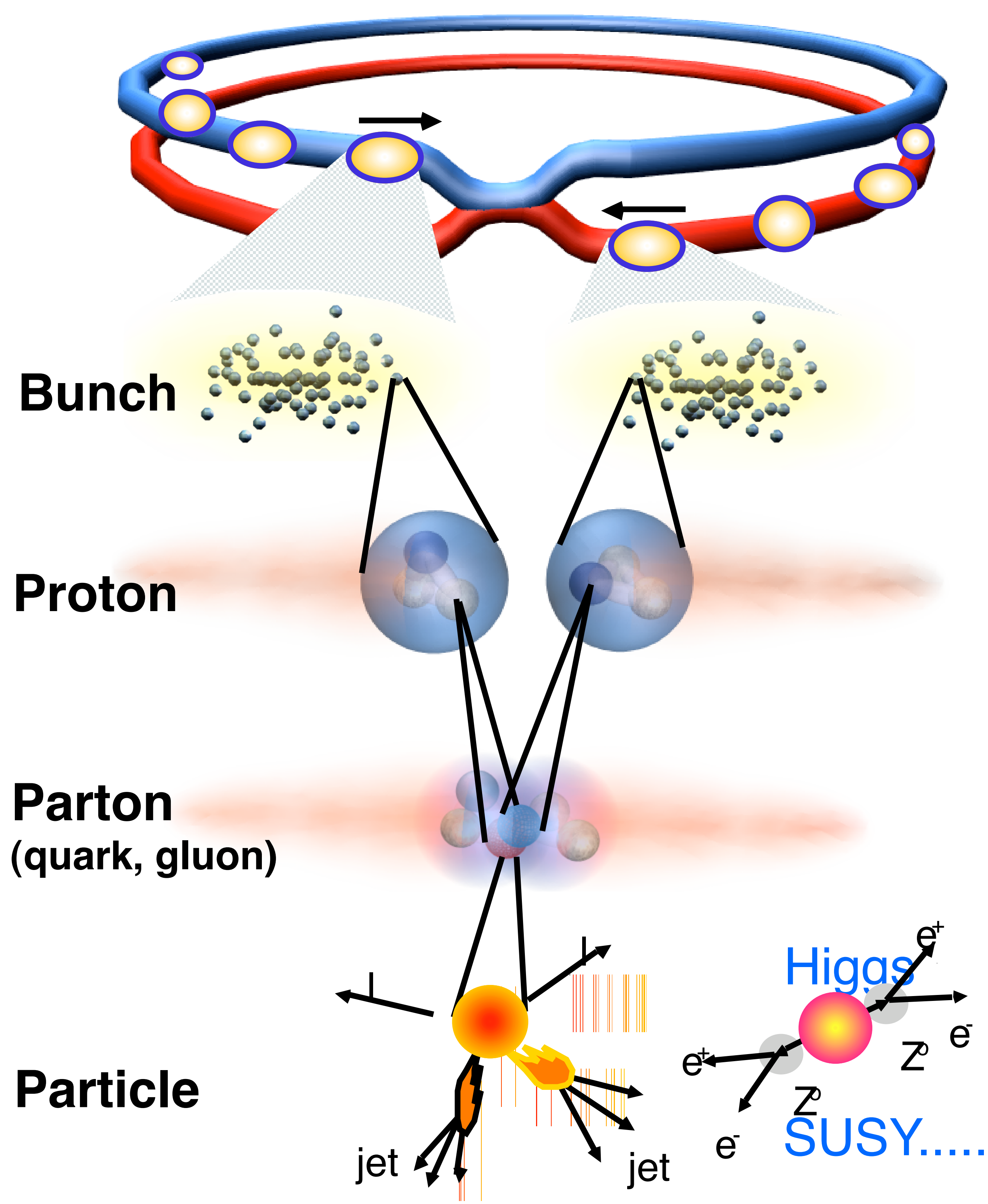
The CMS Computing System: Successes and Challenges

Ken Bloom
University of Nebraska-Lincoln
for the CMS Collaboration
July 27, 2009
Thanks to: Ian Fisk, Oliver Gutsche, Frank Wuerthwein

The problem



Proton- Proton	**2835 bunch/beam**
Protons/bunch	**10^{11}**
Beam energy	**7 TeV (7×10^{12} eV)**
Luminosity	**10^{34} cm^{-2}s^{-1}**
Crossing rate	**40 MHz**
Collision rate	**~10^9 Hz**

New physics rate ~ 0.00001 Hz

Event Selection:
1 in 10,000,000,000,000

Setting the scale

Currently expecting a long LHC run, 6M seconds running time in 2009-10

CMS will record data at 300 Hz

- Total 2.2B events, once dataset overlaps accounted for

Event sizes

- 1.5 MB for raw detector data
- 2.0 MB for simulated raw data

Event generation/reconstruction times

- 100 HS06-sec/event for data
- 1000 HS06-sec/event for simulation

Plug it all into CMS computing model

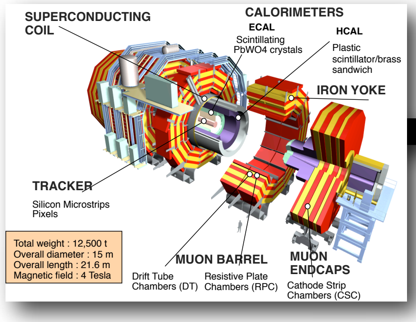
- 400 kHS06 CPU
- 30 PB disk
- 38 PB tape

to handle all CMS computing needs (production, calibration, analysis, etc.)

CMS Distributed Computing Model

CMS has been developing a distributed computing model from early in the experiment

- Variety of motivating factors (infrastructure, funding, leverage)
- Many challenges in making the distributed model work



Tier-0

Prompt Reconstruction,
Calibration Streams
Archival Copy of RAW and
First RECO data

Tier-1

Archival and Served Copy of RECO
Re-Reconstruction, Skimming
Archival Storage for Simulation

Tier-2

Simulated event production and primary resource for analysis by physicists

Getting ready for a new world

While the scale of Tevatron Run II computing is impressive, CMS computing will be something still different:

- Not enough resources at any single location to perform all analysis
  - cf. CDF, FNAL has ~equal resources for reconstruction and analysis
- In fact, CMS computing depends on large-scale dataset distribution!
- All reprocessing resources will be remote
  - cf. D0, much reprocessing off site, but after other elements commissioned
- Commissioning of distributed computing model will be simultaneous with detector commissioning, not to mention search for new physics

Need to take all steps possible to be ready before colliding beams!

STEP 09

STEP = Scale Testing of the Experimental Program

A multi-VO exercise in the context of WLCG -- make sure that all experiments can operate simultaneously, esp. on shared sites. All VO's agreed to do tests in the first two weeks of June.

For CMS, *not* an integrated challenge!

- This way, downstream parts of the system can be tested independently of the performance of upstream pieces
  - Also much less labor intensive....
- Did not want to interfere with other preparations for data-taking

Focus on pieces that needed greatest testing, and had much VO overlap:

- Data transfers: T0→T1, T1→T1 and T1→T2
- T0: recording data to tape
- T1: processing and pre-staging
- T2: use of analysis resources

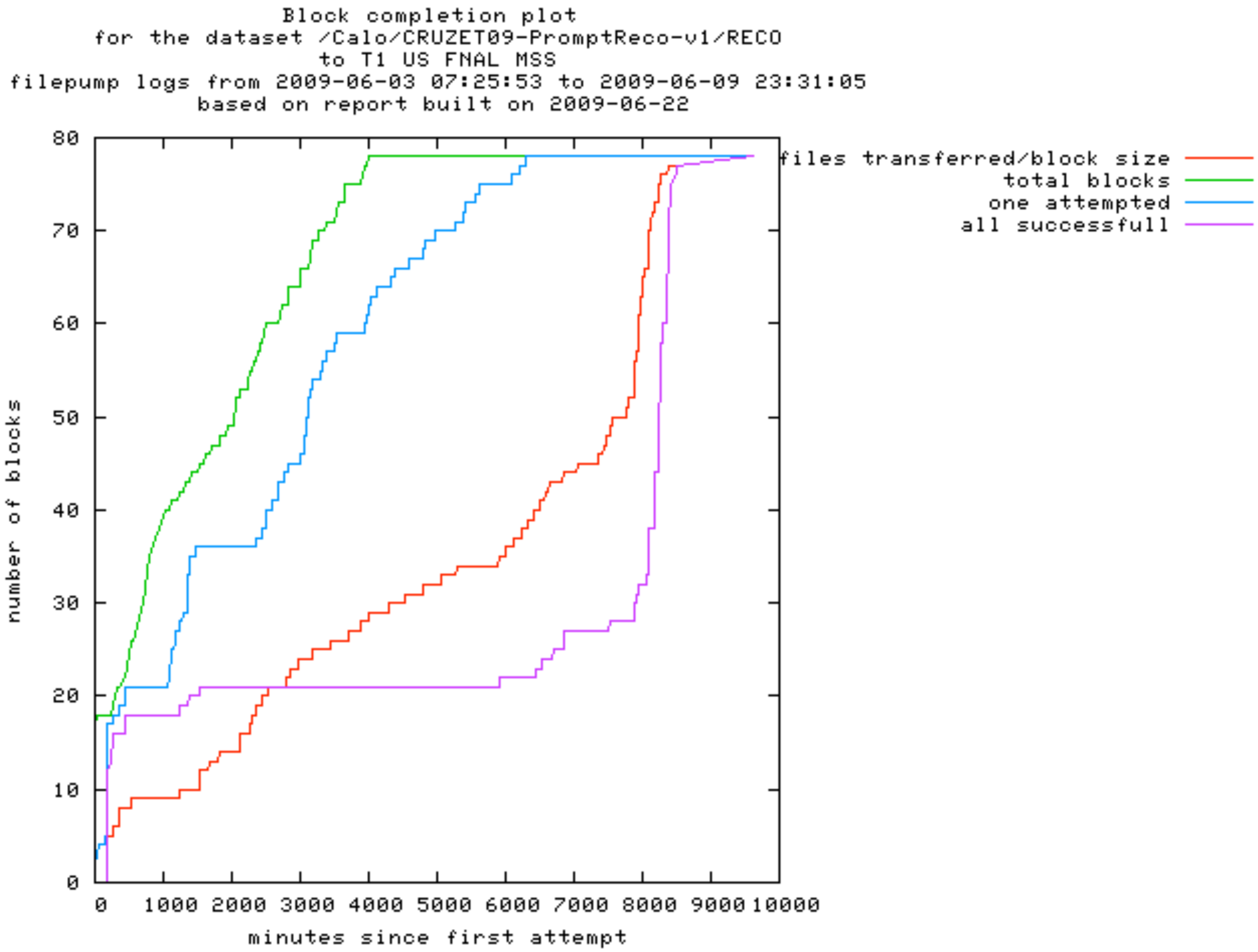
Stress T1 tape writing by exporting data from T0 to archival storage.

Observe latency between start of transfer and file written to tape, sometimes with long tails.

Latency impacted by state of tape system at given site.

- FNAL case -- correlation with tape migration backlog

T0→FNAL tape



Transfers: T1 → T1

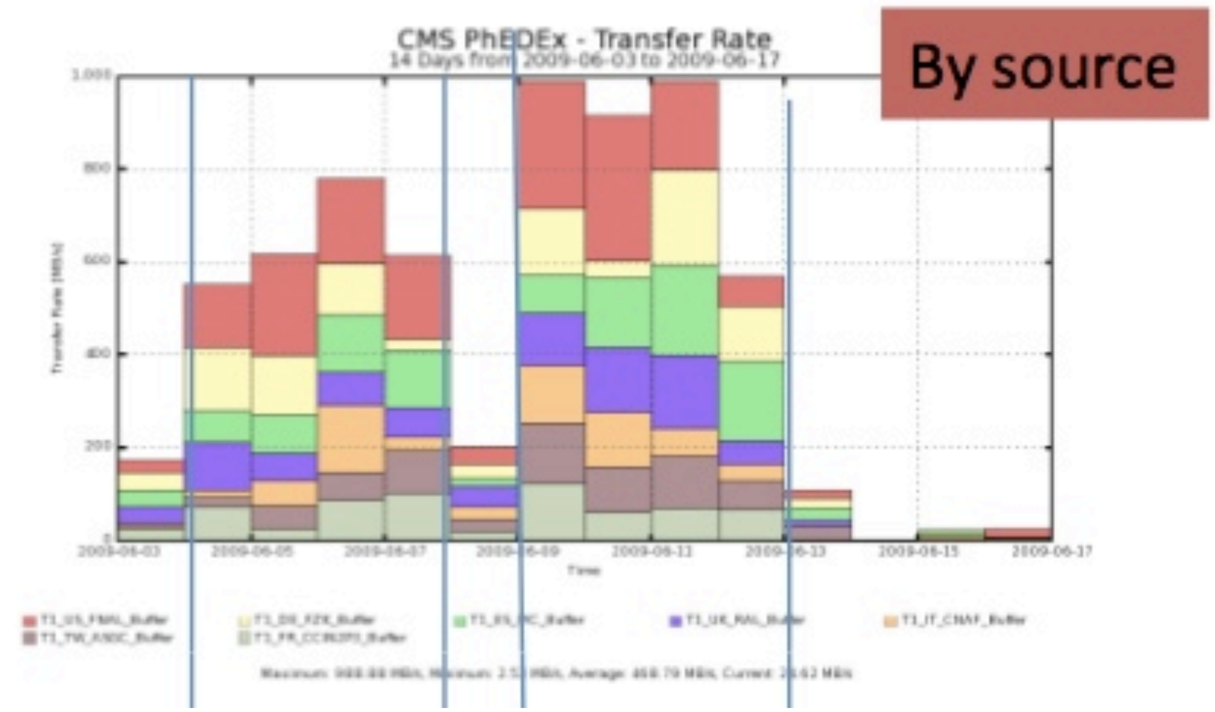
All T1's host full copy of AOD. When a T1 reprocesses their custodial fraction of AOD, the new dataset must be synchronized across all T1's.

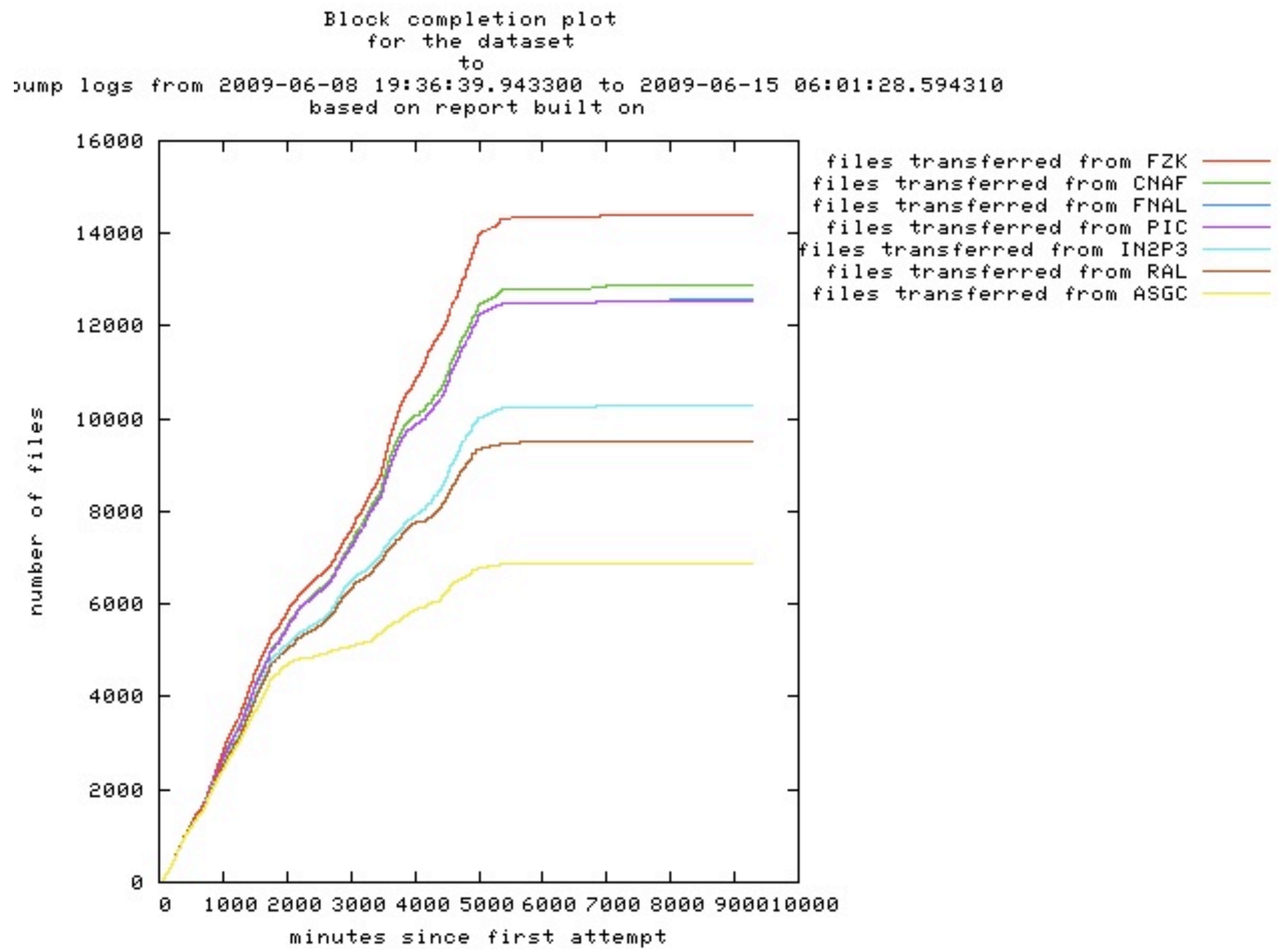
Tested synchronization with 50 TB dataset; goal was to complete in 3 days.

- ➡ Requires 1215 MB/s sustained
- ➡ Achieved 989 MB/s

Clever rerouting: Files routed over fastest links, so once B has A's files, C will get from B instead of A if former transfer is faster.

Learning how to take advantage of this, reduce site configuration issues, etc.



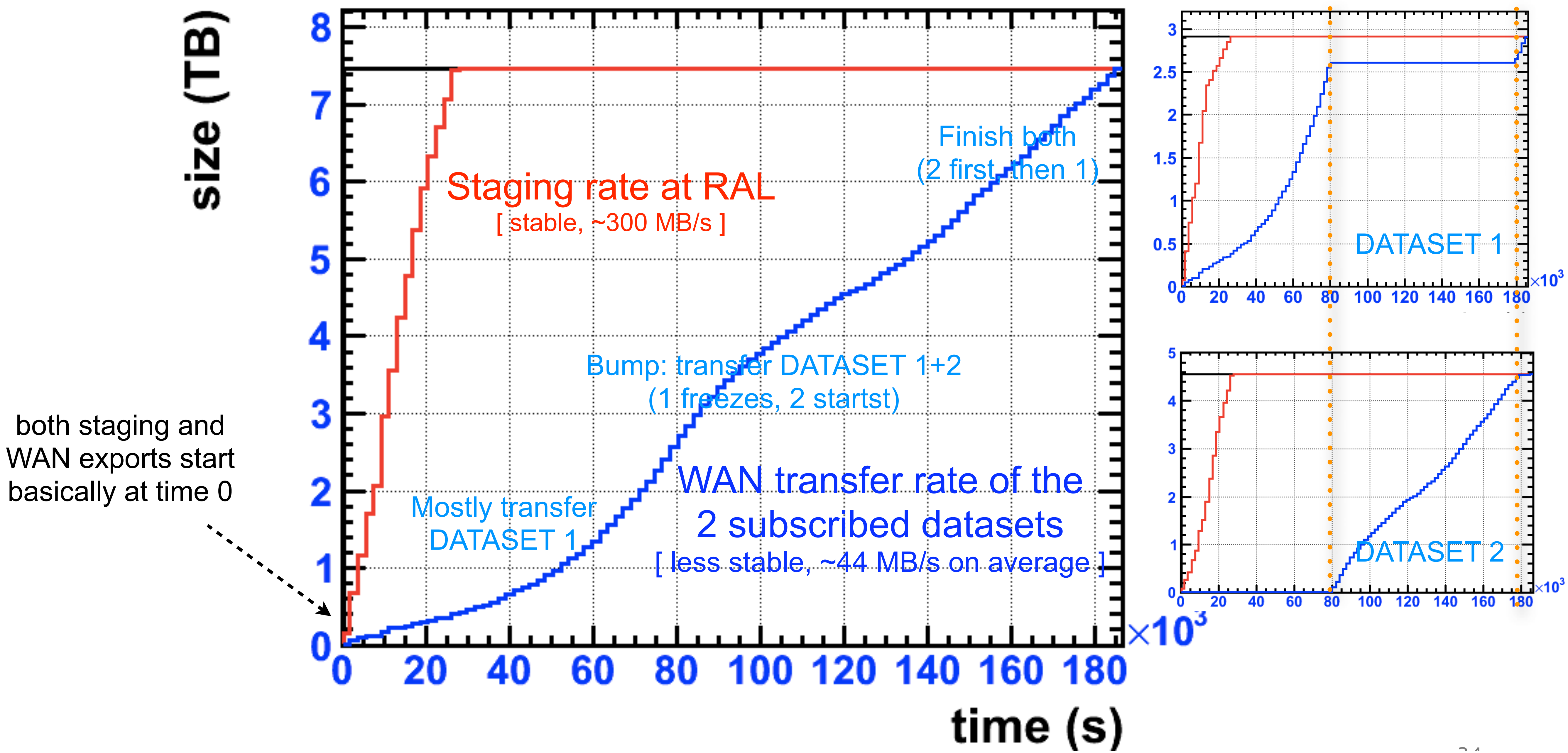


Transfers: T1 → T2

Stress T1 tape by activating T1 → T2 transfers of files on tape, not on disk.

Rate targets achieved, additional load on tape systems observed.

Pre-staging techniques and organization of files on tape will improve this.



T0 tape writing

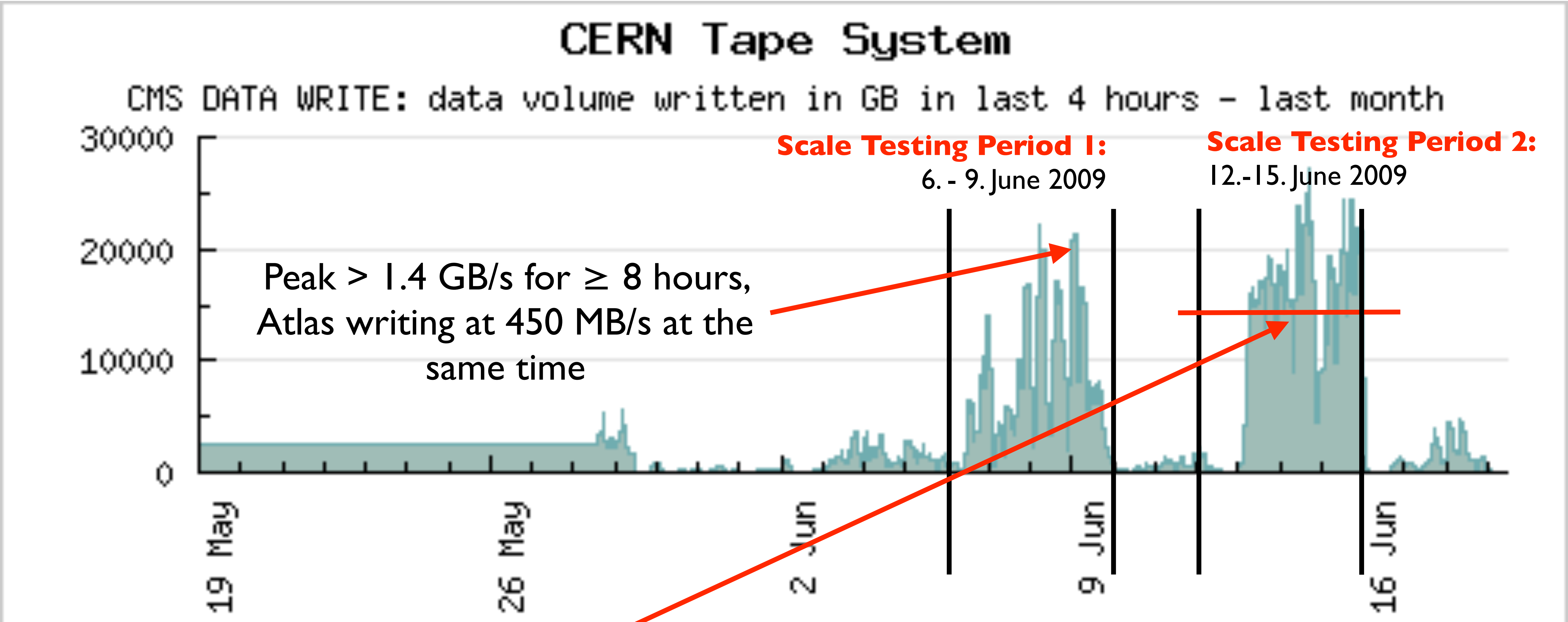
T0 does first-pass reconstruction, then saves archival copy of RAW+RECO

- Repacking of detector streamers to RAW format is I/O intensive, while reconstruction is CPU intensive → do repacking exercise to maximize tape rates.
- Can CMS archive data to tape at sufficient rates, while other experiments are doing the same?

"Sufficient" is difficult to define, as ~50% duty cycle of machine allows catch-up time -- estimate 500 MB/s.

Test schedule constrained by need to handle real detector data from cosmic ray runs; tested for a 4-day and a 5-day period over two weeks.

T0 tape writing results



Sustained > 1 GB/s for 3 days, No overlap with Atlas

500 MB/s easily exceeded in both testing periods.

Main lesson learned: need to improve monitoring of T0, especially that of reading and writing rates by VO.

T1 processing and tape staging

T1's hold custodial copies of datasets, and will re-reconstruct them multiple times.

- In 2010, envision 3 re-reco passes, each 4 months long -- overlapping!
- During early data taking, all RAW data and several RECO versions will fit on T1 disk → efficient processing.
- But as collected dataset gets bigger, it will have to be staged from tape to disk for reconstruction → potentially inefficient processing.
- Pre-staging required to maximize CPU efficiency -- never tested by CMS on this scale or with such coordination

STEP09 at T1 investigated

- tape system pre-stage rates and stability of tape systems
- ability to perform rolling re-reconstruction

Pre-staging

STEP09 test established a rolling re-reconstruction scheme:

- Day 0: Pre-stage amount of data that could be re-reconstructed in one day from tape to disk.
- Day 1: Process Day 0 data, pre-stage Day 1 data
- Day 2: Purge Day 0 from disk, process Day 1 data, pre-stage Day 2 data
- And so on....
- "one day of re-reconstruction" varied by site custodial fraction

CMS does not (yet) have a uniform way of handling pre-staging within the workload management system:

- Three different implementations emerged across the seven T1 sites.
- All three did work, and experience gained will be used to design a final pre-stage system for long-term use.

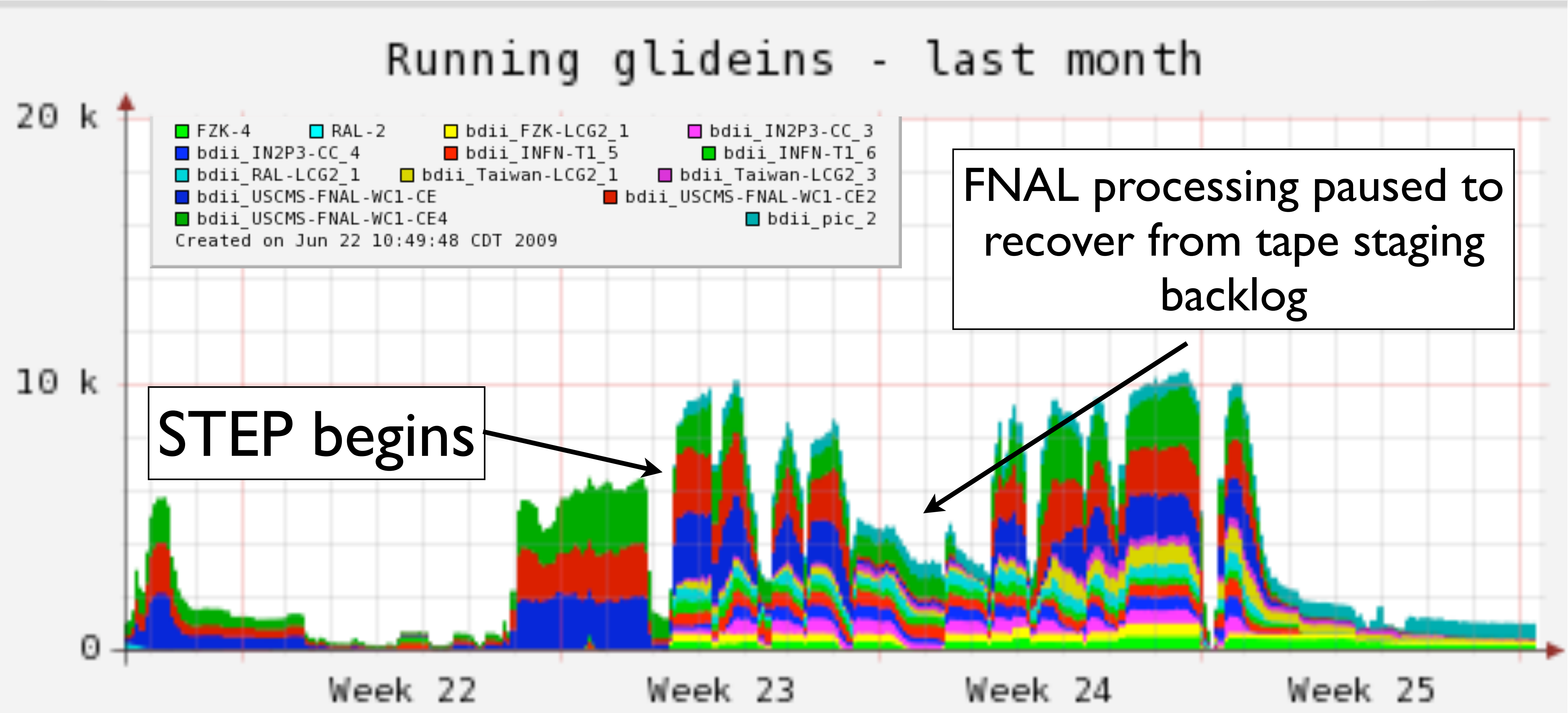
T1 tape performance

Site	Target [MB/s]	2-Jun	3-Jun	4-Jun	5-Jun	6-Jun	7-Jun	8-Jun	9-Jun	10-Jun	11-Jun
FZK	85	Tape system not available							Participated in pre-staging but performance not clear		
PIC	50	60	61	106	83	Samples not purged	Samples partially on disk	99	142	123	142
IN2P3	52	Tape system not available, scheduled downtime						96	99	120	103
CNAF	56	380	300	160	240	240	270	105	80	125	240
ASGC	73		140	170	190	160	145	150	140	150	220
RAL	40	250	230	160	140	135	190	170	100	220	180
FNAL	242	280	200	200	120	Still staging previous day	Recovering from backlog		379	380	400

Most sites had very good performance

- IN2P3 had scheduled downtime, FZK tape system unavailable at first
- Large scale at FNAL triggered problems that were quickly solved

T1 re-processing performance



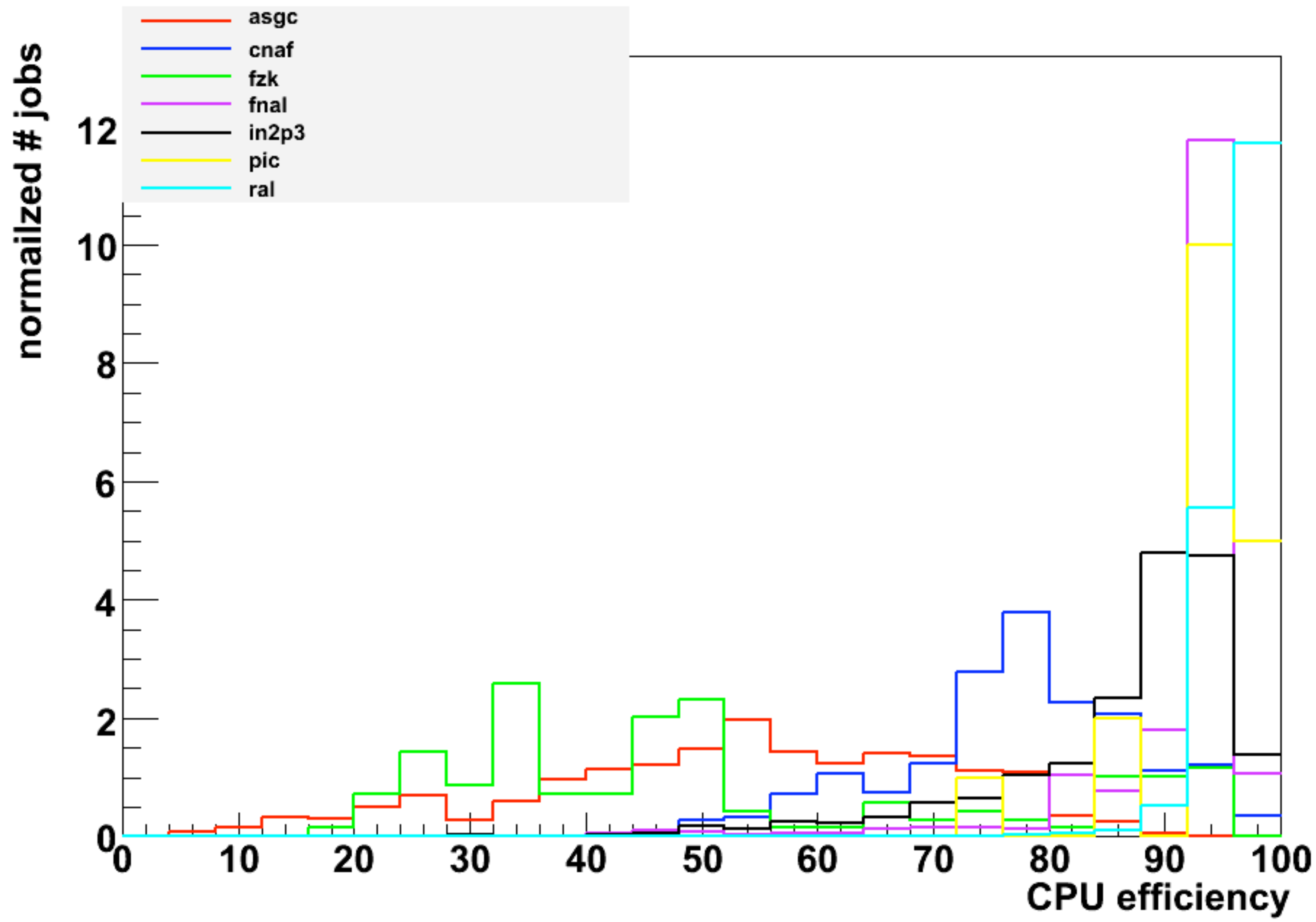
All re-proccessing jobs run by single operator using glideIn pilots

No trouble getting pledged number of batch slots from sites, fairshare between experiments functional

T1 CPU efficiency

CPU time/wall-clock time is a measure of job efficiency; want to spend more time processing than waiting for files to come off tape.

CPU efficiency for a typical day:

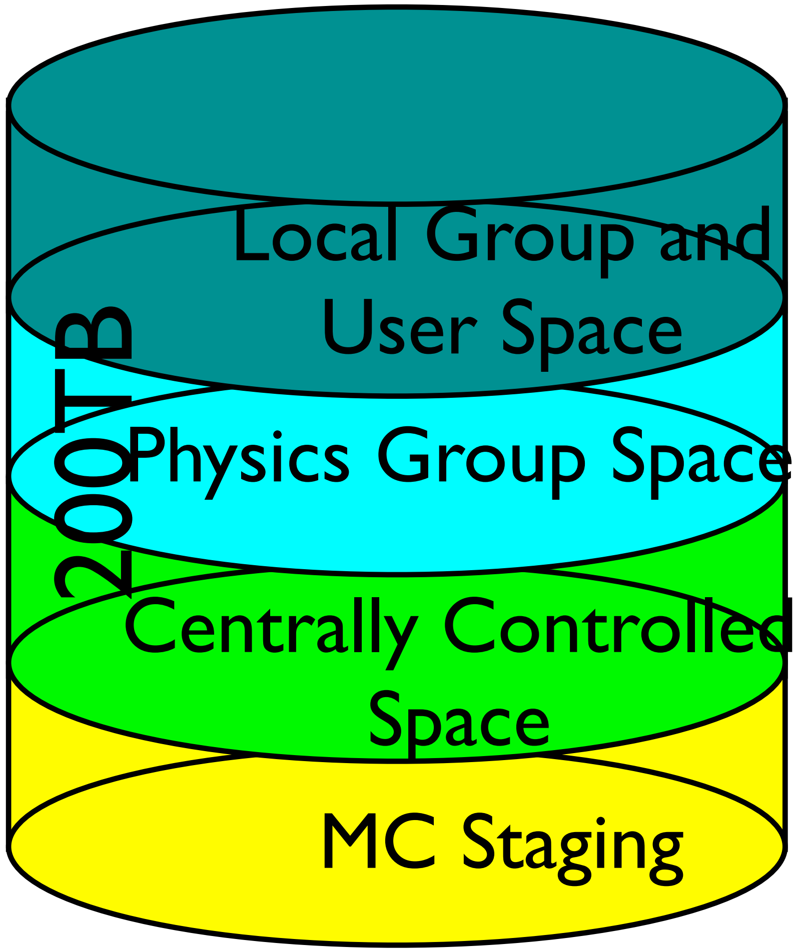


Great variability across T1 sites. However, pre-staging generally observed to greatly improve efficiency.

CMS T2 analysis model: data distribution

In CMS, jobs go to the data -- distribute data for efficient resource use.

Nominal T2 storage is 200 TB, x ~40 T2 sites = huge!



Some amount set aside for centrally-controlled activities (e.g. distribution of datasets of wide interest) and local activities (e.g. making user-produced files grid accessible.)

But bulk is allocated to the various CMS analysis groups for distribution of "their" interesting data.

T2_US_Nebraska Group Usage

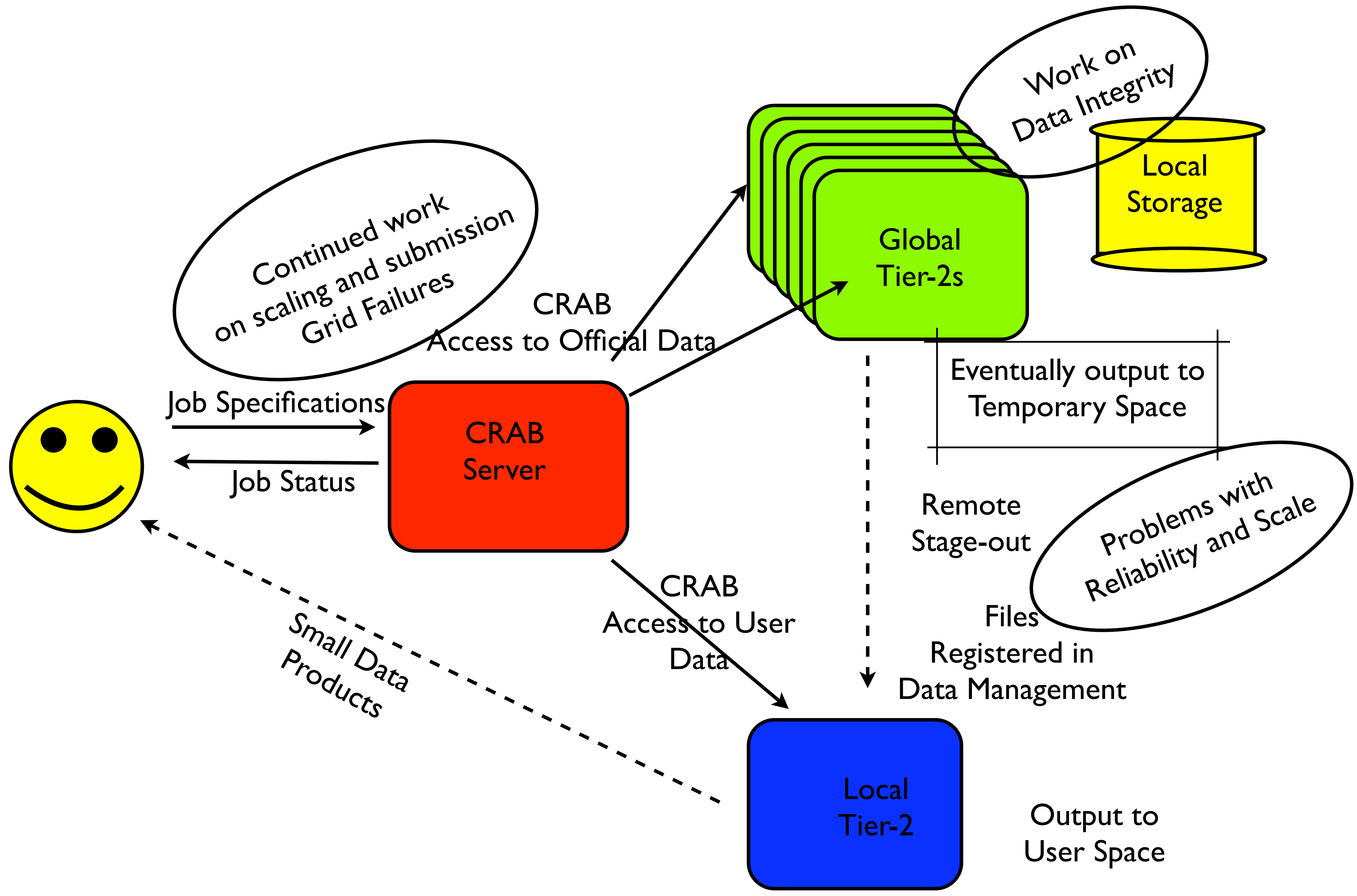
Group	Subscribed	Resident
DataOps	5.12 TB	5.12 TB
FacOps	1.04 TB	1.04 TB
b-tagging	11.47 TB	11.47 TB
local	39.34 TB	39.34 TB
qcd	3.04 TB	3.04 TB
top	25.83 TB	25.74 TB
tracker	4.59 TB	4.59 TB
undefined	37.39 TB	37.39 TB
	127.82 TB	127.73 TB

- 17 such groups in CMS
- Currently no site supports more than 3 groups, no group affiliated with more than 5 sites, manageable number of communication channels
- 7 US T2's support all groups

CMS T2 analysis model: workflow

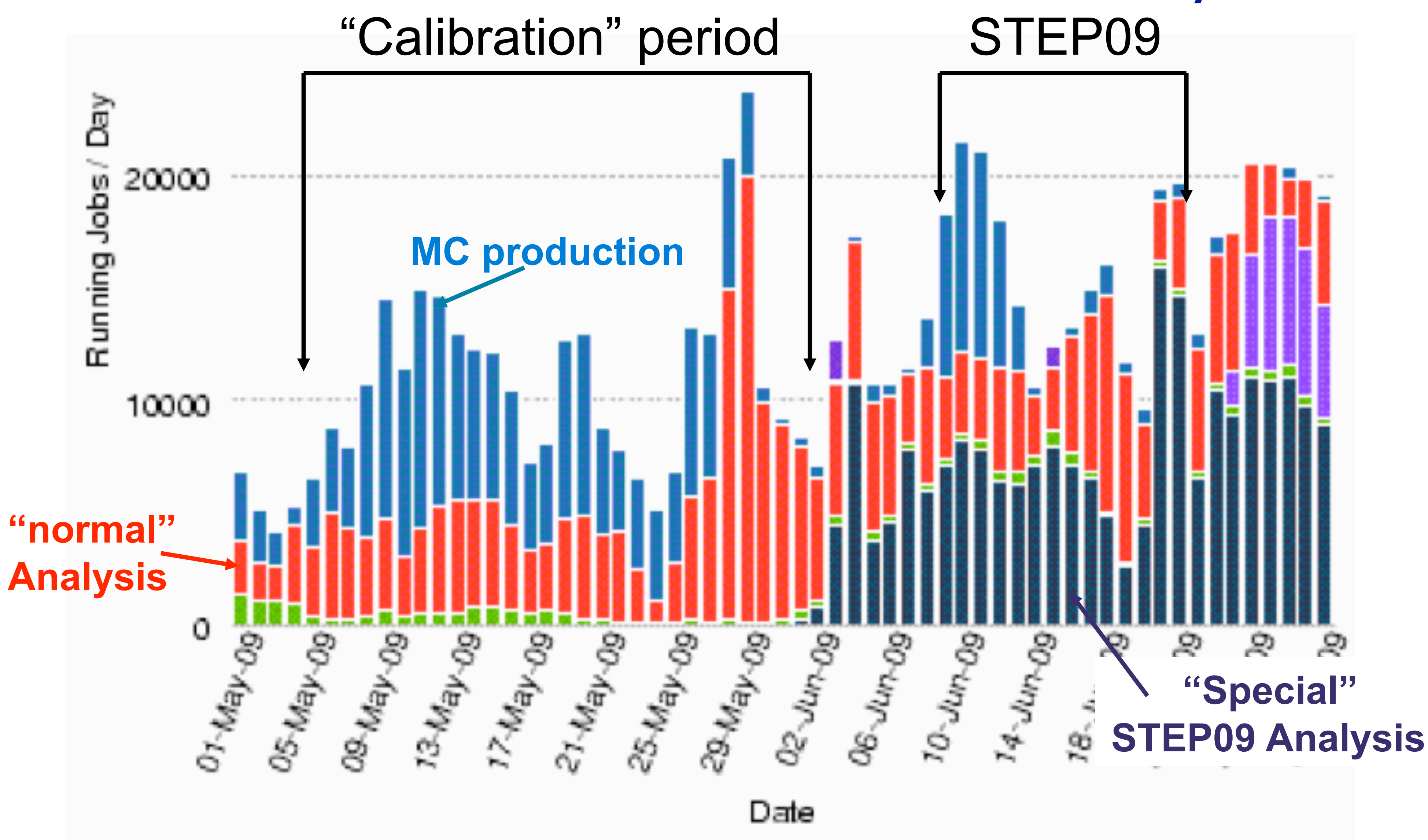
How will 2000 collaborators interact with T2 sites through the grid?

- CMS Remote Analysis Builder (CRAB) shields the user from the underlying complexity, but many things have to succeed



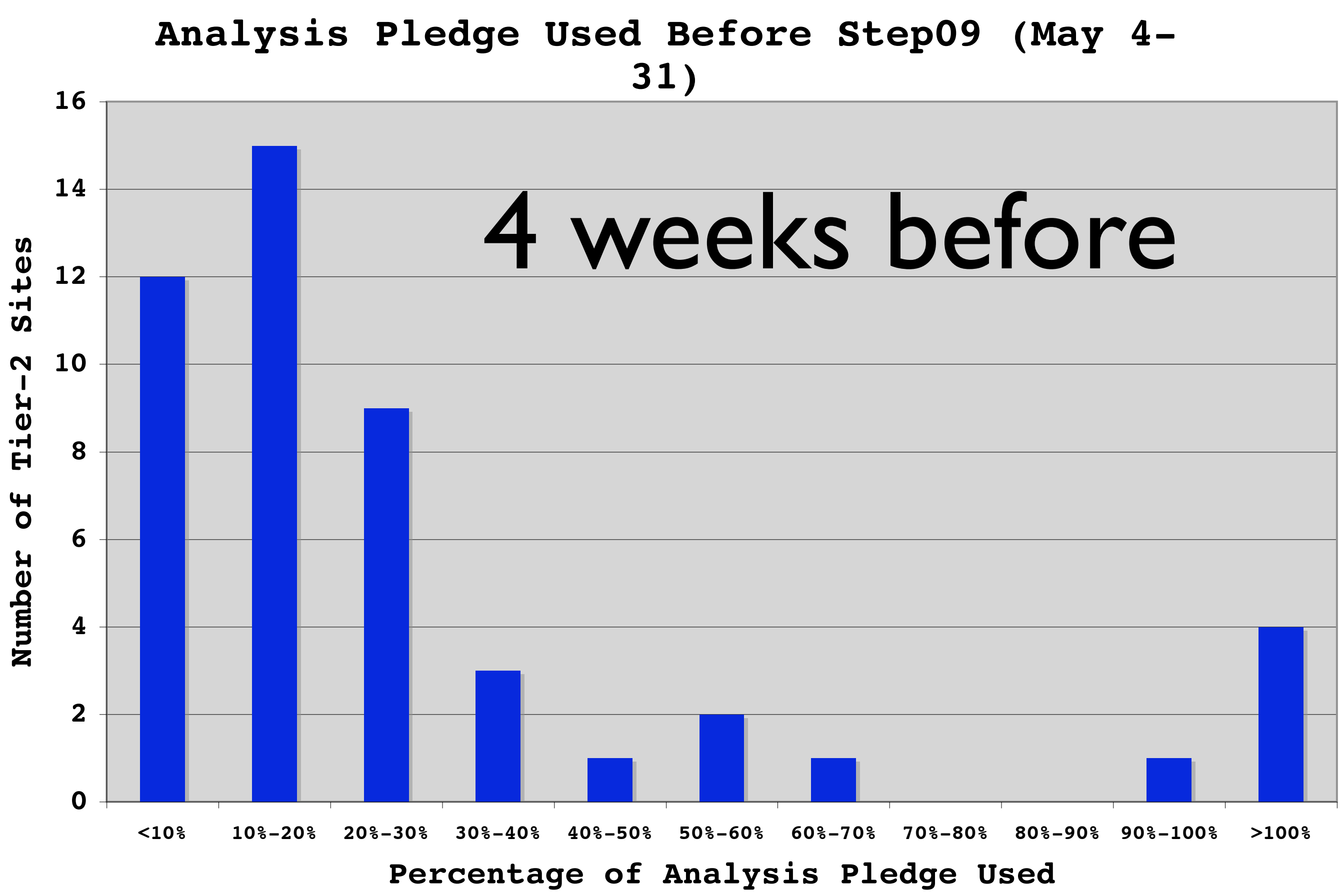
STEP09 T2 analysis tests

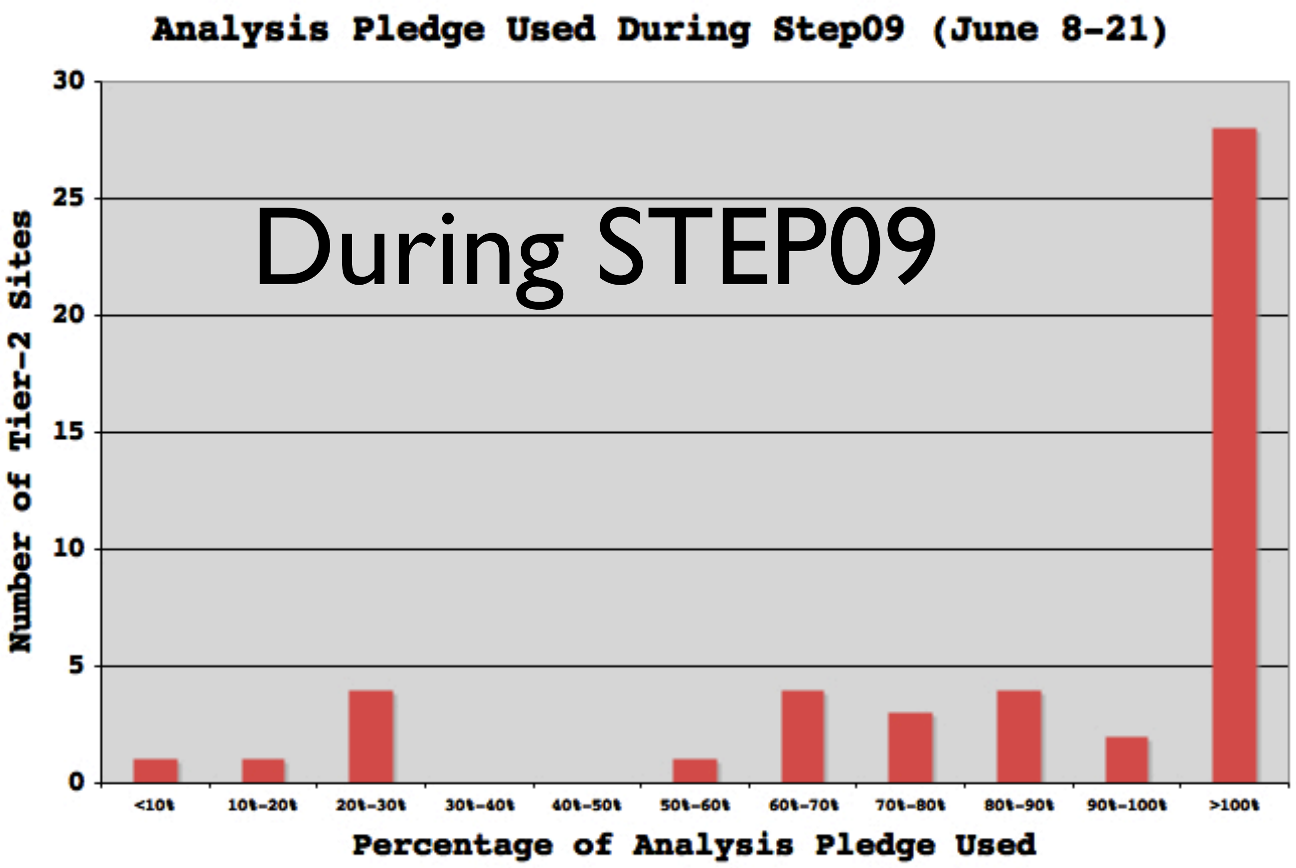
50% of T2 pledged processing resources are targeted for user analysis. 8K batch slots at the moment! STEP09 tried to fill that many slots.



During STEP09, more than saturated the pledged analysis resources, without triggering operational problems at sites. Suggests that there are resources out there that we could be using better?

STEP09 T2 analysis tests





Put another way, easily went from low utilization during typical period to high utilization during STEP09 -- bodes well for future onslaught of jobs.

Analysis jobs success rate

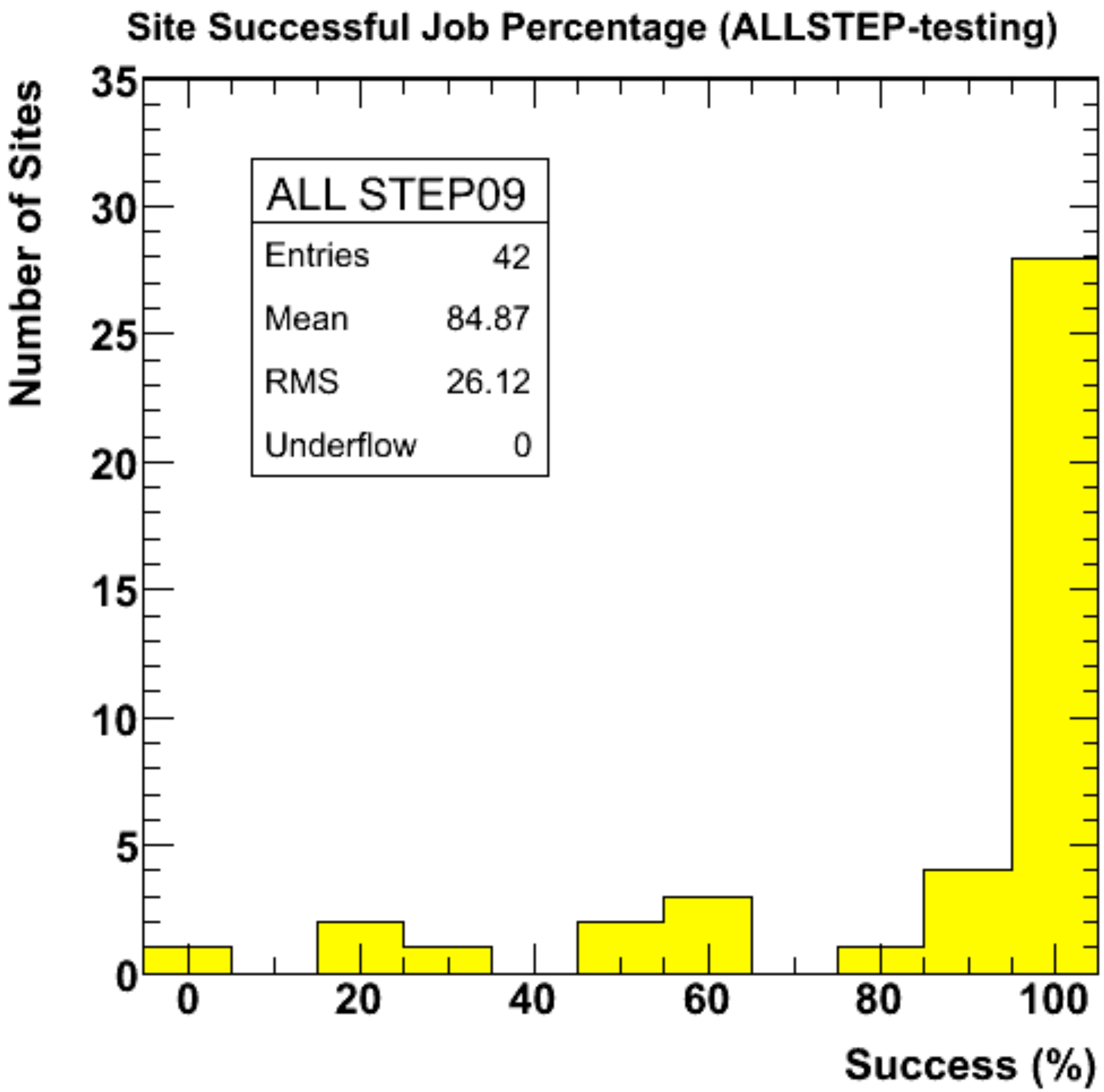
STEP09 jobs read data, used CPU but did no stageout.

Majority of sites handled these jobs perfectly; overall 80% success rate for jobs.

90% of failures were due to read failures at sites -- a clear area that needs improvement.

- The grid works?

Another issue briefly probed in STEP09 -- does data placement matter? Tried moving additional copies of popular datasets to under-used sites; did see some increase in activity there. Promising for the future....



Conclusions from STEP09

STEP09 allowed us to focus on specific key areas of the computing system (tape at T0 and T1, transfers, analysis at T2) in a multi-VO environment.

Most T1 sites showed good operational maturity:

- May not yet have deployed the resources necessary at LHC startup, but no indication of any problems scaling up.
- Not all T1 sites attained the goals; will re-run specific tests after improvements.

Tests of analysis activities at T2 were largely positive:

- Most sites were very successful.
- Easily demonstrated that we can use resources beyond the level pledged by sites.
- Have some indicators that we can use resources more efficiently.

Looking ahead

STEP09 gives us confidence that the CMS computing system will work, but there are still many challenges ahead of us:

- Long LHC run: operational impacts?
- If LHC duty cycle is low at the start, will be pressure to increase the event rate (300 Hz → 2000 Hz) and overdrive the system: will it work?
- Datasets: will divide triggered events into streams to be custodial at various T1's. Can prioritize reprocessing, but can we satisfy local interests at each T1?
- Read errors: can we make disk systems more reliable and maintainable?
- Remote stageout: present system will not scale. What will?
- Resource limitations: during a long run, will we be able to keep multiple copies of RECO data available at T2? If not, how will people adjust?

We will be learning a lot in the next year!

But confident that we are well-positioned to succeed.