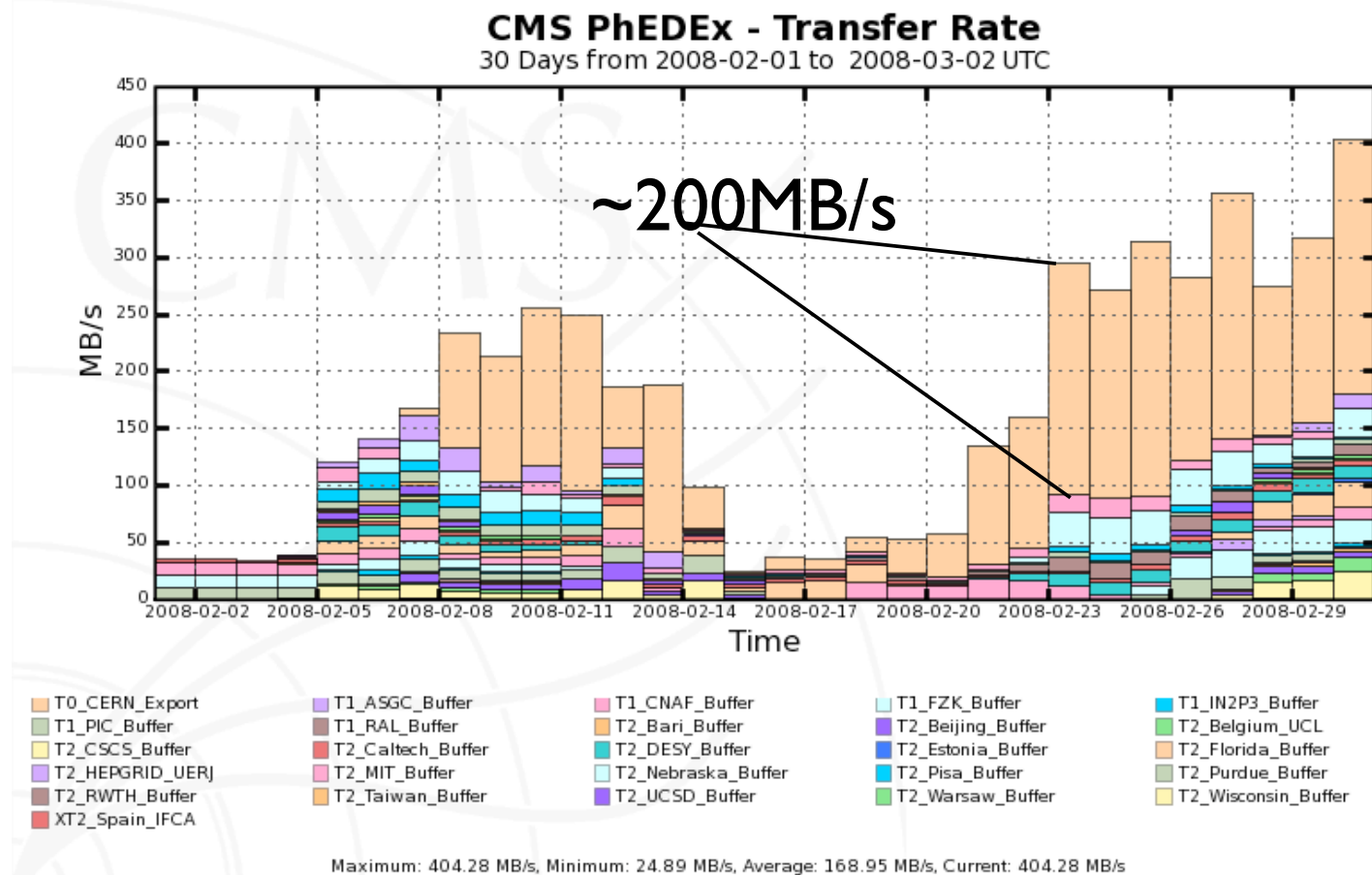
- ➔ Transfers
  - Download from CERN at 40% of the peak rate expected in 2008
  - Transfers data from T1 to T1 at the rate expected for T1 synchronization
  - Transfer to regional Tier-2s
- ➔ Restoration from Tape
  - Demonstrate the ability to recover 10TB
- ➔ Data Processing
  - Reprocess the data staged from tape



# File Import to FNAL

Data rate expected from CERN was 149MB/s

➔ Achieved on enough days to satisfy the metric

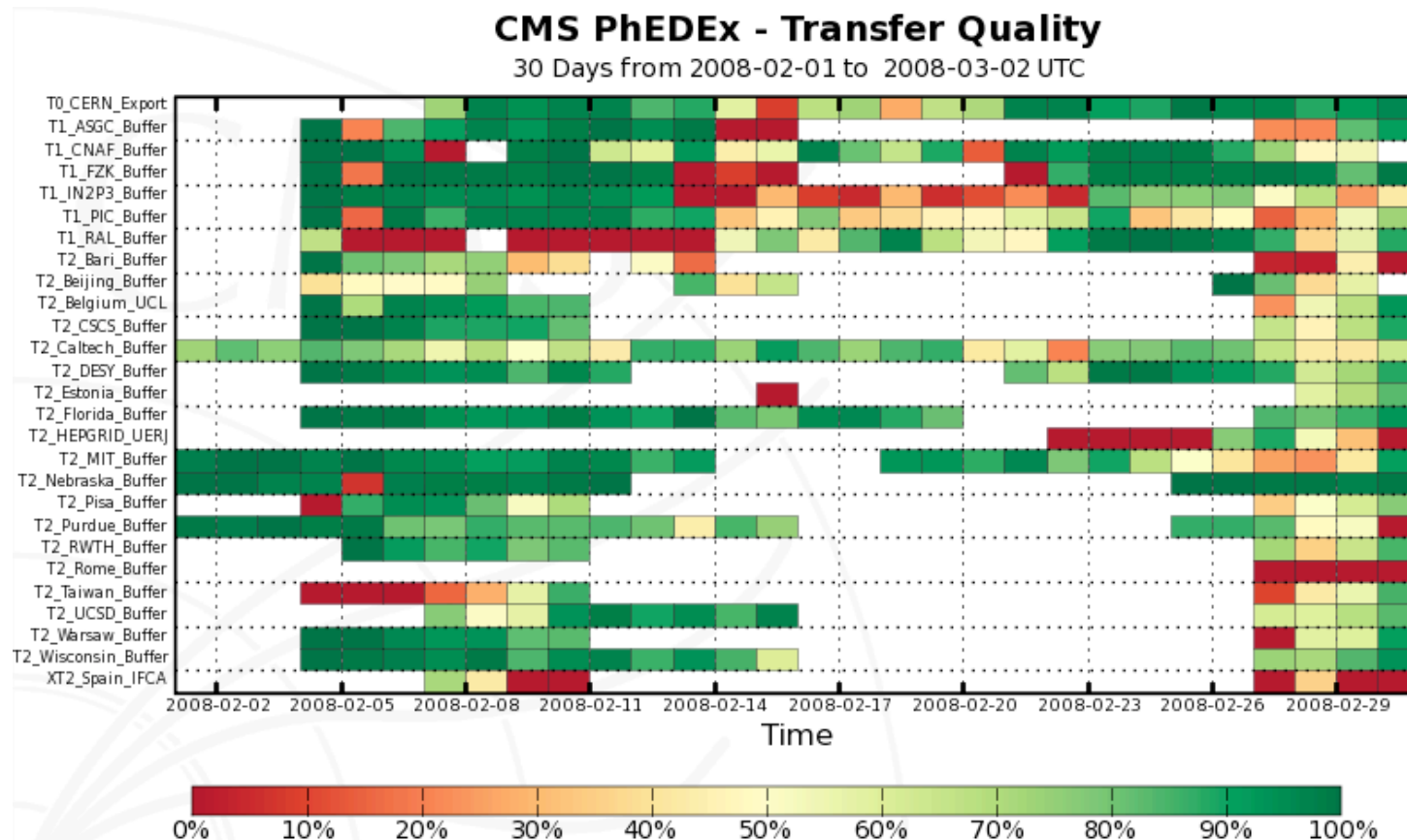




# Transfer Quality

Looking at the transfers quality (the number of retries needed to complete the transfer)

- ➔ There are not obvious site problems during the dip in the middle

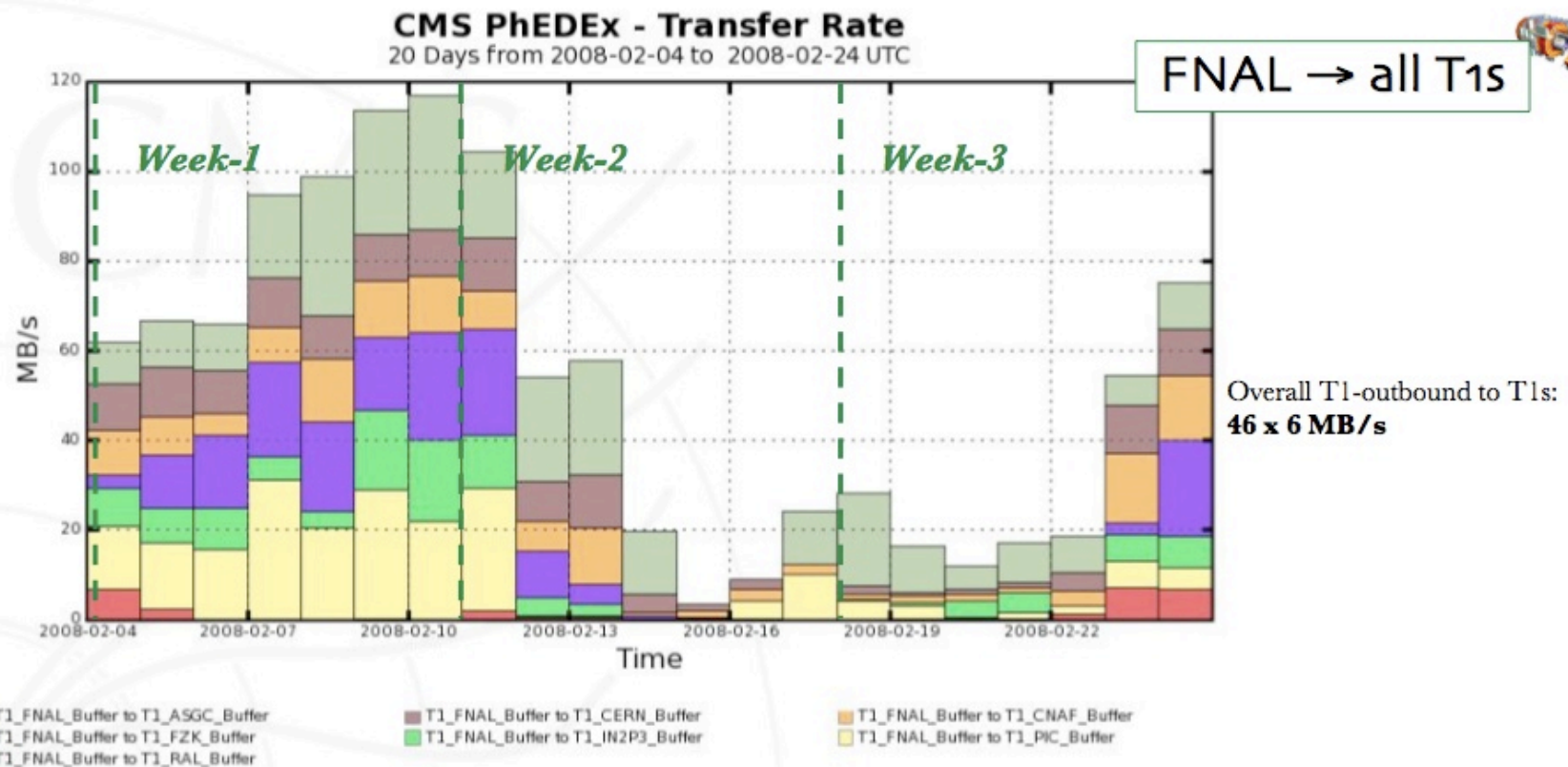




# Transfers T1 to T1

CMS T1 to T1 estimates are based on the ability to synchronize the AOD data in a limited period of time (2 weeks)

- ➔ Since FNAL has a lot of raw data for custodial storage it means there is a lot of AOD data to export 46MB/s x 6 sites
- ➔ Not achieved



Maximum: 116.75 MB/s, Minimum: 3.61 MB/s, Average: 52.99 MB/s, Current: 75.25 MB/s



# Issues with T I to T I Transfers

## Still working to demonstrate the full rate

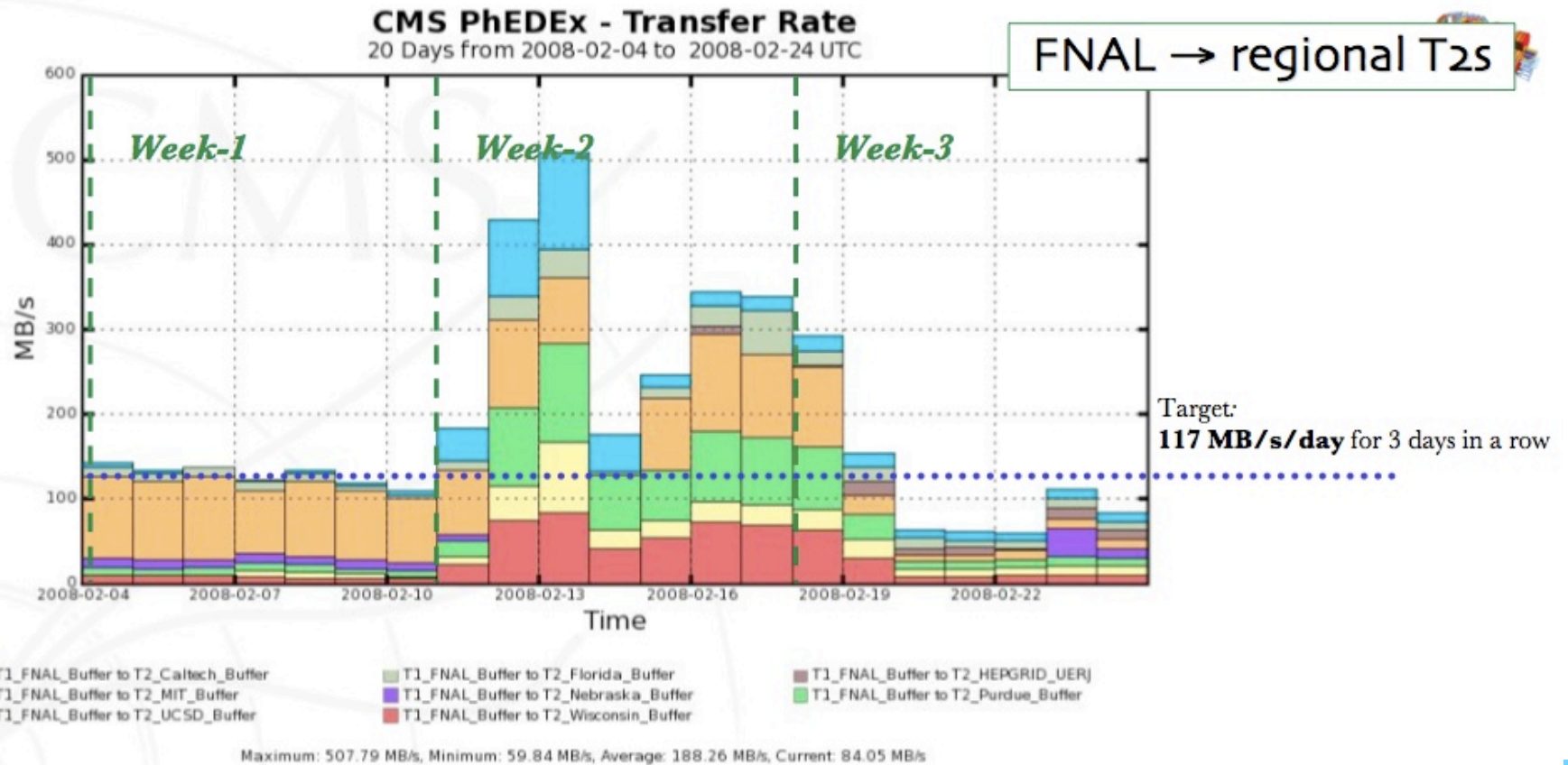
- ➔ The FNAL rate of around 250MB/s on export has been demonstrated frequently as a total but not to the 6 T I simultaneously
- ➔ The 46MB/s export rate from FNAL is comparable to the CERN to T I rate for some links
  - For periods of data synchronization during data taking it will be a higher load
- ➔ CMS the experiment needs to assess the priority of the high T I to T I synchronization rate.
  - It should be achievable, but it's not clear that it's extremely well motivated
    - May be more worthwhile to concentrate on commissioning other sections



# Transfers Tier-1 to Tier-2

FNAL transfers to regional Tier-2s was successful

➔ This is an area we have concentrated on





# Tape Restores

CMS devised a tape restore test where 10TB of data were to be restored and then reprocessed and the time measured

- ➔ FNAL wished to demonstrate that a concentrated pre-staging of data would not be necessary for only 10TB
- ➔ Basically restore the files as needed while other production processing requests were in the queue
- ➔ Reasonably successful
  - The numbers from FNAL look reasonably similar to Tier-I sites that prestaged the data
    - Roughly 24 hours to complete the exercise
    - Rate from tape was ~110MB/s
    - Processors were also performing other organized processing requests





# Some recent mass storage lessons

## Arranging the layout of files onto pools is hard

- ➔ Initially we just used the dCache cost functions
  - These are not updated rapidly and pools tend to be selected for periods of time
    - If all the files for a time period are written to a pool when production begins to run 3000 batch slots can make requests of a single or small group of pools
- ➔ We switched to random assignment to pools
  - Random gives a reasonably flat access profile, but has too problems
    - As a pool files up files are kicked out of cache. If those files are needed again, dCache stages them back. It's a random collection of things that need to be restored
      - Very hard on the robot
    - If a pool goes offline, a random collection of files need to be restored
      - Very hard on the robot
- ➔ We're looking at things between random and persistently one pool



# Processing

During the staging and re-processing activity CMS made use of about 1200 slots

- ➔ We have demonstrated up the nearly the full farm of 3000 batch slots for other challenge activities
  - Not especially challenging

## A couple of processing lessons

- ➔ Currently we are running 4 CEs
  - Seems to be scaling reasonably well
- ➔ We currently have one SE endpoint
  - CMS applications initially staged out all files through the SE interface
    - Some applications have multiple outputs
    - Can easily drive the system beyond the scaling of the SE
      - For the moment we have been modifying the production workflows to stage out with a lighter less authenticated dccp stage-out
        - Fewer srm-get-meta data calls



# Outlook

Not everything went as smoothly as we would have liked during the February tests

- ➔ Transfers from CERN were slower to ramp up
- ➔ T1 to T1 synchronization isn't at scale yet
- ➔ While transfers to associated T2s look OK we've got work to do on the other T2s

A lot of metrics were met

- ➔ Looking forward to May tests.