US CMS Tier-2 Storage Experience

Ken Bloom

WLCG DM BOF

September 1, 2007

US CMS T2 storage

All seven US CMS T2 sites use dCache for our storage management, and it pretty much works.

Sites generally use the release that comes through VDT, along with whatever patches are needed.

- Currently on dCache v1.7, no upgrade of major version planned for a few months yet
- But patches keep coming; most sites are at #34, which has fewer problems than previous versions

Takes about 1 FTE to operate and maintain; not a trivial system.

Probably the most complicated software at US CMS T2, but it works!

- Demonstrated sustained 200 MB/s from storage to applications
- Can handle large disk pools
- Can handle large WAN transfers

dCache manages a lot of disk

We are able to manage disk pools of significant size:

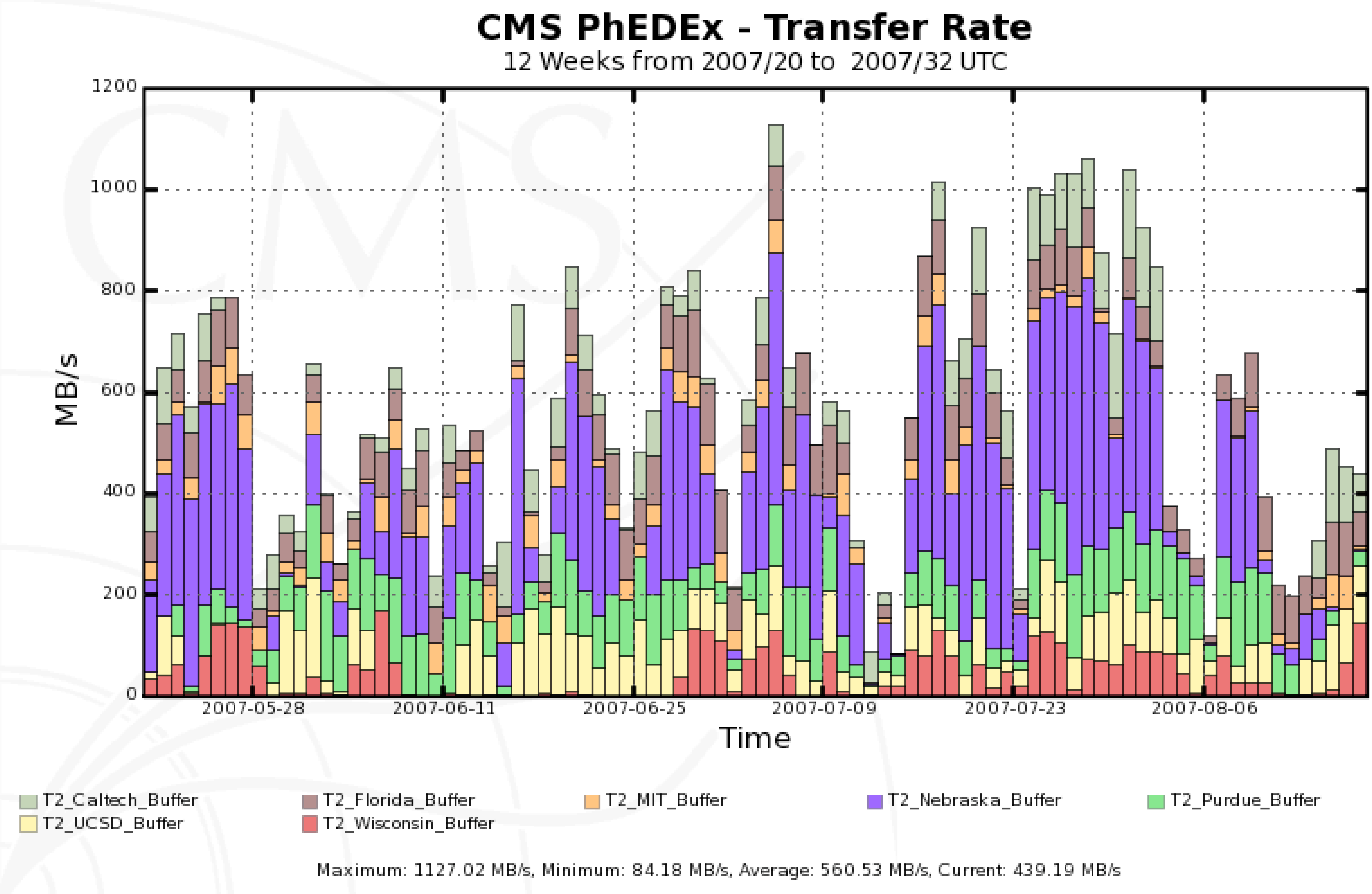
Site	CPU (kSI2K)	Disk (TB)	WAN (Gb/s)
Caltech	586	60	10
Florida	519	104	10
MIT	474	157	1
Nebraska	650	105	10
Purdue	743	184	10
UCSD	318	98	10
Wisconsin	547	110	10

This is "raw" disk -- actual usable amount smaller after RAID, "resiliency" for keeping multiple copies of the same file, etc.

All sites will be at >200 TB by mid-2008; no scaling problems anticipated (or none that I know of).

Network Performance

Systems can handle significant inflows of data for extended periods:



Not just about network speed, but about ability to put the bits somewhere!

Why are we successful?

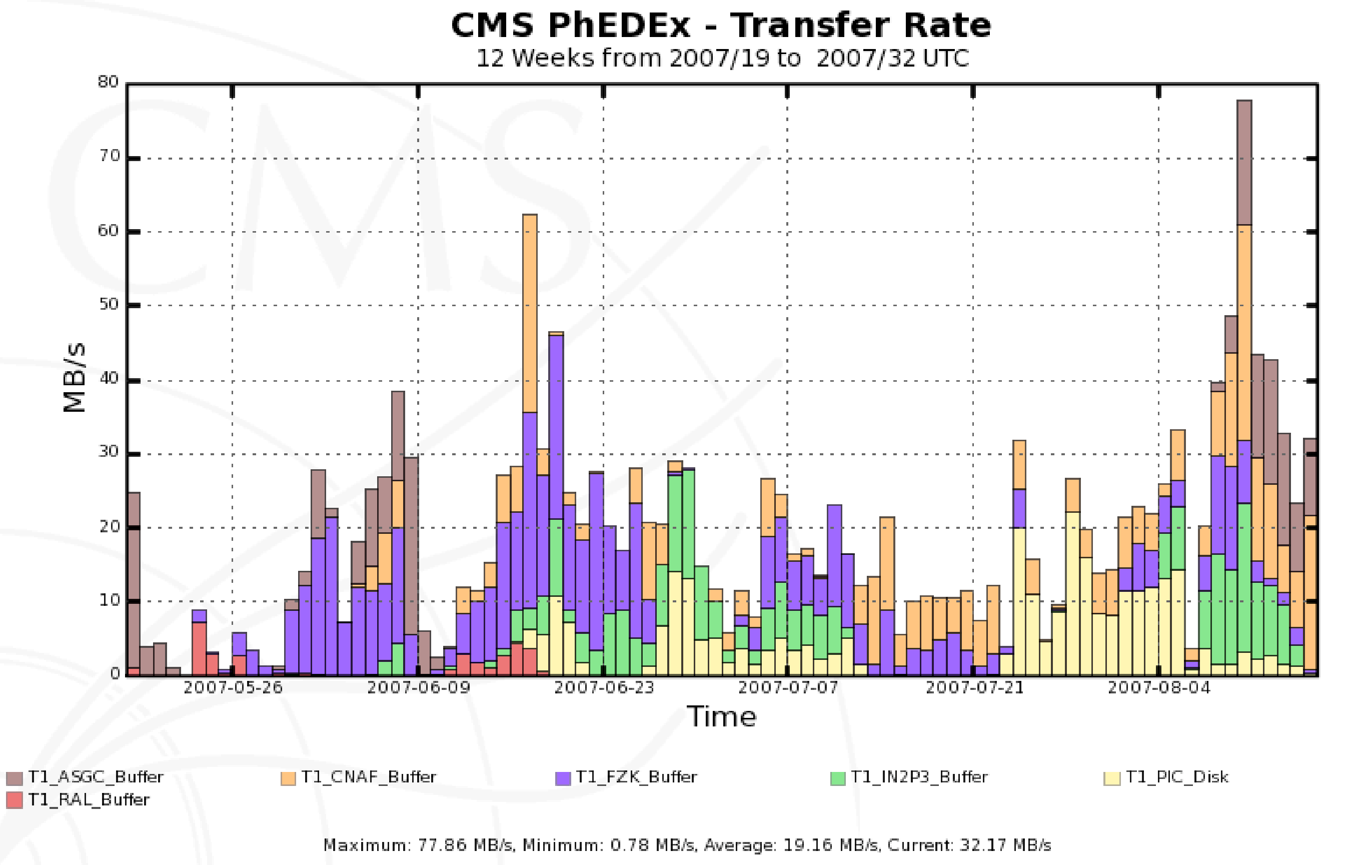
Some of it is technical -- flexibility in configuration, etc.

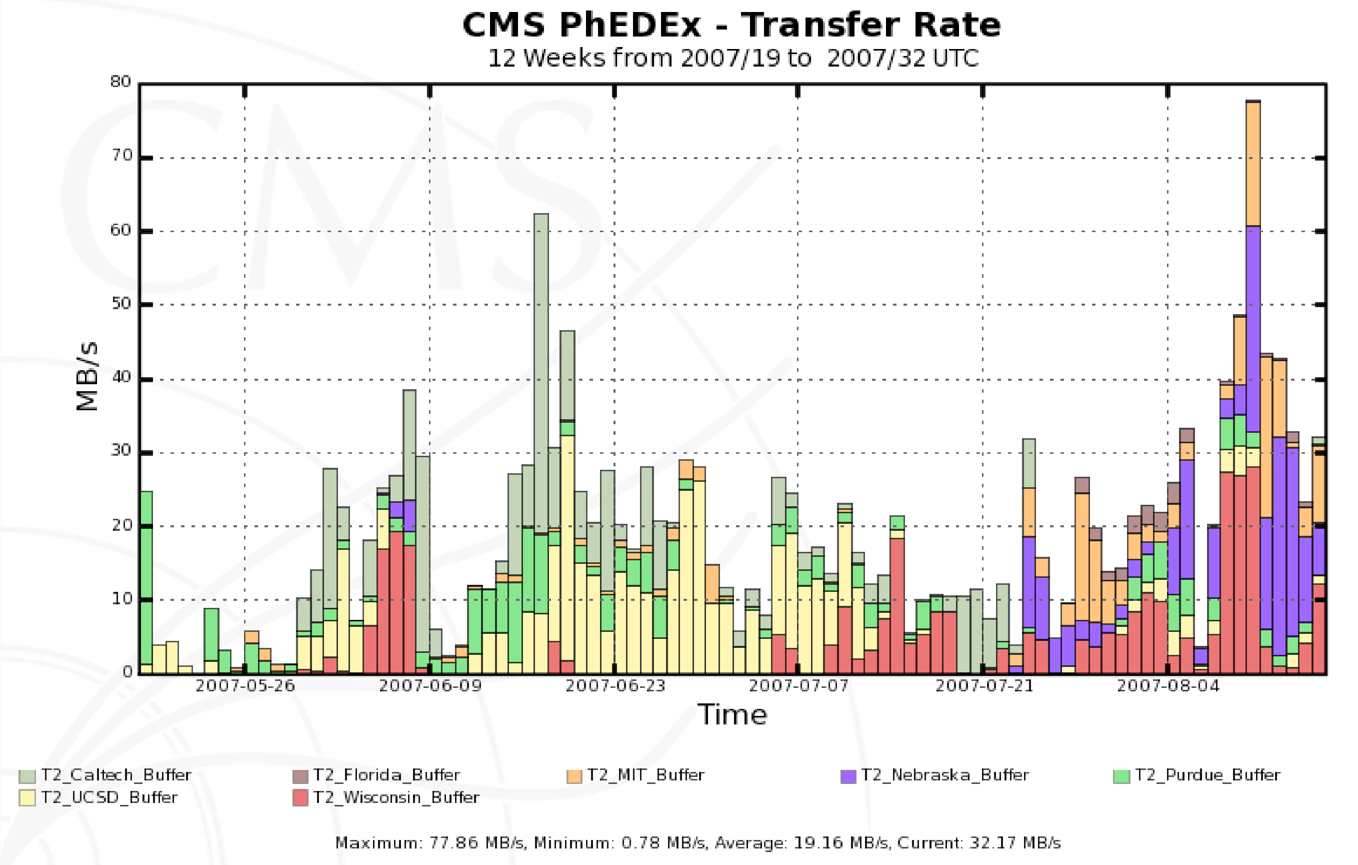
But a lot of it is social:

- ➨ Uniformity of architecture throughout US CMS sites allow us to share experience, utilities, etc.
- ➨ Biweekly US CMS T2 meetings allow us to identify common issues and find common solutions.
- ➨ We have good relations with the FNAL-based dCache developers
  - Great support from FNAL!
  - And give them (i.e. OSG storage group) useful feedback
    - E.g., problems with ReplicaManager have been worked out in that context.

In contrast...

Those great transfer rates are dominated by transfers from FNAL. From other places, the story is not as good:





Despite recent concerted efforts, have found it difficult to maintain even a few MB/s from non-US T1 sites -- see James Letts's talk tomorrow.

We believe that it's partly due to storage issues elsewhere (Castor works, but less reliable?), but mostly due to social issues -- need concerted efforts at the T1's on CMS specific demands and problems.

Our #1 concern in CMS T2 computing at the moment!

Summary

While the flexibility and complexity of dCache is a challenge, the US CMS T2 sites have had a lot of successes with it:

- ➨ Good performance for data serving and data transfers
- ➨ Good support and interaction with developers, and across the T2 and OSG communities
- ➨ We are hoping for more improvements yet, but we are confident enough in the product

What do we have to do to improve our interaction with other storage systems on the WLCG?

- ➨ US sites are eager to help as much as possible.