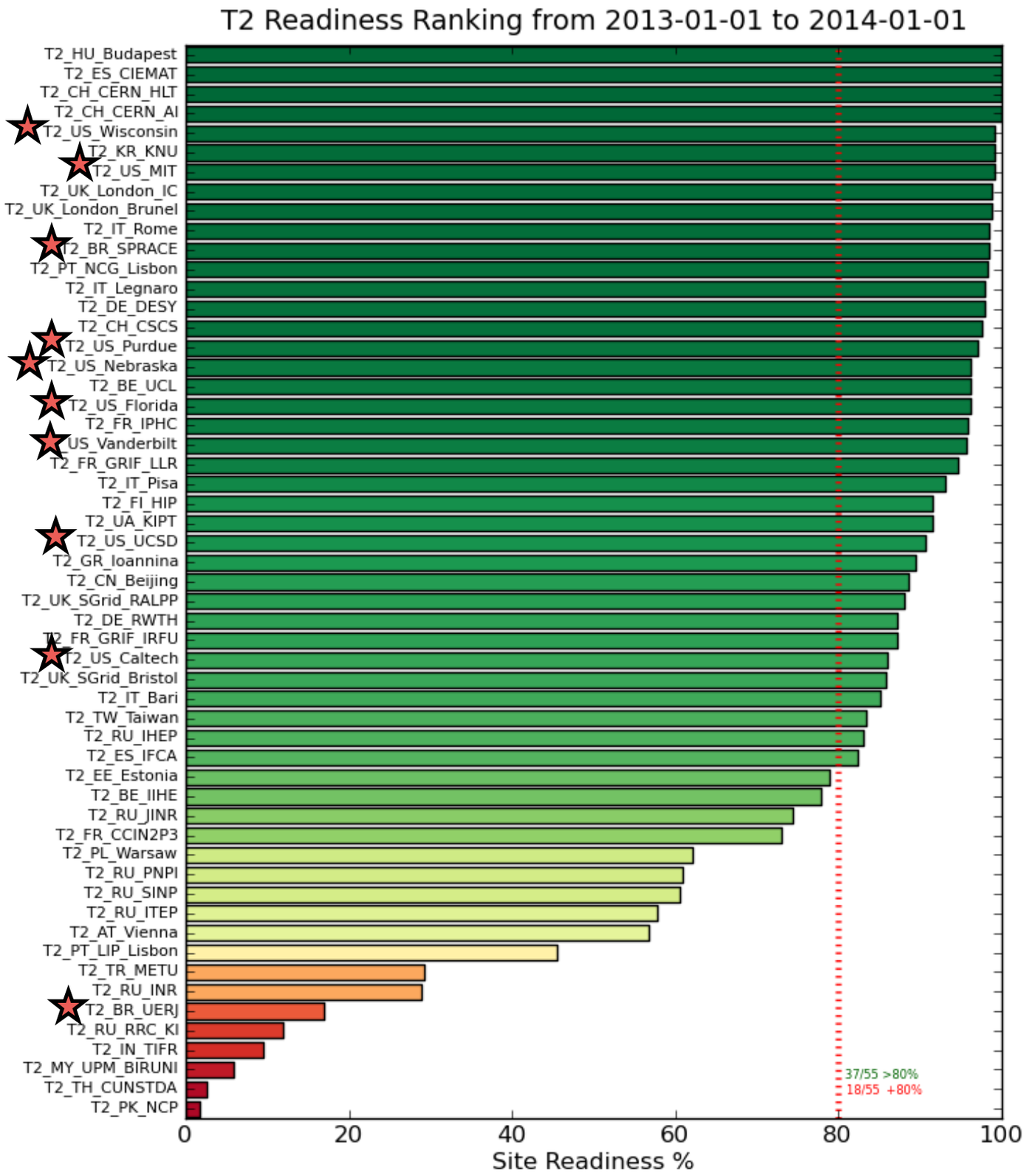
US CMS Tier-2 Status and Future

Ken Bloom

April 7, 2014

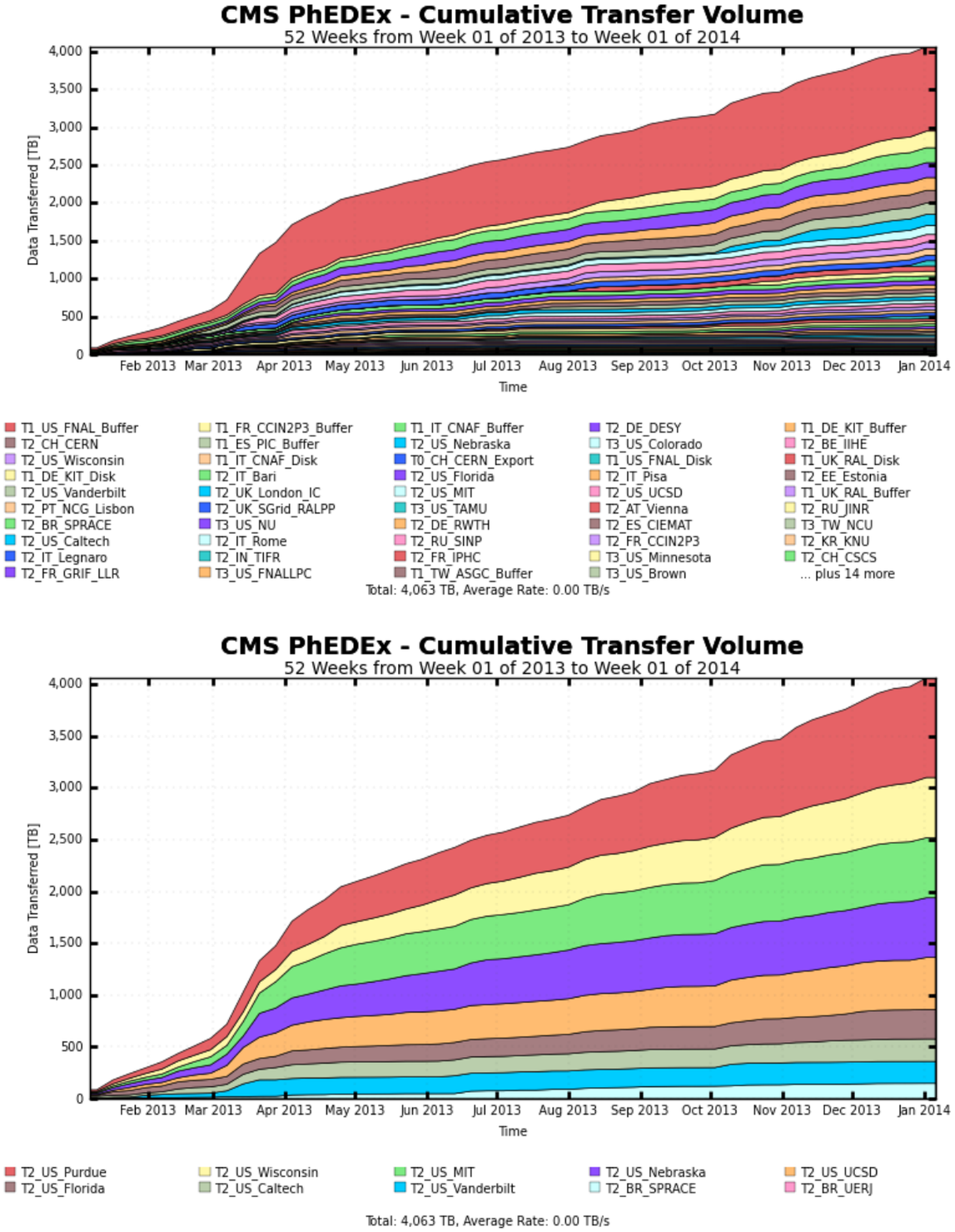
2013: another successful year

- The site readiness metric has its quirks, but is still our best measure of the usability of each site on any given day, on the basis of ability to execute basic functions, including data transfers
  - Makes allowances for short-term problems, scheduled downtimes, weekends
- CMS goal for this is >80%, all but one OSG site above

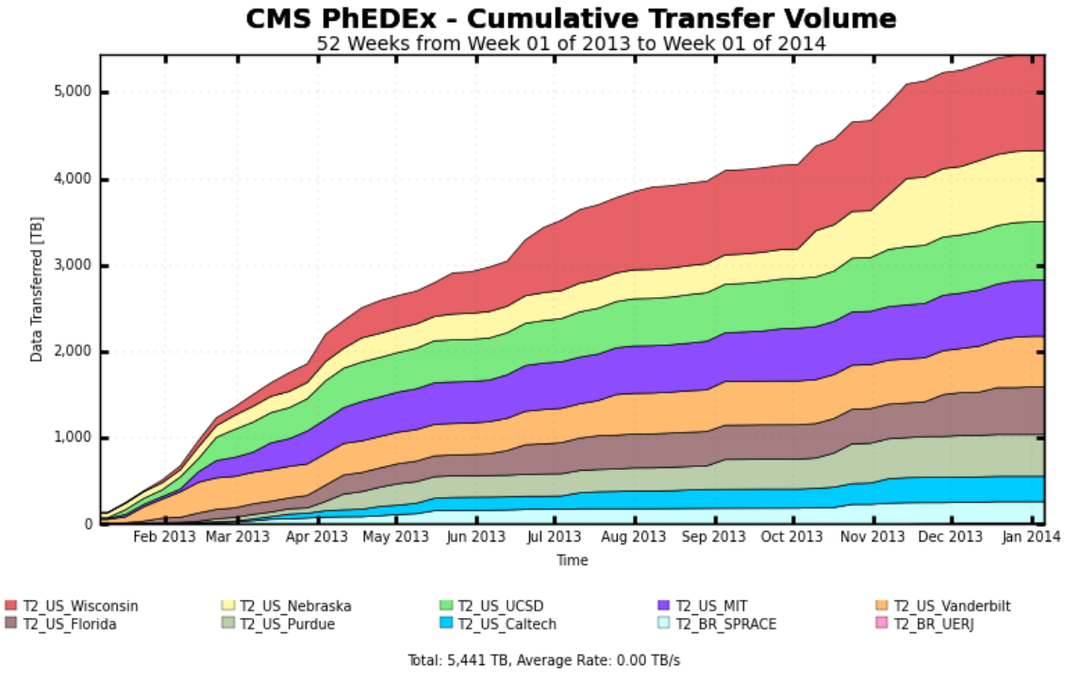


2013: another successful year

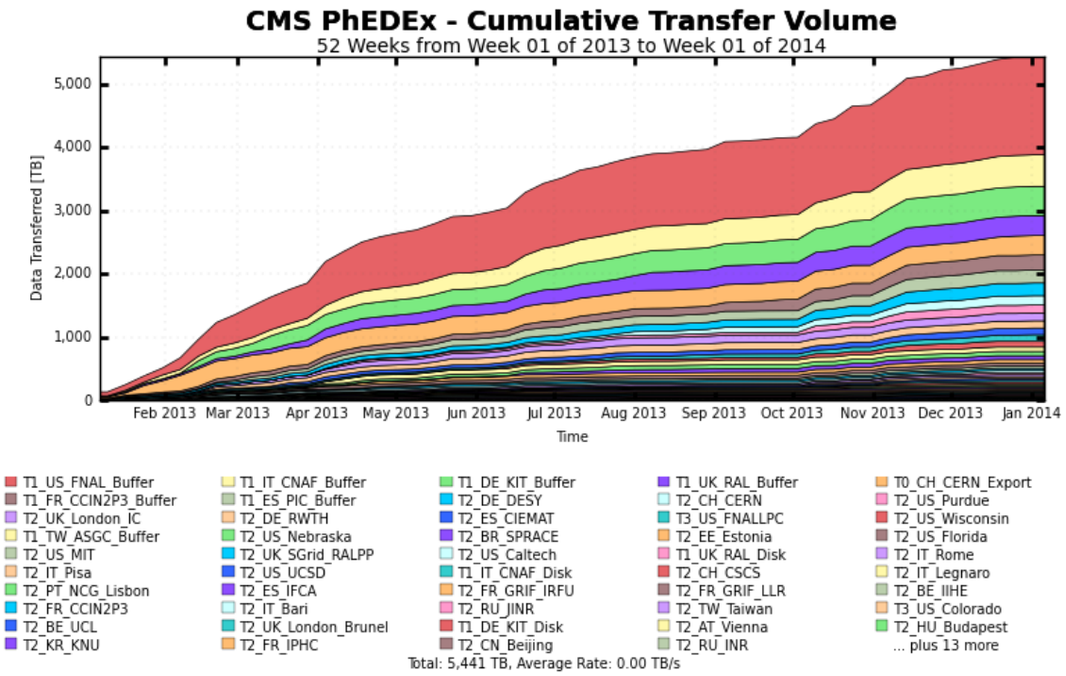
Outbound: 4.1 PB



Inbound: 5.4 PB



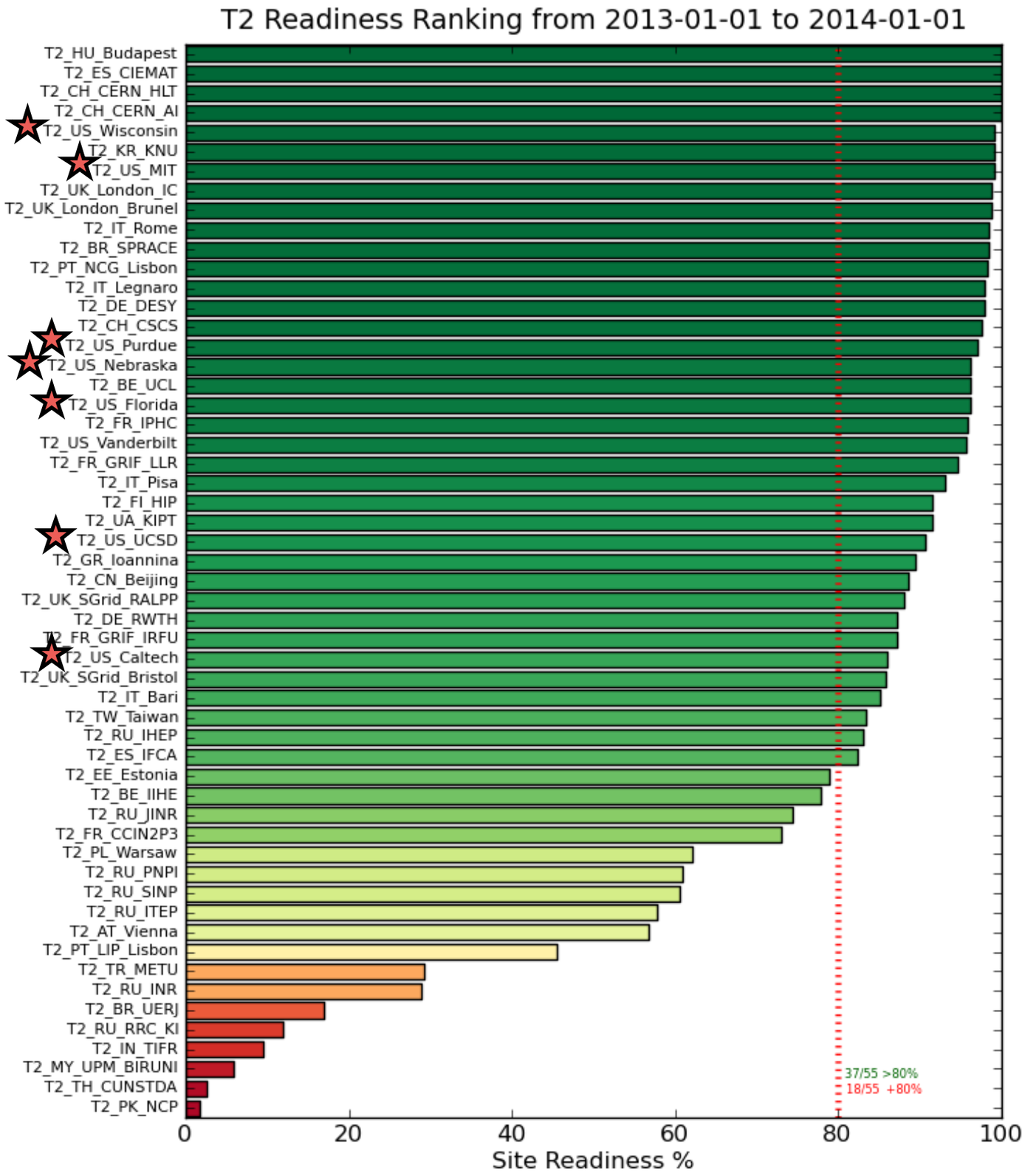
by destination



by source

2013: another successful year

- The site readiness metric measures the usability of each site on any given day, on the basis of ability to execute basic functions, including data transfers
  - Makes allowances for short-term problems, scheduled downtimes, weekends
- CMS goal for this is >80%, all US sites are above

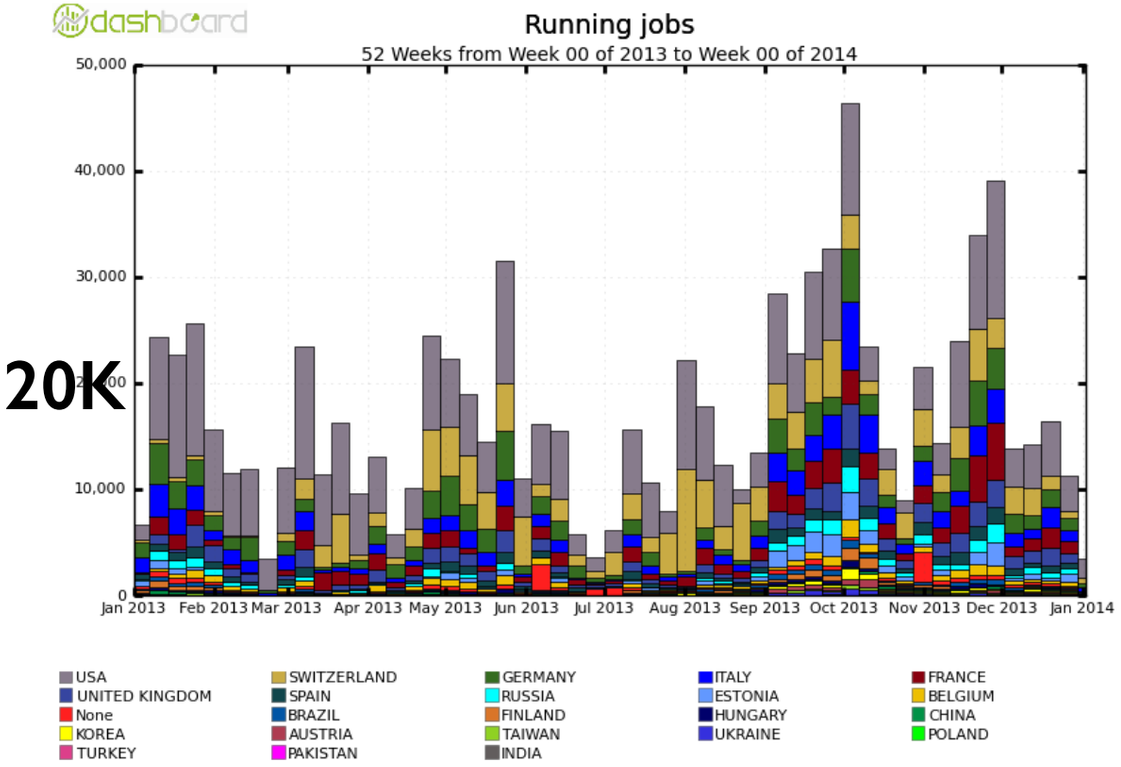


2013: another successful year

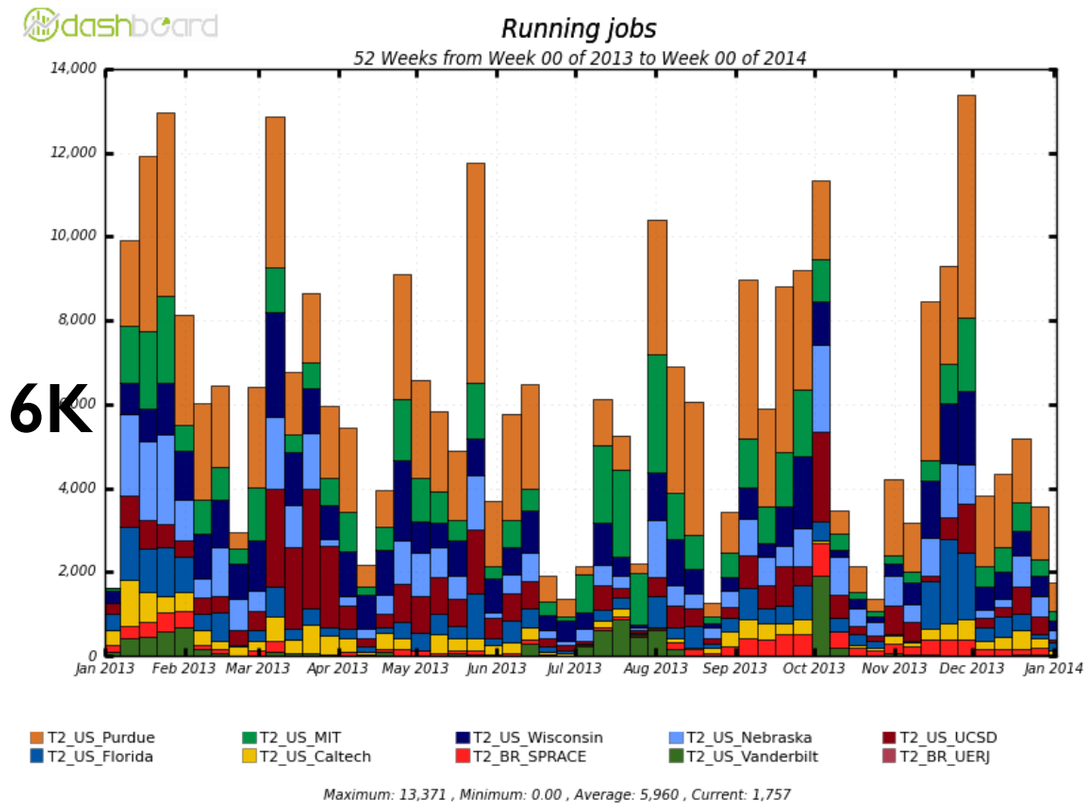
Production at T2s

Analysis at T2s

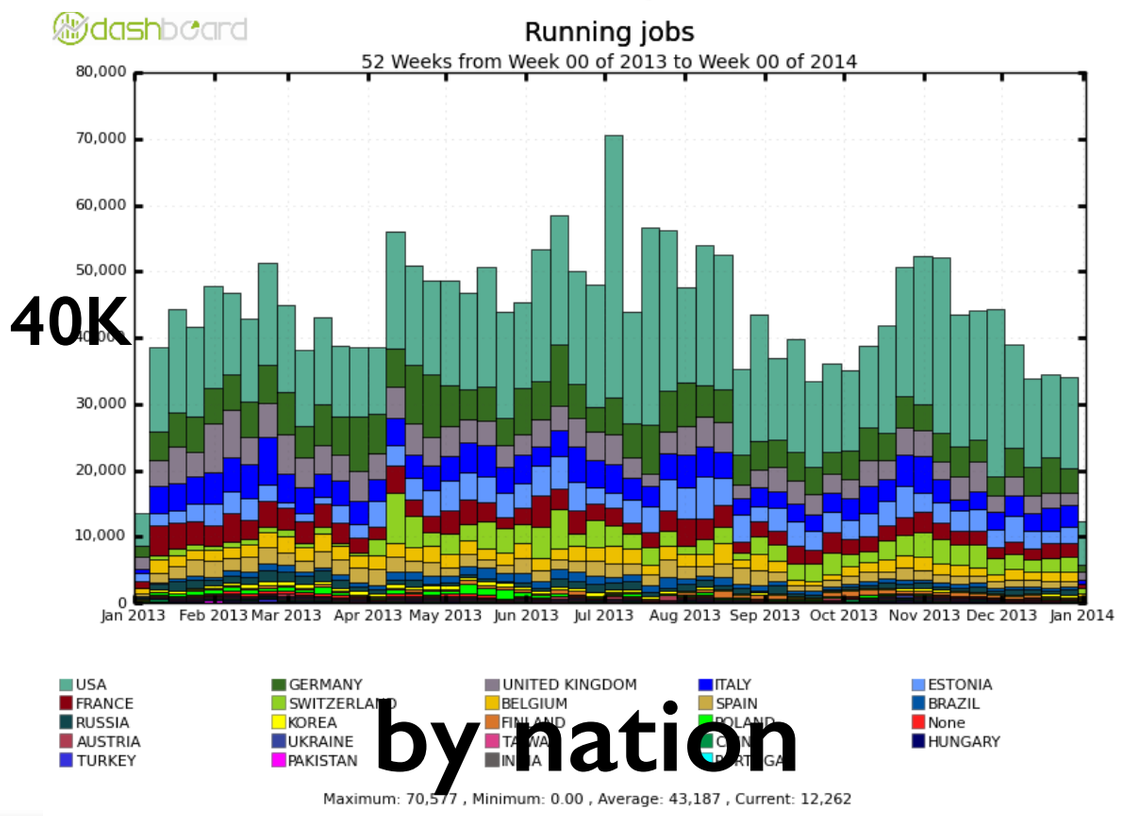
20K



6K

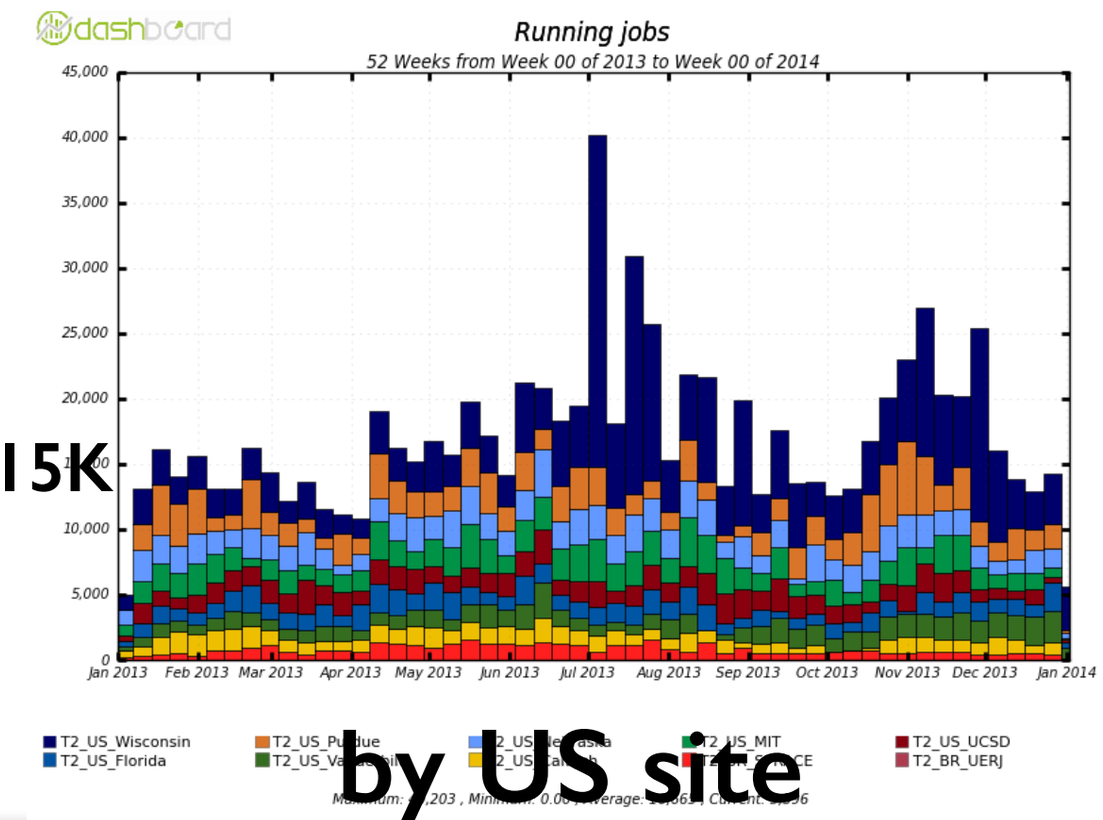


40K



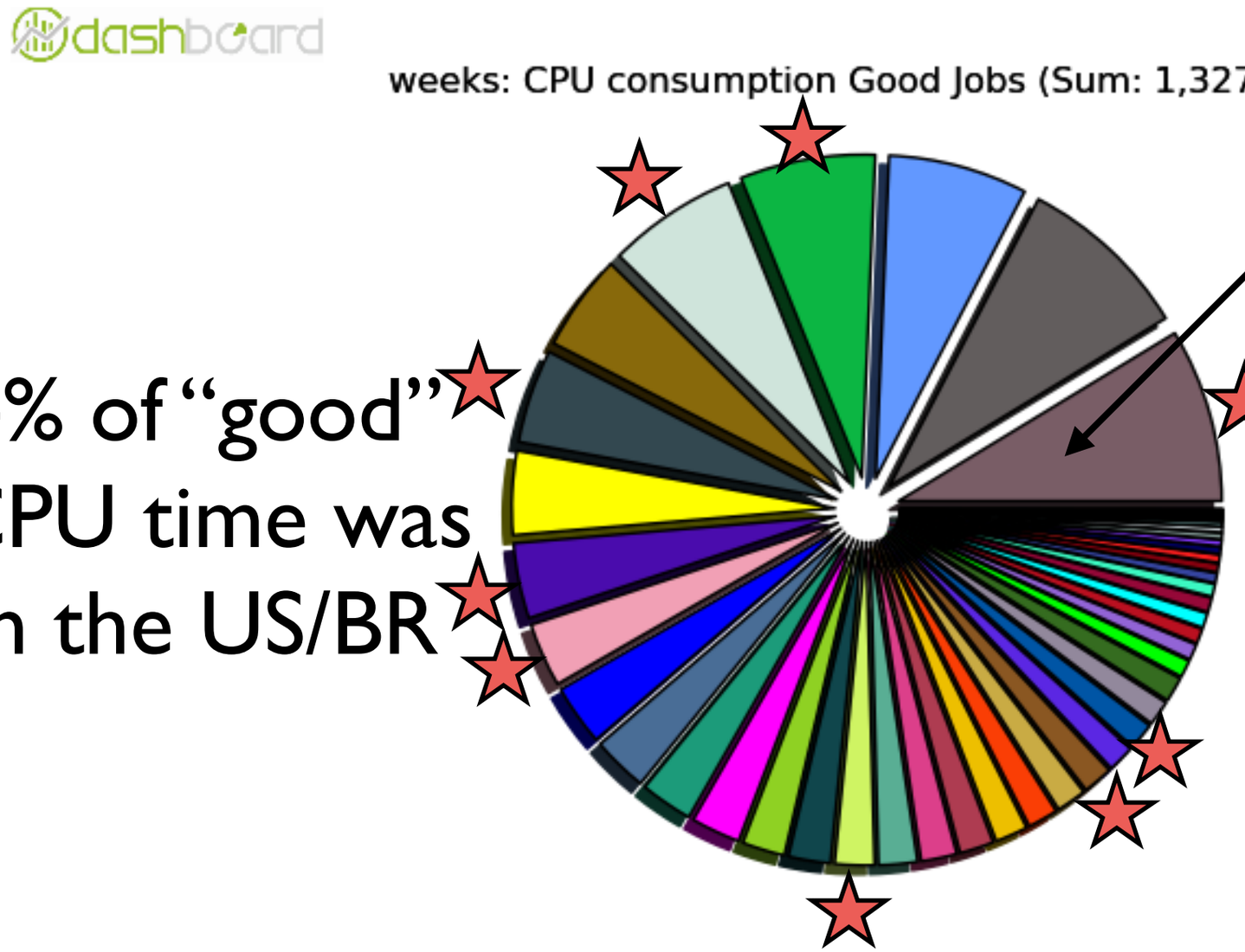
by nation

15K



by US site

2013: another successful year

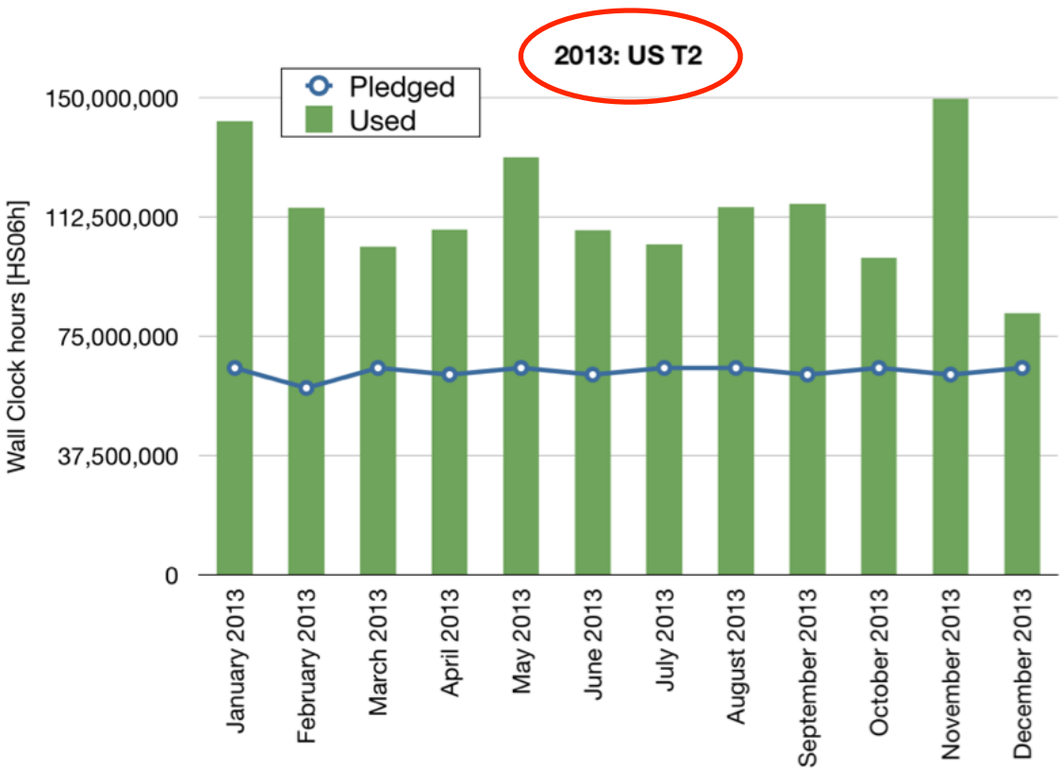


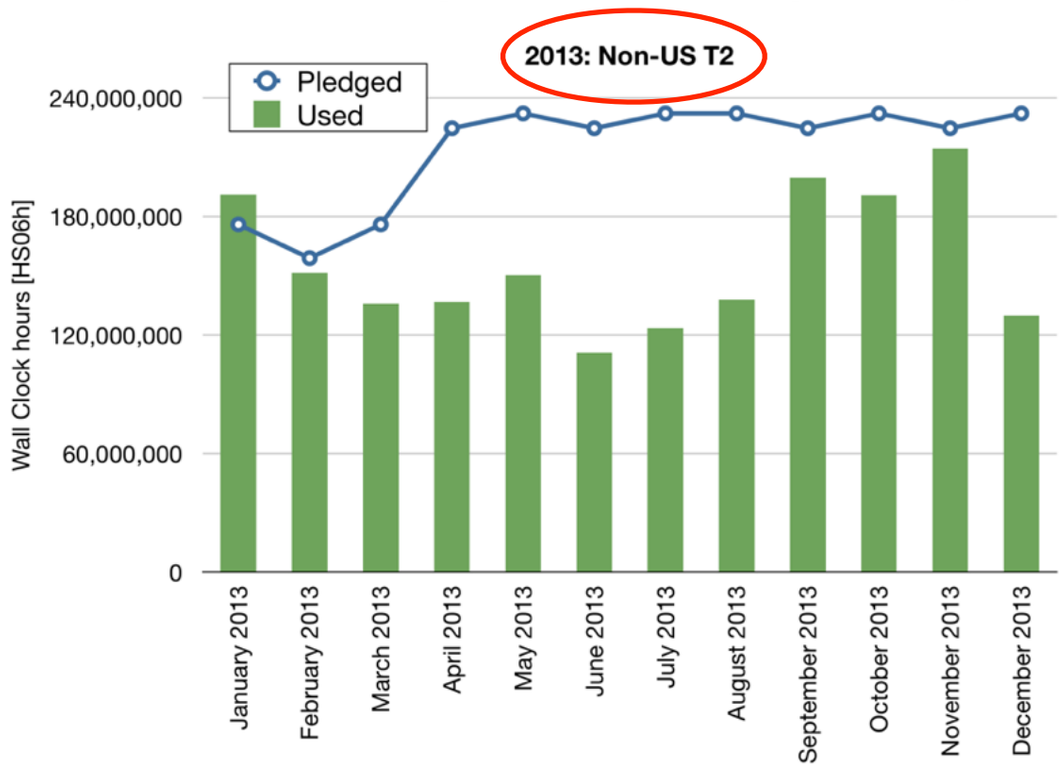
38.4% of "good" T2 CPU time was from the US/BR

UCSD is biggest because of special opportunistic use of SDSC resources this year!

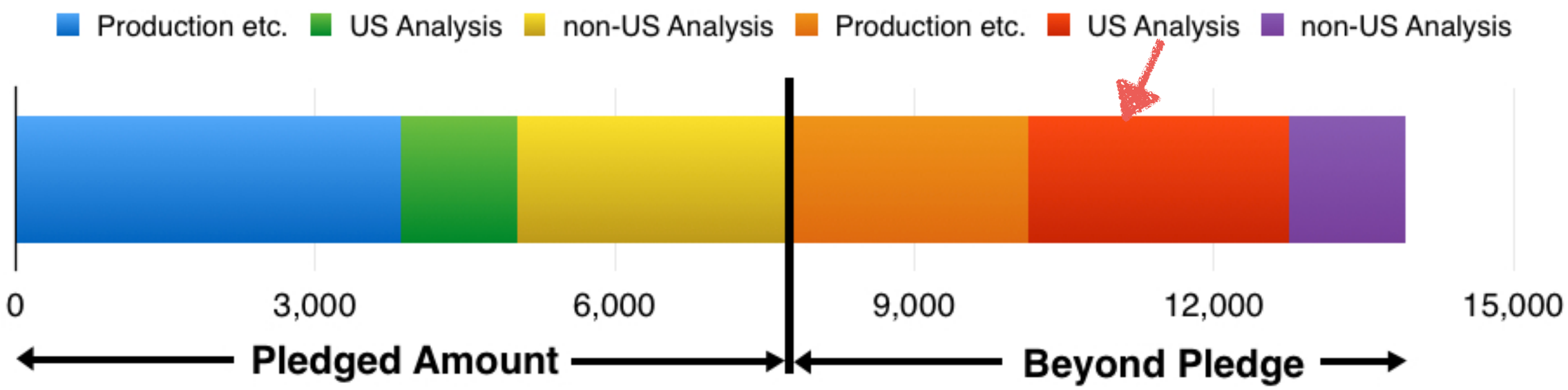
Processing usage

- Unsurprisingly, CMS uses US resources well beyond the pledge:



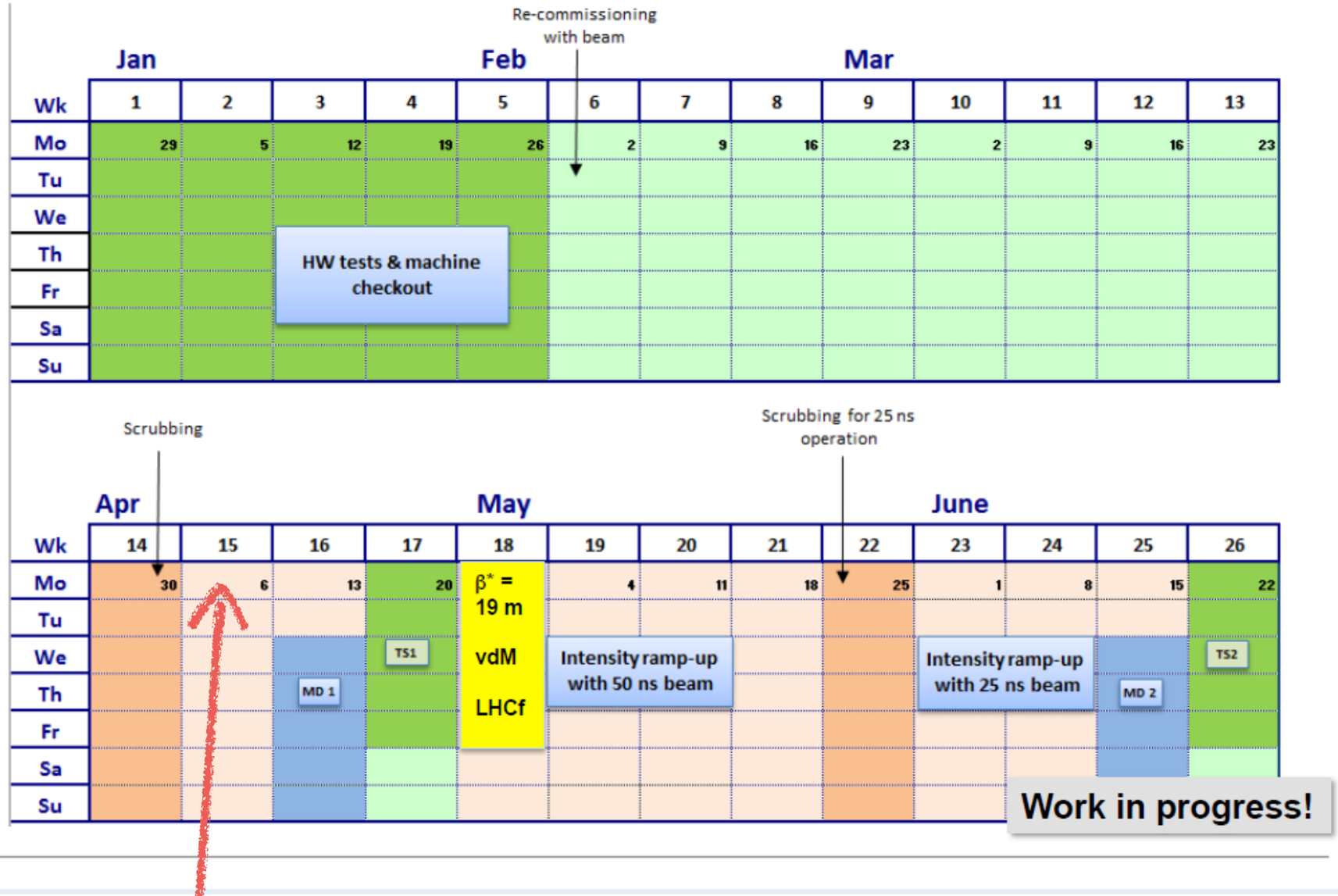


- How are these resources used? Best estimate (by J.Letts):



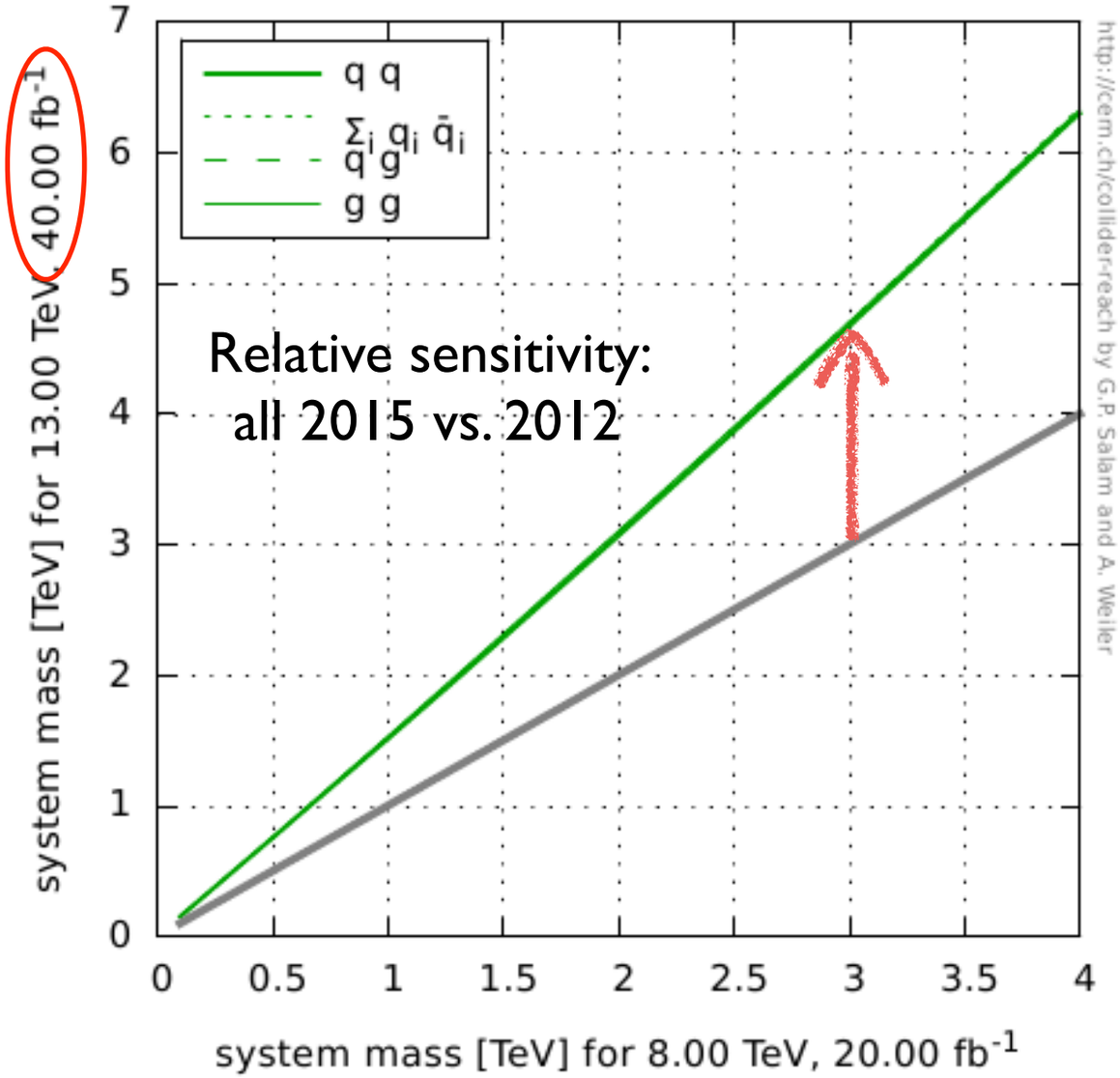
One year to go

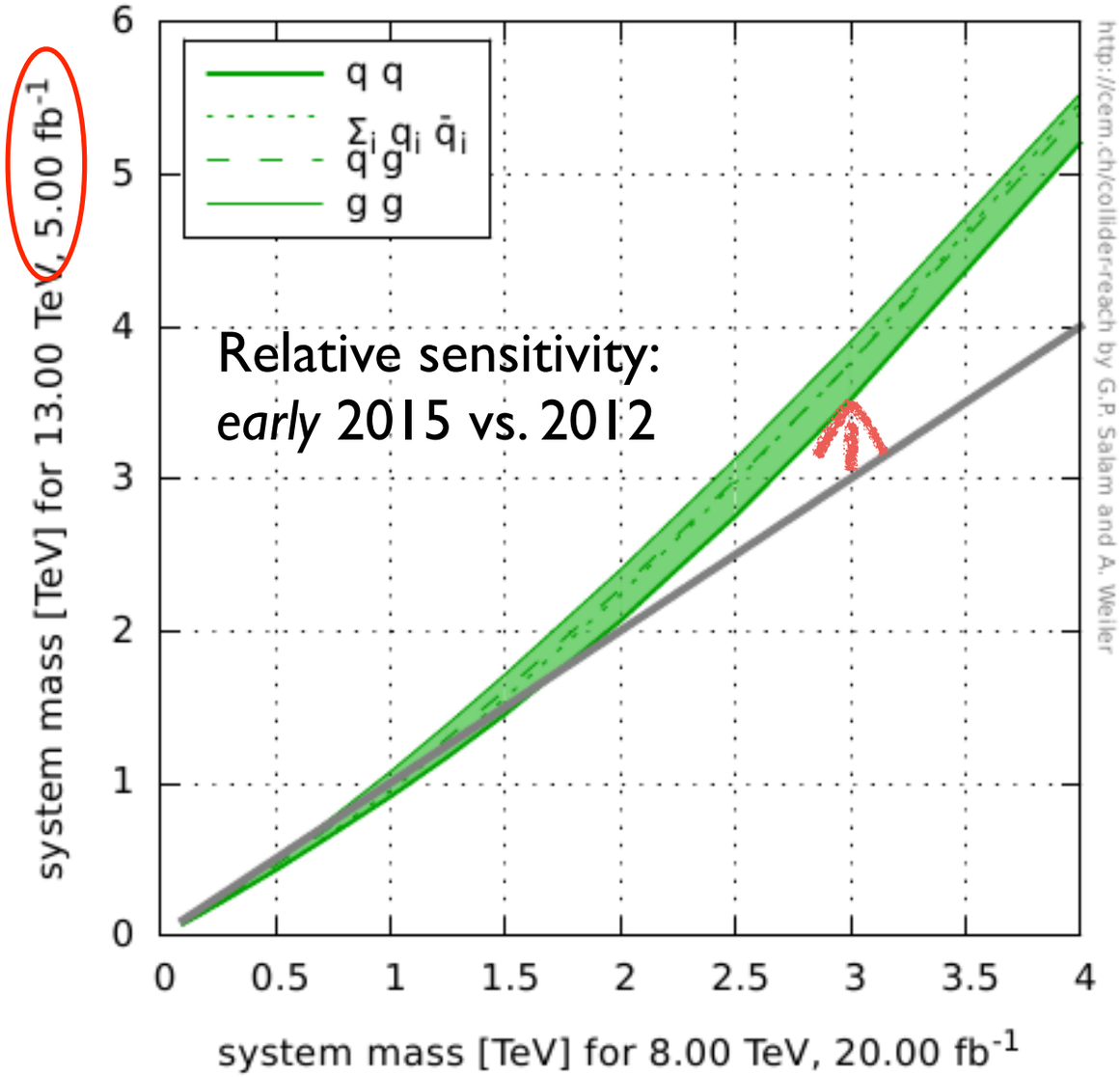
Early 2015 schedule → "Nominal" 25 ns Physics from July'15



- First Monday of April: LHC collisions resume
- Higher beam energy, higher beam intensity, higher pileup rate
- Larger trigger rates, larger events and processing times

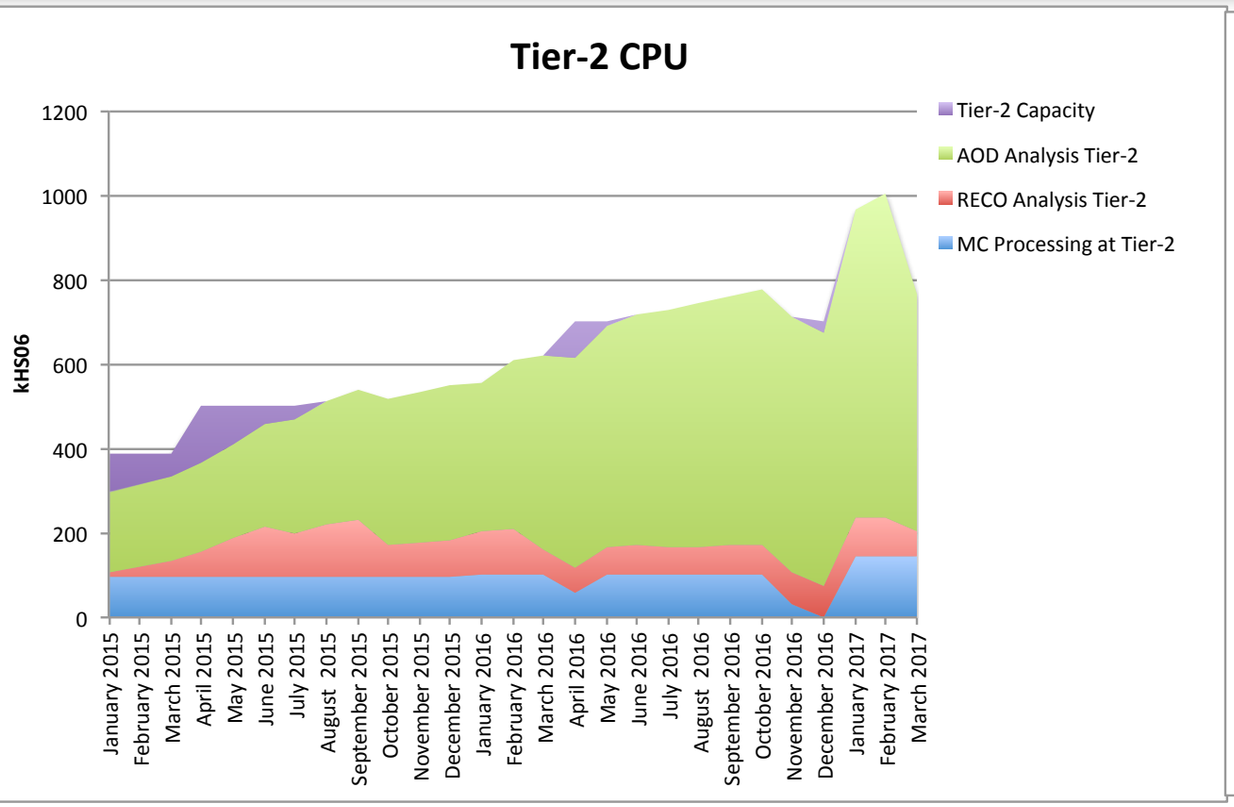
Last chance at new physics until….

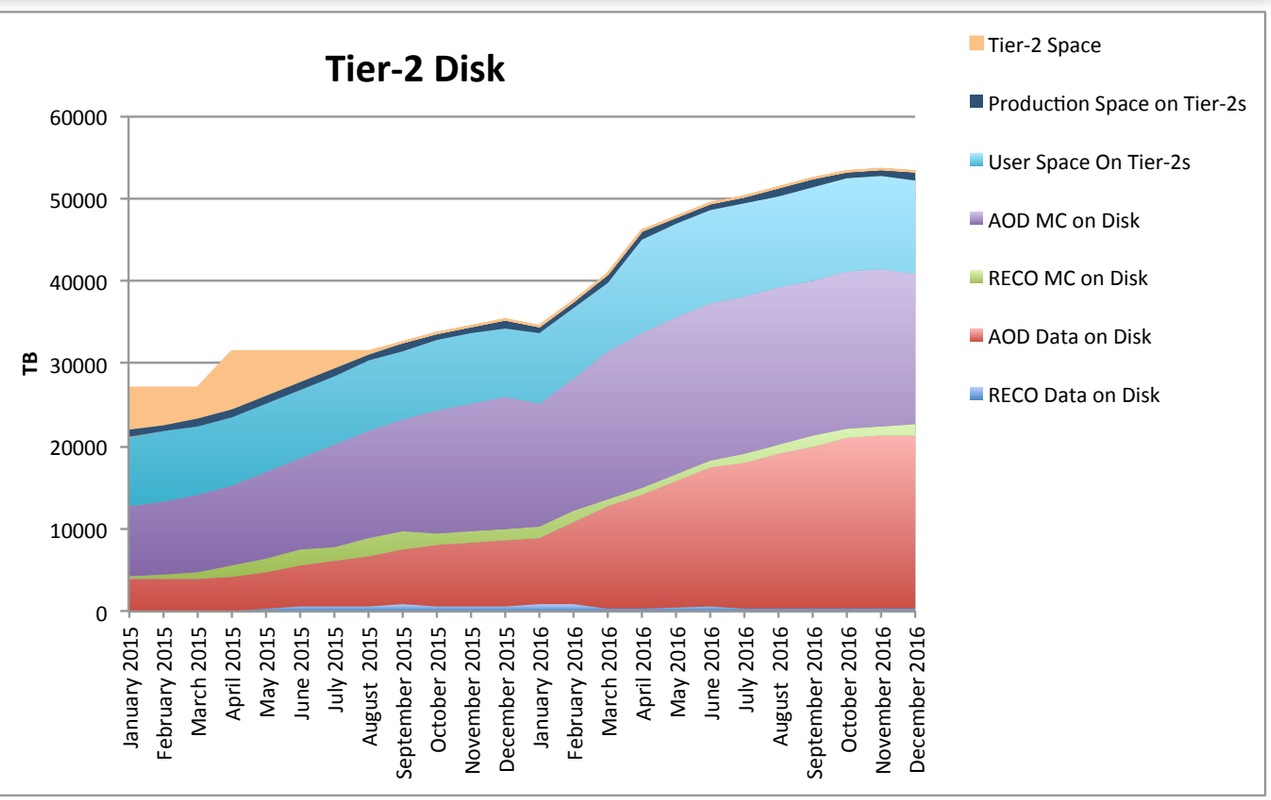




- We could see new physics in the first year — or first months
- There is no next major step in collider energy on the horizon
- We must be ready and we must execute!

CMS resource needs





- Resource needs increase as LHC resumes operations and data accumulates once again

	2014	2015	2016
CPU	+14%	+25%	+40%
Disk	+4%	+16%	+20%

- Year-to-year increases relatively moderate at T2, as we assume they function as data caches, older data doesn't stay around long
  - Note: very different at T1 — ~70% CPU increase in 2015, ~30% disk increases in 2015 and 2016

US T2 deployment plan

- Currently expect to maintain build-to-cost model with $250K/site/year. This should be more than sufficient to meet CMS needs.
  - We continue to be CPU-rich, already have deployed far beyond the ~20 kHS06/site needed by CMS in 2016
  - Cost of disk drives what we can deploy
    - Have used the LHC shutdown to invest in disk, with the goal of getting ahead of the CMS resource needs
      - Sites added "extra" 0.5 PB last year and will add 0.5 PB this year
    - This will allow us to deploy 1 PB/site that is not pledged to CMS but is available to US CMS users through /store/user etc.
      - Can be accessed via AAA, space management needs development effort
    - To maintain the extra 1 PB/site and meet CMS request, need to deploy *total* of 2.2 PB/site in 2015 and 2.4 PB/site in 2016 — quite do-able!
- Likely deployment strategy is then to add disk as described above and beyond that buy as much CPU as possible

Contingency (or lack thereof)

- But: the CMS computing resource request is designed to fit into a flat-ish budget, under the assumption that we can take the operational measures to get the job done with resources available
- This leads to an operational plan with ~no contingency!
  - Example: end-of-year reprocessing of that year's data can only be completed in time with HLT, which is only available in a particular time window → all reprocessing ingredients must be on time
    - HLT is as big as the entirety of CMS T1!
- CMS will be eager to have access to opportunistic resources
- Your site can make a difference!
  - Now more than ever, we need your help in finding extra resources on your campus that can be incorporated when CMS needs them
  - Have had some notable successes here already, but we need more

Operations

- For Run 2, we will be operating a much more flexible system of distributed facilities that will help meet the computing needs
  - Different tiers are becoming more similar in functionality
  - Wide-area data access — AAA — is a key enabling technology
- Many examples of new operational modes and capabilities
  - Prompt reconstruction at T1, use of HLT during shutdowns (and between fills?), global scheduling/prioritization through glideinWMS, improved user analysis tools (CRAB3), dynamic data management
  - Oli, Tapas and Christoph will discuss some of these today
- Thanks to tools such as AAA, Parrot, and CVMFS, CMS is in a much stronger position than ever to make use of any opportunistic resources that sites can provide to us, and we might well need them
  - E.g. if we miss the HLT window, do re-processing at T2 sites?

Milestones for 2014

- CMS computing has an extensive set of milestones for this year; US T2 milestones are aligned with the major ones of CMS
  - 1 April: Deployment of T2 computing resources sufficient to meet CMS resource request for 2014
    - Timed to start of WLCG resource year; we have done this
  - 30 April: Achieve full participation of T2 sites in CMS data federation
    - Also done, of course
  - 30 June: Demonstrate appropriate fraction of full production scale of CRAB3 distributed analysis tools at T2 sites: 5K cores, 50K jobs/day, average 80 users/week
    - The numbers are 25% of the total CMS goal
  - 31 October: Complete participation in exercise of full system for organized production: 5K cores at T2 sites for simulation
    - Also 25% of the total CMS goal
- All of these milestones should be readily achievable

Technology advances

- One of the reasons the US T2 sites are so successful in CMS is that we try to stay at the leading edge of technology
  - US T2's are always the first to try anything new in CMS computing
- Today each site will be discussing how they are getting ready for the future, in terms of technologies and otherwise
- Brian will be offering us some challenges for us to meet over the next few months
- And we'll hear about specific projects underway at sites that will be of interest to everyone:
  - Jeff on automatic failover inside FUSE
  - Manoj on experience with whole-node scheduling
- We're always happy to put such a presentation into a monthly US T2 meeting, too

Other technical issues on the horizon

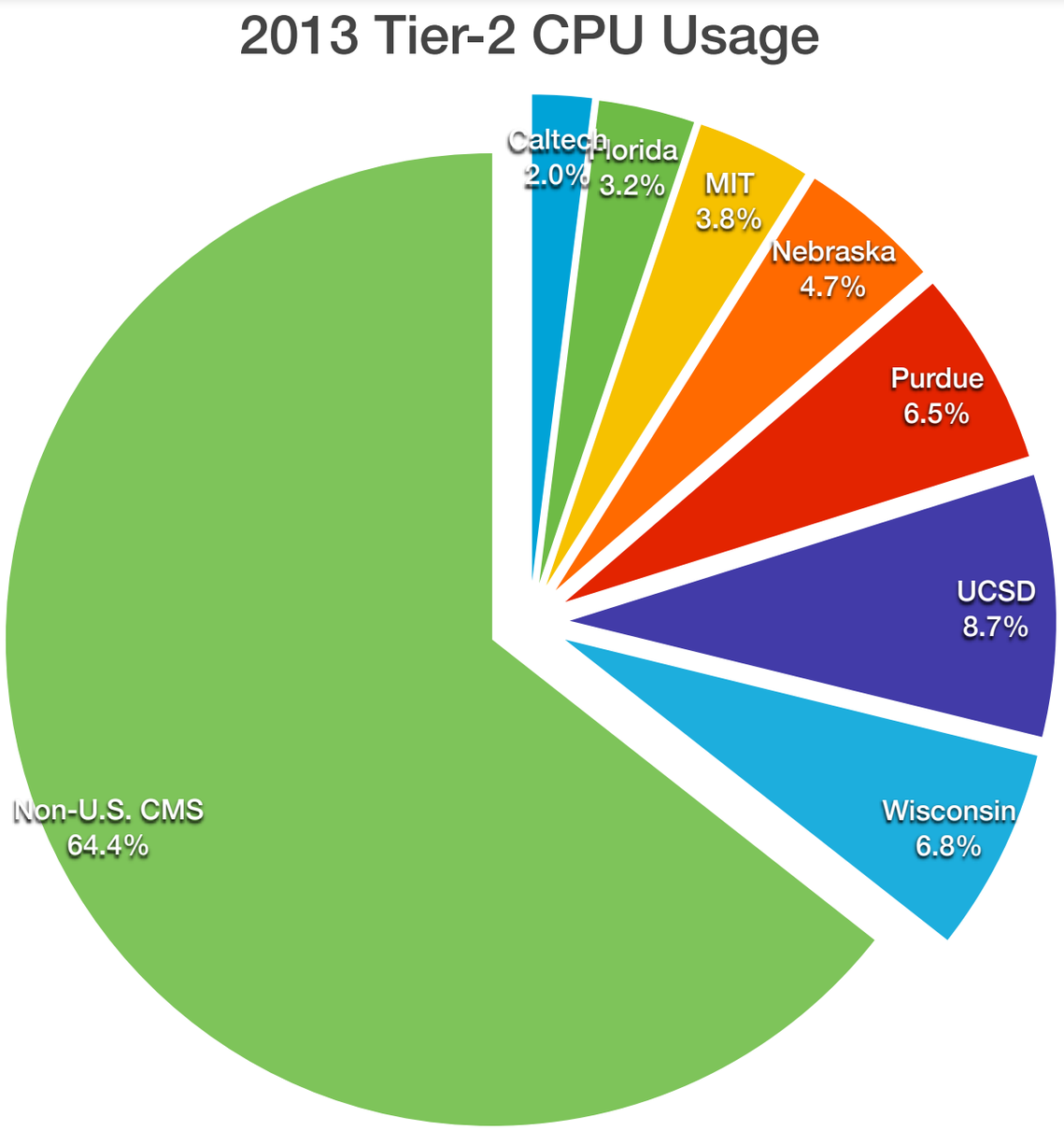
- Had hoped for a presentation on experiences with 100 Gbit networking today, but we are not quite ready yet
  - Thanks to sites that are starting to explore this
  - To my knowledge, almost every site has a plan to get to 100 Gbit
  - This may soon become our baseline network connectivity!
- We now have the technology to give priority to "national groups" of users on suitable portions of sites
  - We first talked about this idea at our 2010 workshop at FNAL!
  - US users should have priority on compute resources above our pledge to CMS
  - However, this apparently needs a lot of work on the part of sites to partition the cluster appropriately
  - I am not sure that it is worth it, but it might be necessary in case of a resource crunch

Notes on budgets

- The US CMS Operations Program had its annual external review three weeks ago, and as usual it went well for the program as a whole and for software and computing in particular
- This is the agencies' annual opportunity to reveal budget guidance for the next few years — may not be the actual funding, but a sense of what we need to plan for
- Scenarios offered were either flat-flat or a 5% cut
  - Neither one has pretty outcomes, and a budget reduction could endanger the program as a whole
  - We could see a shifting of priorities away from the LHC this year….
- Even though it is agreed that T2 sites are a good investment, reducing the T2 hardware budget is an easy cut to make
  - We have strongly articulated the consequences of reducing site staff or reducing the number of sites that we operate

Value

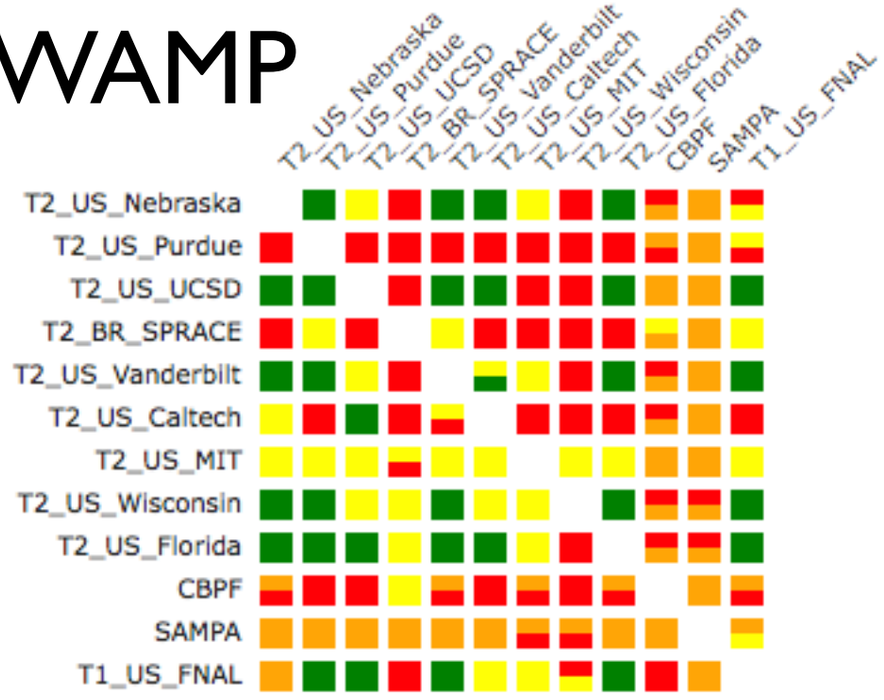
- The 2012-13 T2 cost/benefit review gave a set of metrics that could be used to evaluate the value delivered by sites
  - Sites in fact gave a substantial return on investment thanks to the participating universities
- Should we need to reduce hardware purchases, we would use those metrics to decide how to allocate funds to sites that provided the most value
- It goes without saying that we want every site to deliver as much value as possible to CMS



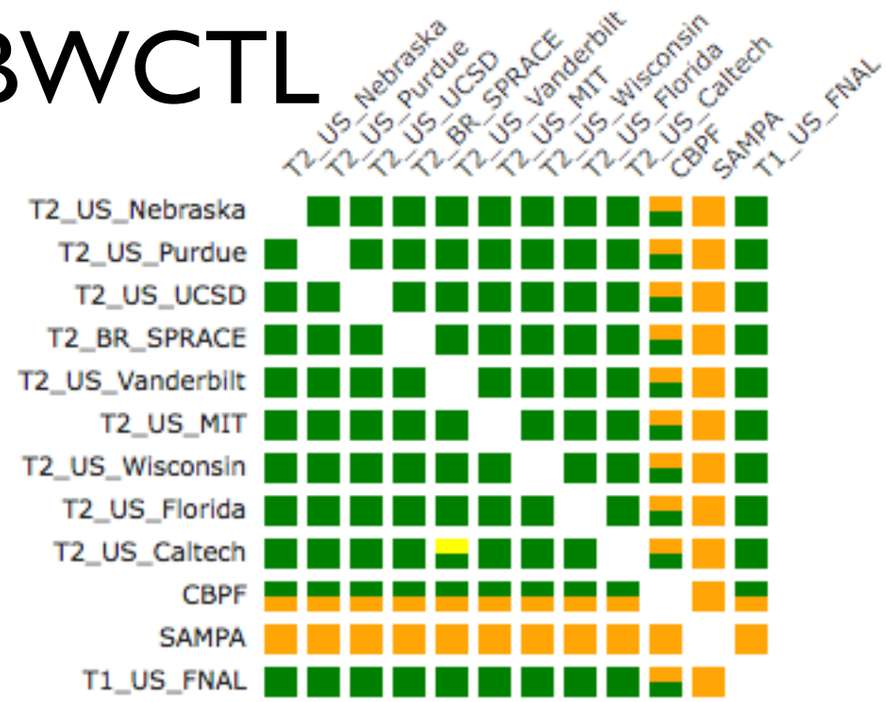
One measure of value; others include quality of site support, contributions to development/operations….

Near-term notes

OWAMP



BWCTL



- These perfSONAR tests look a lot better — thanks!
  - Remember, orange (no data) is worse than red (< 0.5 Gbps)
  - Garhan will give us some pointers on pS tests today
  - (Need to get rid of those sites that aren't ours….)
- We are looking for some sites to test out more pervasive monitoring of CMS disk usage, see here for details

Outlook

- The excellent quality of work performed at the US CMS Tier-2 sites — over the past nine years! — is a key driver of the successful CMS physics program
- Discovering the Higgs was exciting, but there is a lot of physics potential in run that starts a year from today and we must be ready for it:
  - CMS will need all the computing that we can throw at it
  - We need to respond to operational changes
  - We must continue to modernize our facilities
  - All while operating the sites at the highest level of service
- I'm looking forward to working with you as we get ready for the next era of the LHC
  - Even if it means that I have 9 PM meetings — only three more of those until I come home!