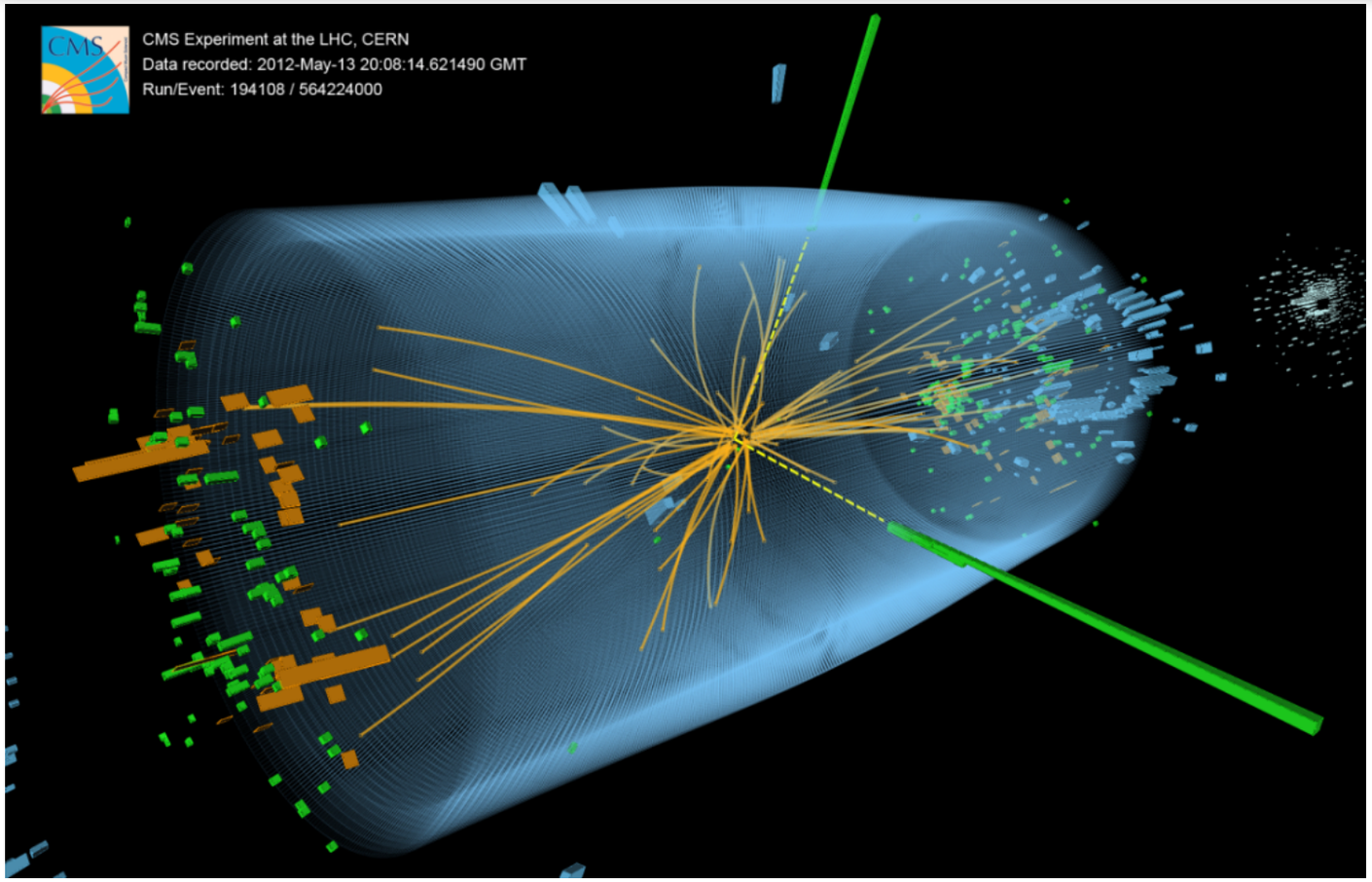
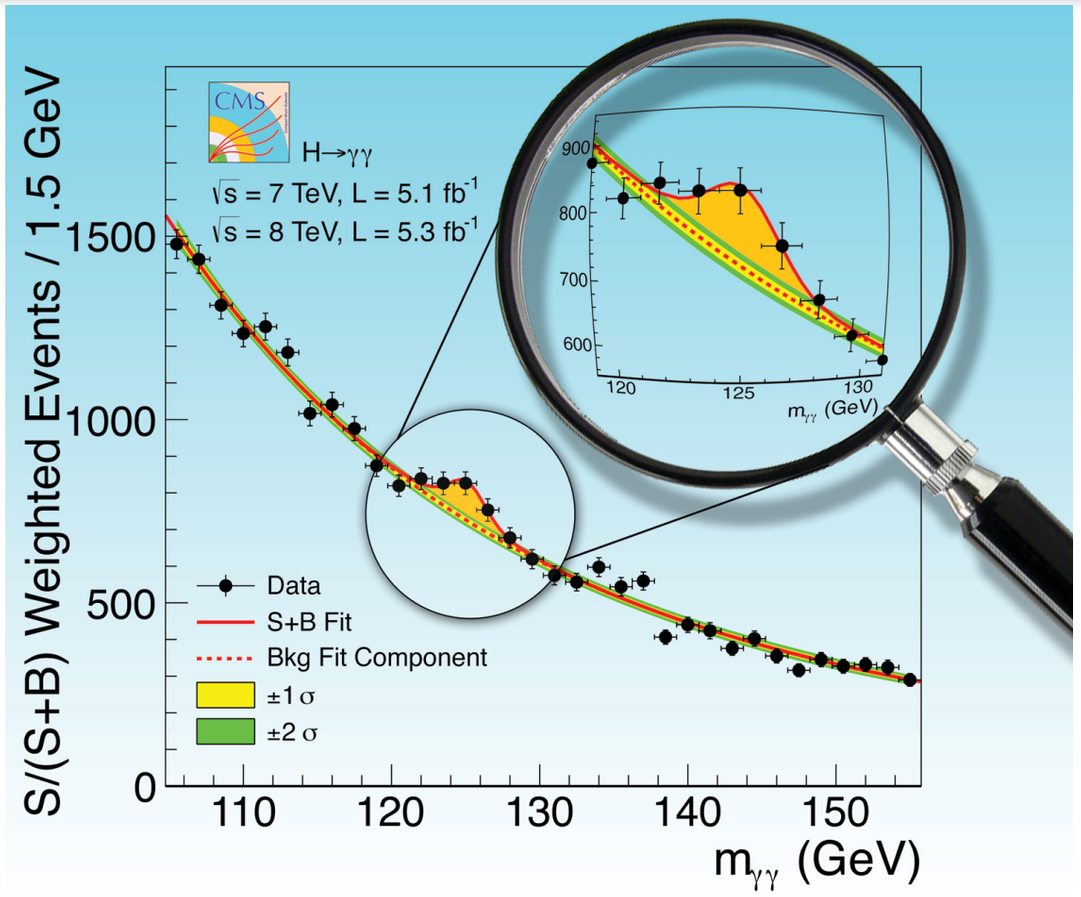
US CMS Tier-2 Status and Future

Ken Bloom

March 11, 2013

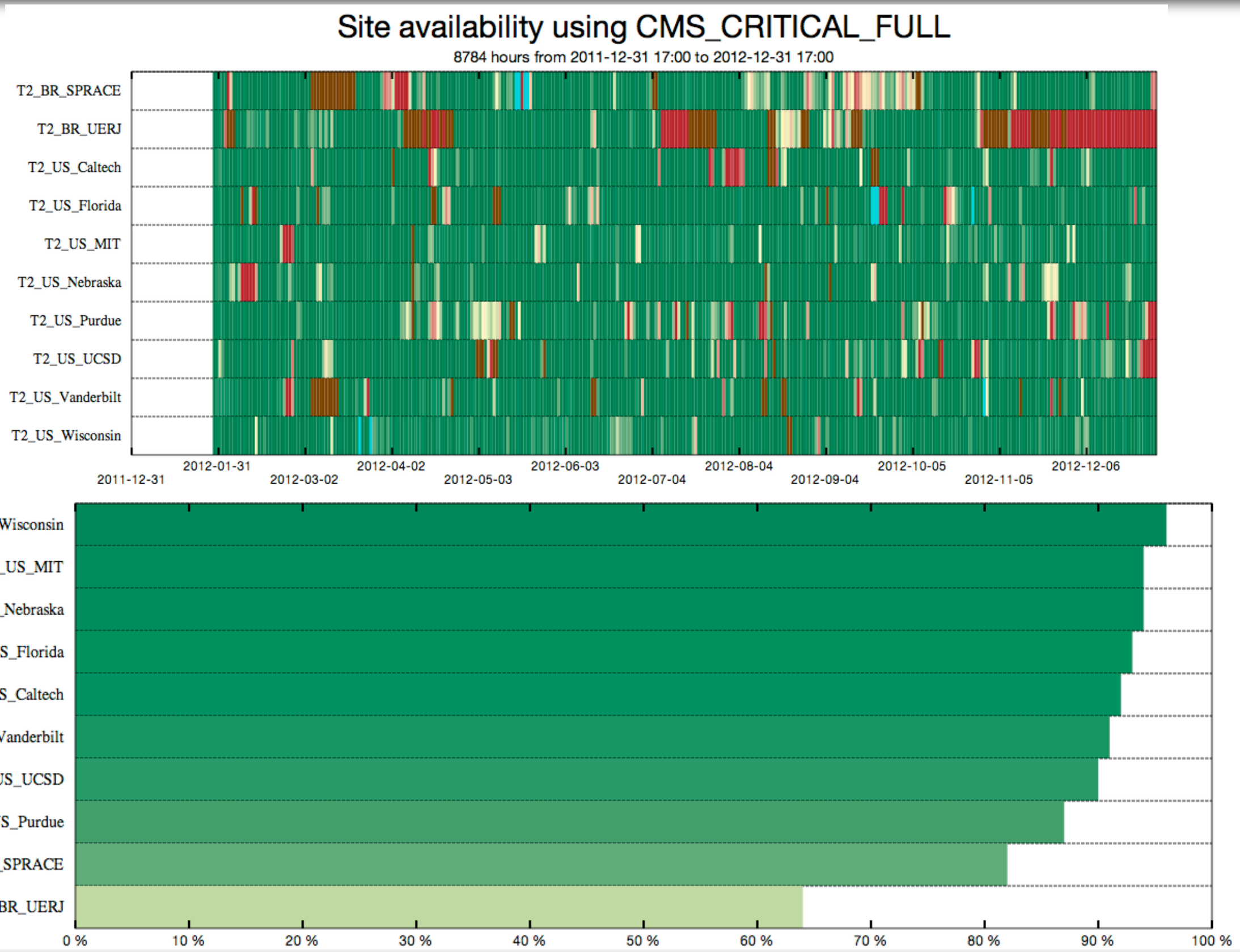
We made history



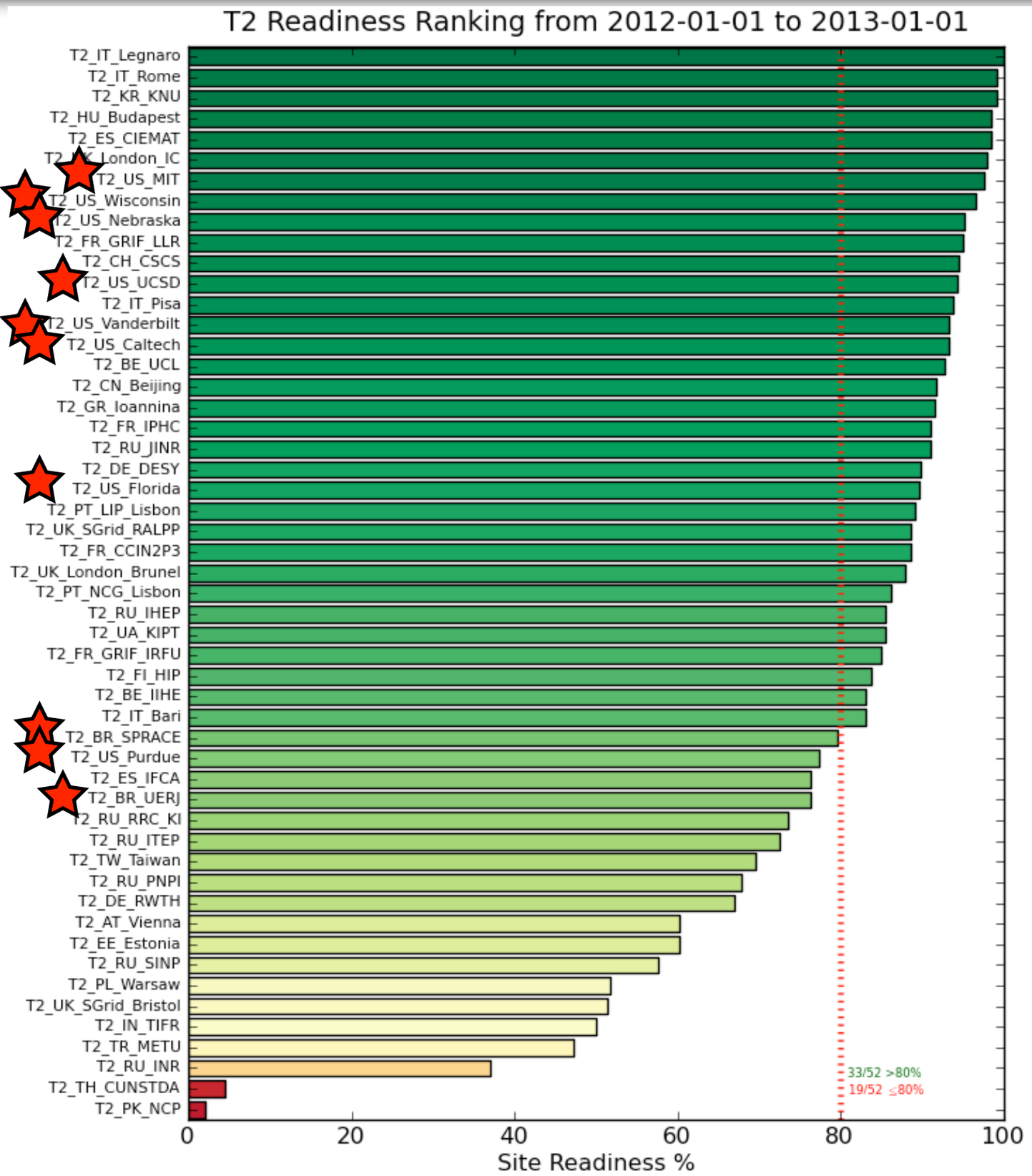


- We are making discoveries that are changing our understanding of physics through work done at Tier-2 sites
  - Be proud of it!
- The US sites are among the biggest and most successful in CMS, and the experiment would not be successful without us

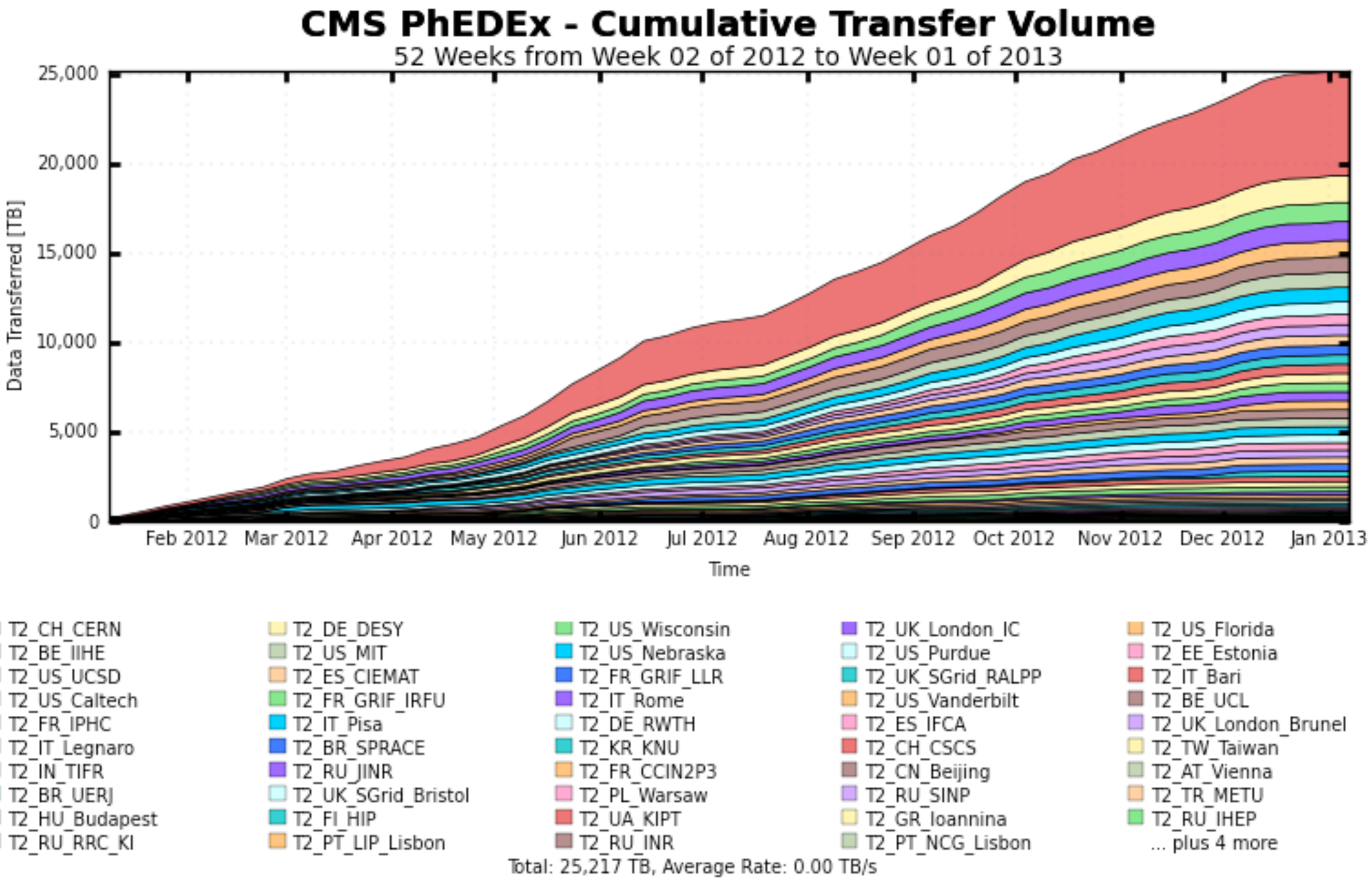
2012 in review: SAM



2012 in review: site readiness

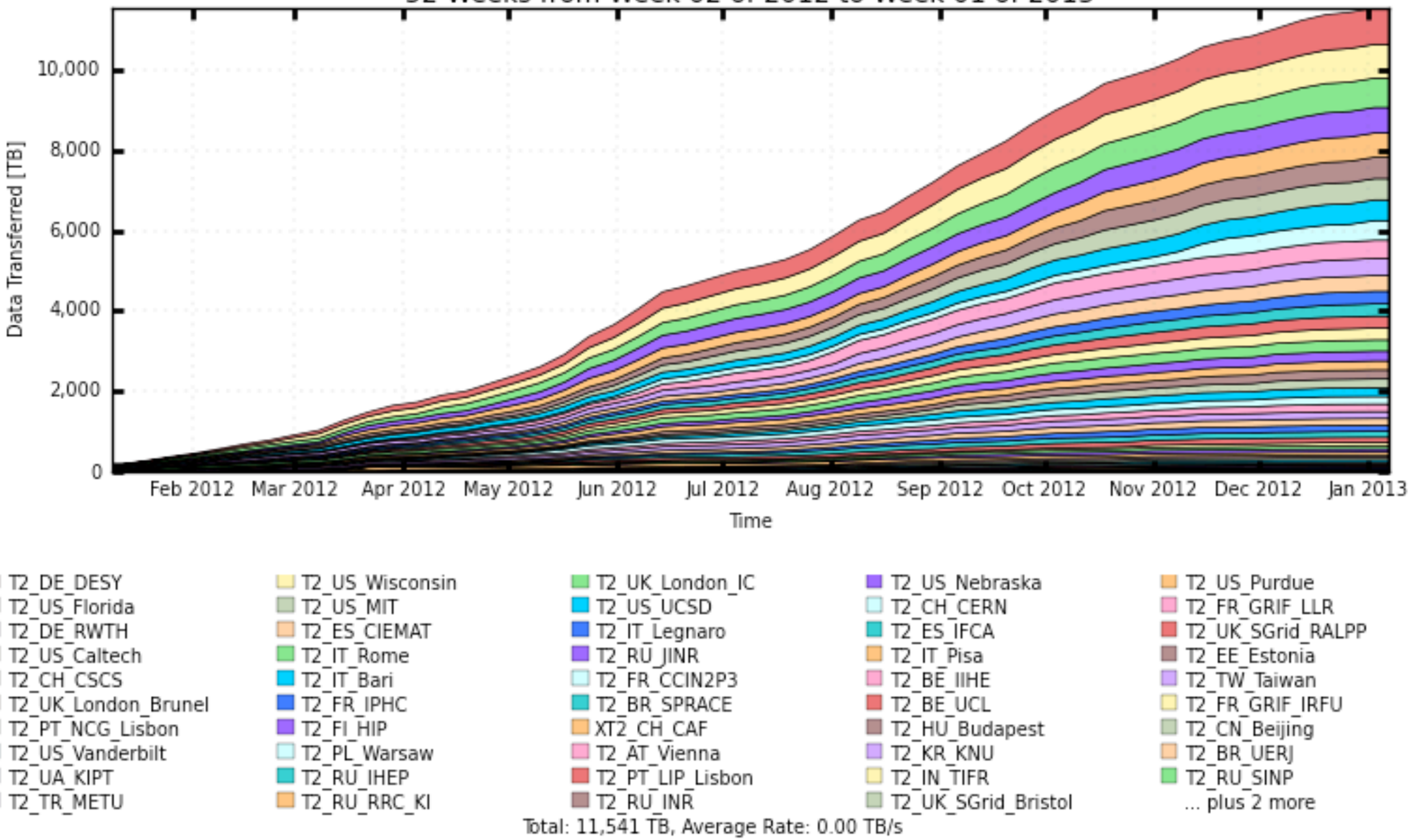


2012 in review: inbound transfers



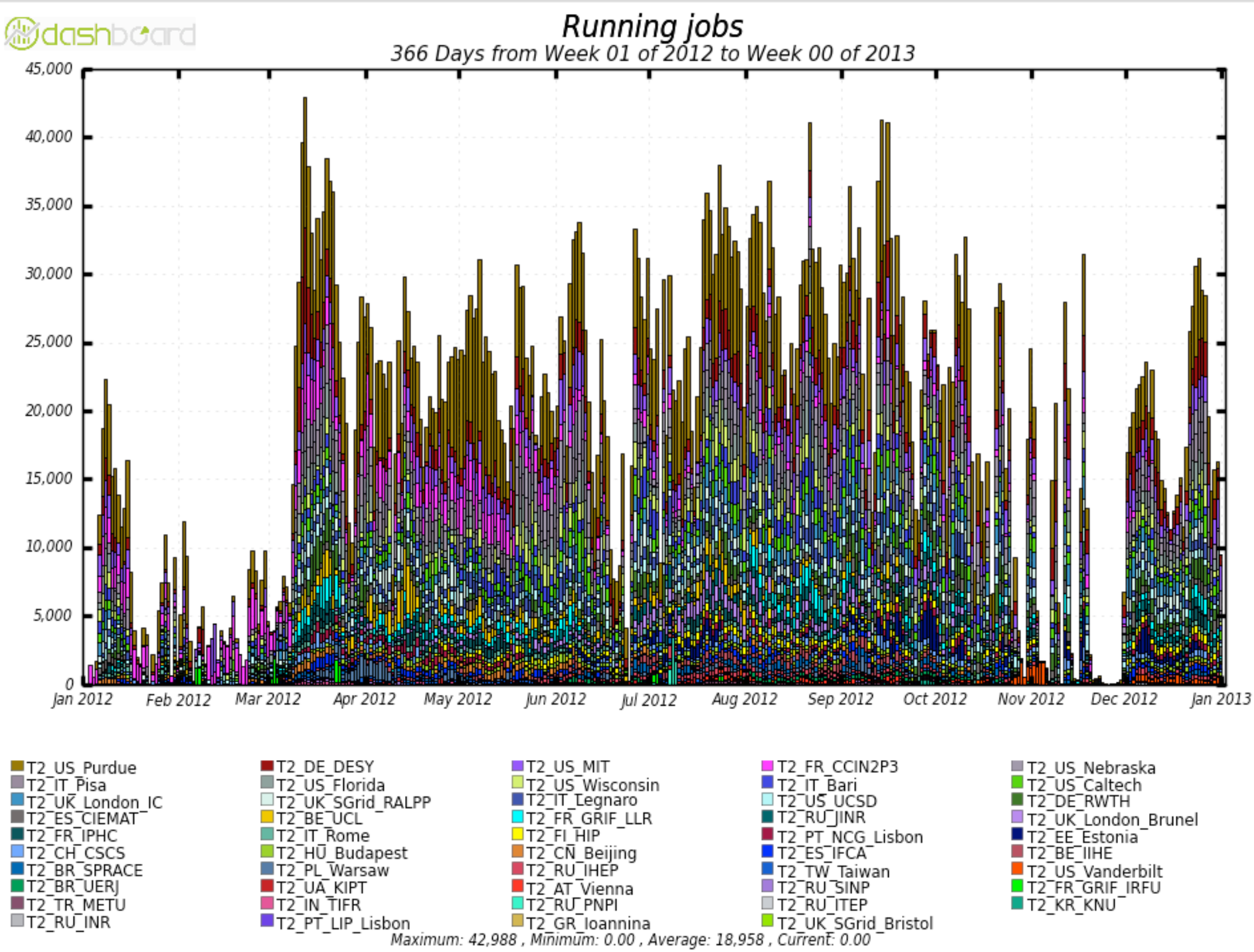
- 25.2 PB into all T2's, 6.4 PB into US/BR T2's

2012 in review: outbound transfers



- 11.5 PB from all T2's, 4.2 PB from US/BR T2's

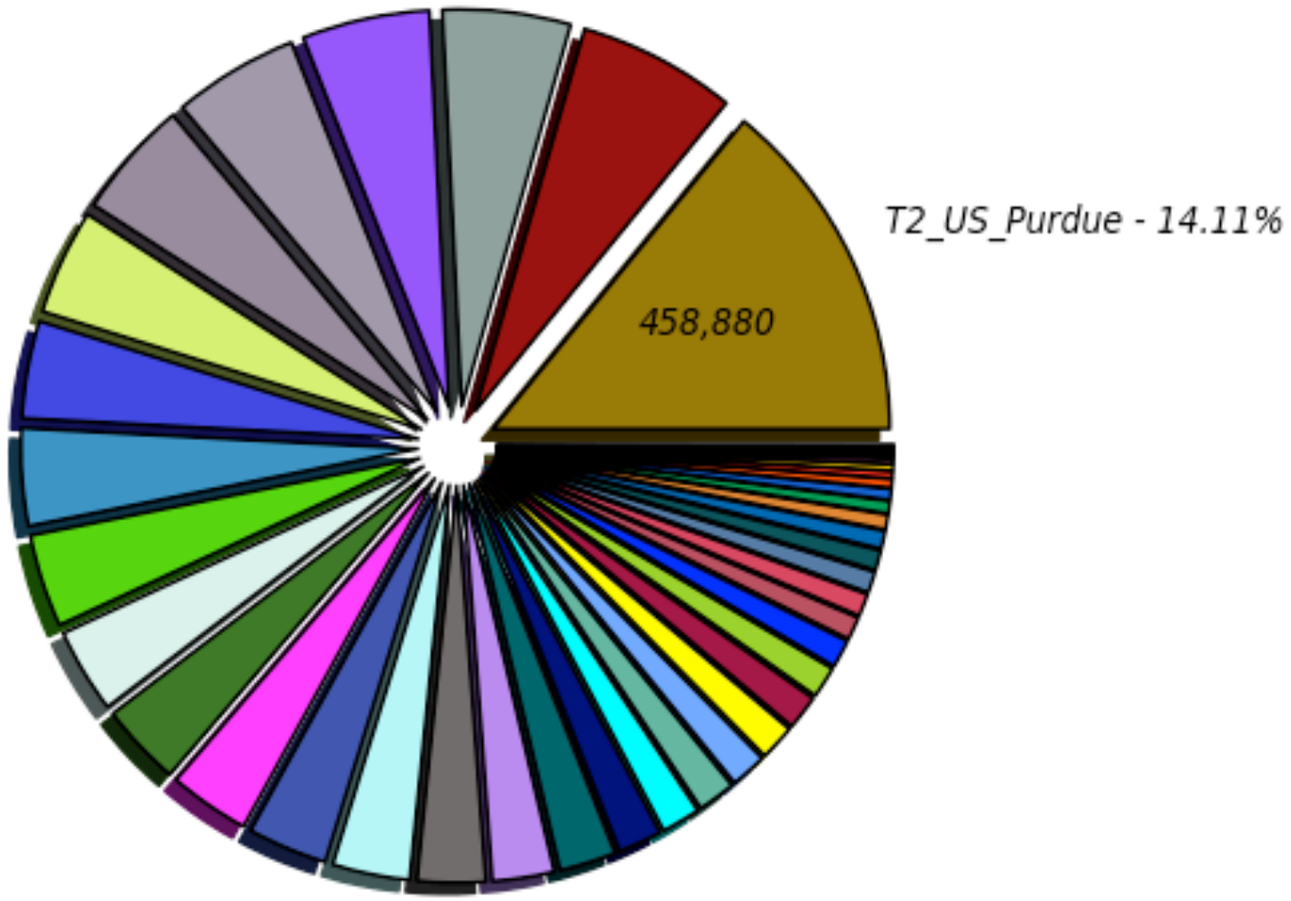
2012 in review: production



2012 in review: production

dashboard

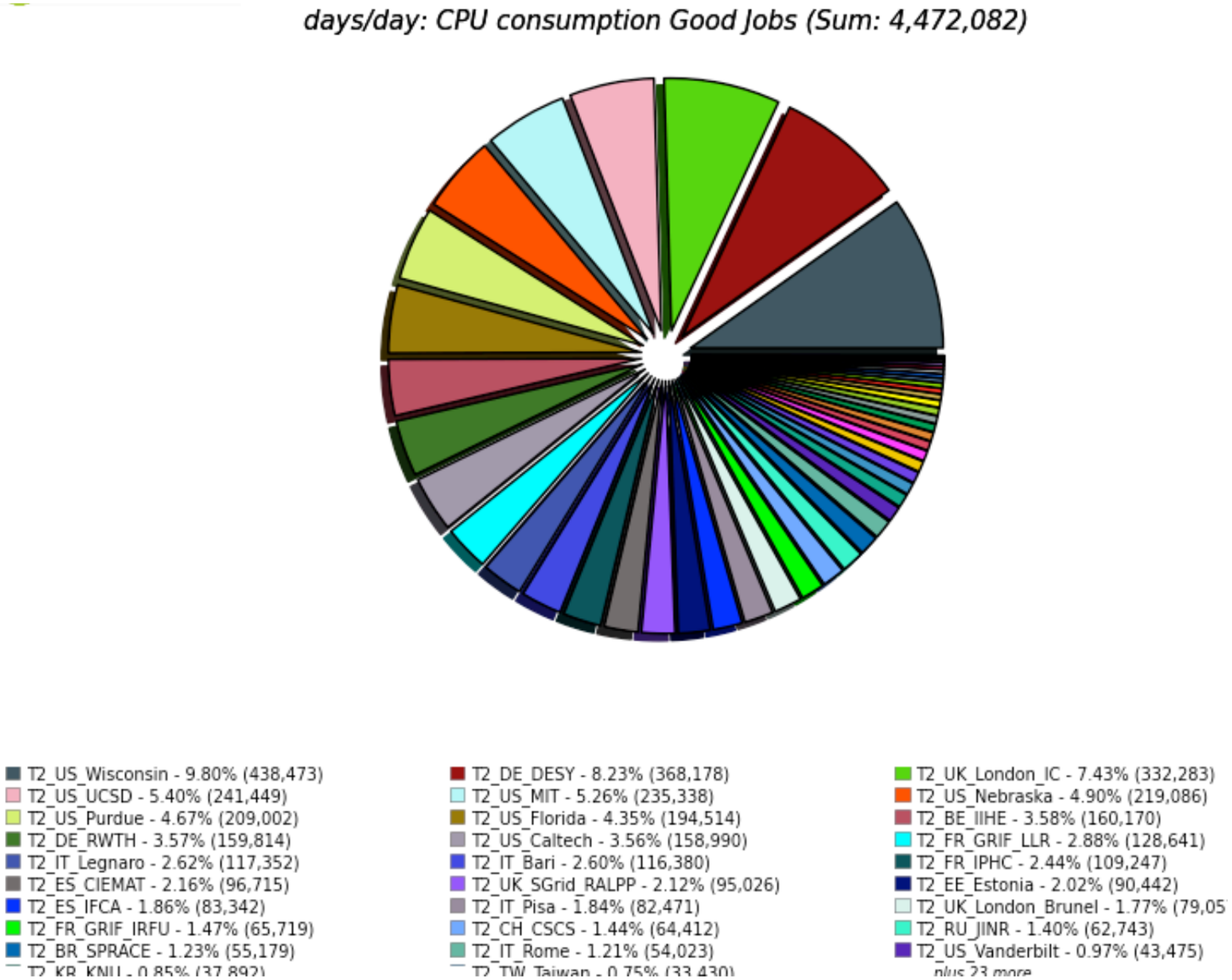
days/day: CPU consumption Good Jobs (Sum: 3,252,777)

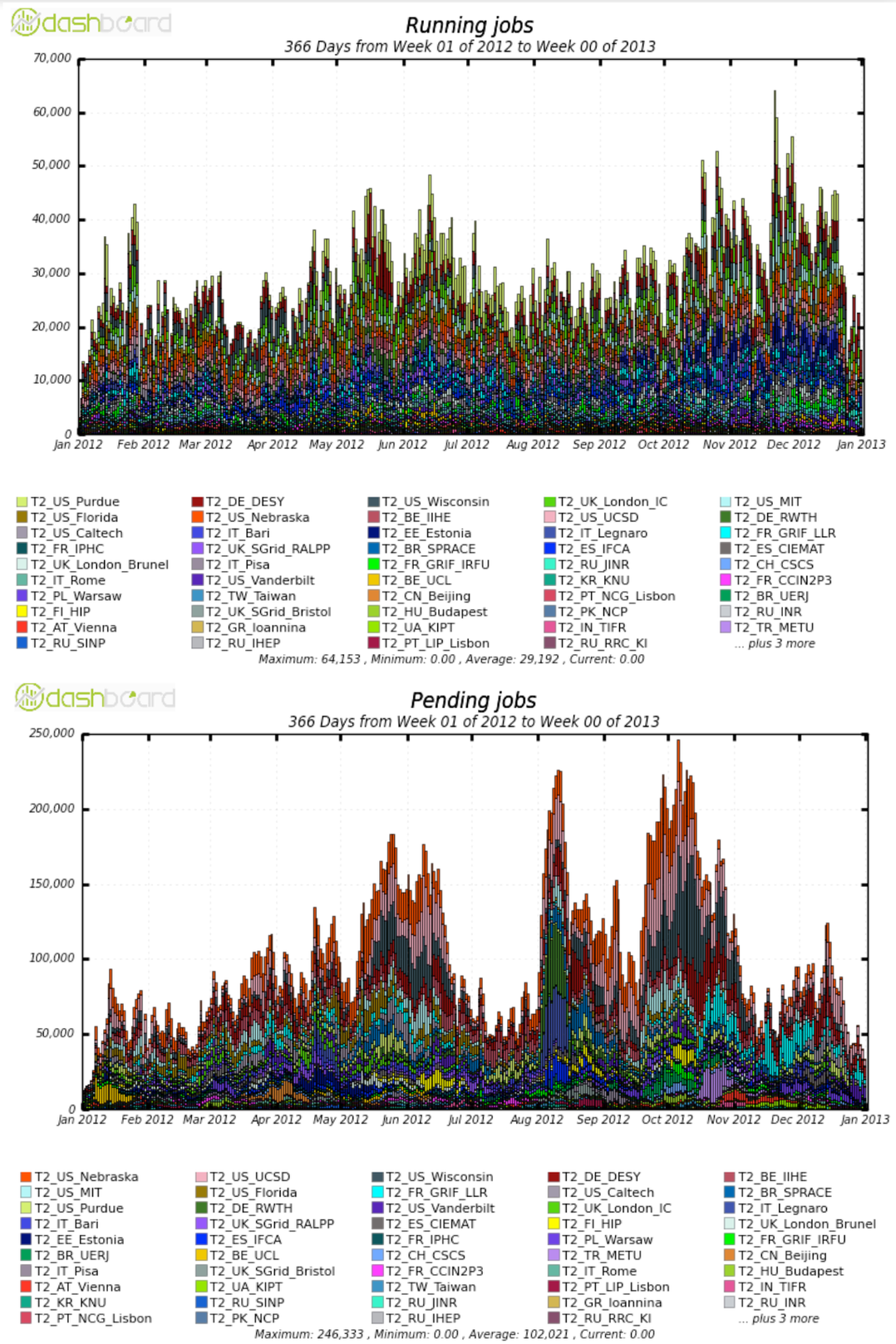


- T2_US_Purdue - 14.11% (458,880)
- T2_US_MIT - 5.20% (169,131)
- T2_US_Wisconsin - 4.30% (139,775)
- T2_US_Caltech - 3.78% (122,935)
- T2_FR_CCIN2P3 - 3.29% (107,138)
- T2_ES_CIEMAT - 2.77% (90,119)
- T2_EE_Estonia - 1.78% (57,952)
- T2_CH_CSCS - 1.47% (47,972)
- T2_HU_Budapest - 1.28% (41,681)
- T2_RU_IHEP - 0.91% (29,700)
- T2_DE_DESY - 6.46% (210,288)
- T2_US_Nebraska - 5.11% (166,335)
- T2_IT_Bari - 3.96% (128,846)
- T2_UK_SGrid_RALPP - 3.45% (112,303)
- T2_IT_Legnaro - 3.25% (105,756)
- T2_UK_London_Brunel - 2.52% (81,854)
- T2_FR_GRIF_LLR - 1.58% (51,544)
- T2_FI_HIP - 1.46% (47,383)
- T2_ES_IFCA - 1.13% (36,871)
- T2_PL_Warsaw - 0.86% (27,898)
- T2_US_Florida - 5.21% (169,529)
- T2_IT_Pisa - 4.81% (156,548)
- T2_UK_London_IC - 3.95% (128,553)
- T2_DE_RWTH - 3.35% (109,020)
- T2_US_UCSD - 3.15% (102,477)
- T2_RU_JINR - 2.31% (75,115)
- T2_IT_Rome - 1.55% (50,291)
- T2_PT_NCG_Lisbon - 1.42% (46,079)
- T2_BE_IIHE - 0.93% (30,400)

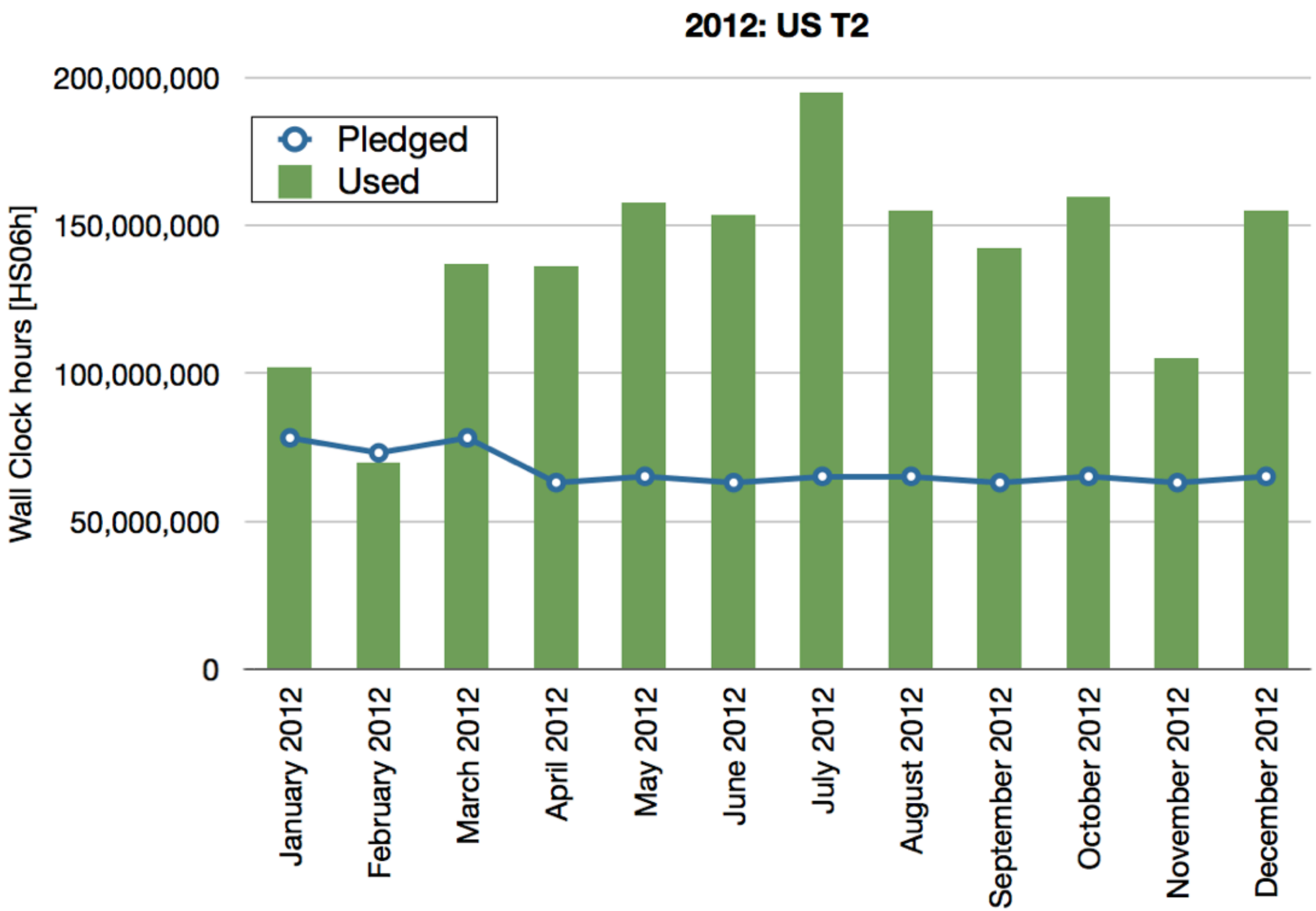
plus 20 more

2012 in review: analysis





2012 in review: broader context



- US sites have provided well in excess of pledged CPU

Looking ahead: operations

- We are trying to wring as much out of resources as possible
  - The "pending jobs" plot suggests inefficient usage
  - We want to use opportunistic resources aggressively
- Thus, computing operations is trying to maximize flexibility, especially in where various workflows can be run
- AAA is a very important piece of this -- removes data locality requirement, can run jobs anywhere
  - Early example: now running MC digi-reco at US T2 sites, reading GEN-SIM datasets over the WAN, pileup locally
- Also developing tools to integrate opportunistic resources on the fly as transparently as possible
  - We could use help from sites in identifying campus resources that we can have opportunistic access to
- We will need sites to stay up to date with various upgrades

Looking ahead: operations

Site	OSG3 CE	OSG3 Workers	OSG3 Aux	GUMS	CVMFS	HDFS	HS06 Spreadsheet	DigiCert	SL6	Register PerfSONAR	Enable t1production role	Enable file monitoring	Last Update
T2_US_Caltech	✓	✓	✗	✗	✓	✗	✓	✗	✗	✓	✓	✓	March 7
T2_US_Florida	✓	✓	✓	partial	partial	N/A	✓	✗	✓	✓	✓	✓	March 8, 2013
T2_US_MIT	✓	✓	✓	✓	✗	✗	✓	✗	✗	✓	✓	✓	March 6, 2013
T2_US_Nebraska	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	March 6, 2013
T2_US_Purdue	✓	✓	✓	✓	✓	✗	✓	✗	✓	✓	✓	Will enable when network operation at site becomes stable	
T2_US_UCSD	✗	✓	✗	✗	✗	✗	✓	✗	Partial	✓	✓	✓	3/6/13
T2_US_Wisconsin	✓	✓	✓	✓	✓	✗	✓	Partial	✗	✓	✓	✓	
T2_US_Vanderbilt	✓	✓	Partial	✓	✓	N/A	✓	Partial	✓	✓	N/A	✓	March 6, 2013

- There will be more to come, e.g. multi-core queues or rsl statements to request multi-core job slots
- And, must operate new stuff well:
  - xrootd deployment should be robust, will be used more heavily
  - xrootd fallback should be activated and working well
  - CVMFS, glexec are needed

Looking ahead: operations

- More items from operations:
  - With pileup files being read locally for redigi at T2's, they need to be replicated sufficiently for heavy access
  - t1production and production roles should both be enabled and mapped to the "production" fraction of resources
  - Sites should support 48-hour jobs and have BDII report it correctly
  - Ops would like to know of any special queues that exist (large memory, very long jobs, etc.) that they could make use of
  - Be sure there is always enough free disk space for operations
  - Monitor pilots for black holes, other problems

Looking ahead: deployment

- The 2013 pledge to CMS is the same as for 2012
  - 12,500 HS06 and 1000 TB per site
  - We provide much more CPU and about as much disk as pledged
- The 2014 pledge is expected to be the same, too
  - LHC not running this year, no new data, not as much to do, have access to CERN resources for analysis

- The LHC will resume operating in earnest in 2015
  - Anticipating that trigger rates will grow by x2.5, and increased pileup will lead to x2.5 in reconstruction time
  - But T2's are relatively insulated from this -- assume that we do not keep any Run 1 data at T2, just the new data
  - Expect to need "only" ~30% increase in CPU, ~15% increase in disk from current levels over the next two years

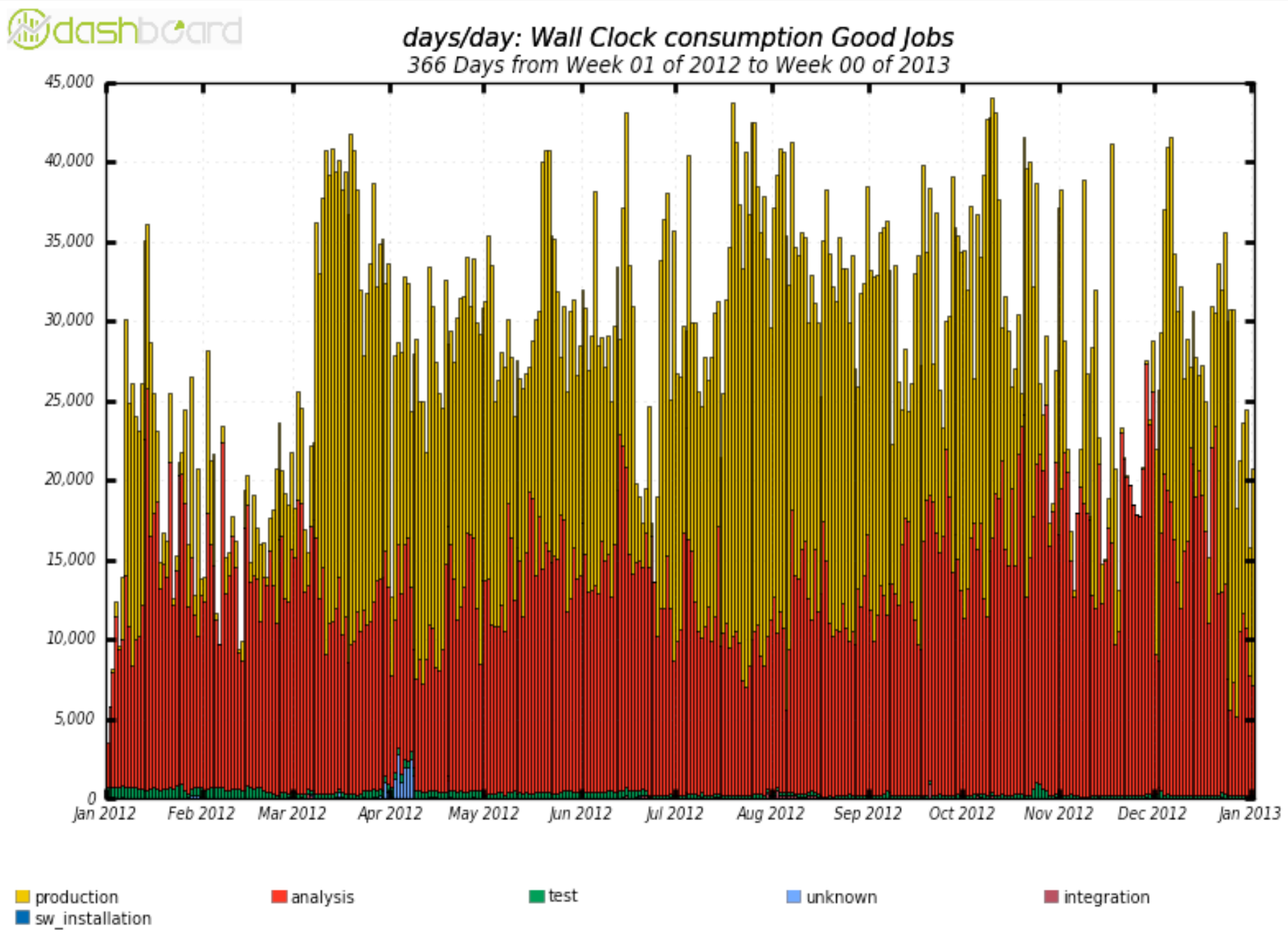
Looking ahead: deployment

- At the same time, we are anticipating that US T2 sites will remain "build to cost" at the current level of funding
  - Should be easy to deliver in excess of pledge, in both CPU and disk
  - We can think creatively about what to do above pledge
  - During this period, less MC production, so focus on analysis?
- One school of thought: focus on CPU
  - Look at the number of pending jobs!
  - Thanks to AAA, we can use the CPU to read data from anywhere
- Another school of thought: focus on disk
  - The pending jobs are just site mismatch inefficiency; fix with AAA
  - Data can be anywhere, but it must be somewhere
  - Users often complain of lack of disk space; thanks to AAA and networks, we can make remote disk more usable for them

Looking ahead: deployment

- I am leaning towards thinking "more disk" is the better answer, as it is the "harder" one to do, but I haven't decided
- We have time to think -- we do not have the 2013 funds in hand yet, and sites might want to wait to get better pricing
- We very much welcome discussion on this!
  - What is your instinct based on observed use patterns at the site?
  - Do you think you can get better pricing with a tilt towards "more CPU" or "more disk"?
  - Do you think you can better operate one or the other?
- We're not a priori opposed to non-uniform deployments across sites, i.e. some sites could focus on buying CPU and others on buying disk
  - Need metrics to capture the utility of disk....

Why we don't need more CPU?



- Analysis doesn't saturate T2 CPU when there is no production?
  - Although that is too simplistic an argument....

T2 cost/benefit review

- Once again, thanks for everyone's help with this!
- Caveat: this is not my report; I am just relating what's in there

- In general, I consider the review report to be quite positive:
  - All sites performing well, outstanding among CMS T2s
  - Provide much more resources than pledged, in part by taking good advantage of university partnerships
  - Good support for physics groups and physicists
  - Much more involved in the development of CMS computing than a typical T2 site -- leading CMS in developing new computing model
  - Generally not hitting infrastructure limits
  - US T2s contribute to US leadership in physics analysis
  - Every dollar spent on site expansion is well spent

My thoughts on recommendations

- Consider a purchasing program for T2 sites not involved in existing "community programs," weighing cost savings against lack of flexibility and administrative effort
  - An interesting idea, but I think implementation would be difficult
- Continue and expand efforts to leverage opportunistic resources, and make information about them available to CMS
  - Already a priority for CMS and US CMS computing
  - Again, we could use help from sites on this
- Look into distributing the user community better across sites; whole package of user facilities offered by a site is important for this (e.g. UCSD, MIT are popular)
  - As an administrative matter, this should not be hard, and with AAA, it should be even easier for users to make good use of T2s, even without login privileges etc.

My thoughts on recommendations

- Consider implementing policies to achieve more a uniform approach to user support within and across T2 sites
  - Not totally sure what this means, but it raises a long-standing concern for me, which is giving preferential login access to certain users. We must be very careful about appearances of favoritism!
  - Sites that do have such policies should consider why they do and whether it is necessary given current computing tools
- Stay on forefront of computing evolution, as it benefits CMS and local users -- an opportunity to increase US CMS influence

Review conclusions

- All sites are doing well, above expectations, tops in CMS. NSF support has been critical. All receive substantial benefits from the institutions that host them, and the institutions think CMS computing is a worthwhile investment
- Provision of CPU and disk resources vary for many reasons, but each site brings something special and unique to CMS computing
- Many intangibles go into assessing the value of a site; putting everything together provides a rough grouping:
  - Purdue, Nebraska and Wisconsin are outstanding, provide significantly more than expected
  - Florida and UCSD are excellent, provide more than expected
  - Caltech and MIT are very good, provide about what is expected
- Excellent aggregate performance provides advantages to the US program in a number of ways

Moving forward

- Given this very positive report, we do not see any reason to make large changes to the US CMS T2 program
  - This could change if there were to be significant budget cuts, but none are foreseen at the moment
- But we would like to use this as an opportunity to:
  - Find more opportunistic resources on your campus -- a way to increase benefits without additional cost
  - Encourage sites to become more involved with the development of CMS computing -- your experience and skills are valuable
  - Remind everyone to be focused on site and user support
- Looking further ahead, it may make sense to deploy more equipment at sites that have good matching arrangements
- This year, US CMS wants to evaluate the distribution of resources across computing tiers -- could be good for T2?

Perspectives

- Eight years into the US CMS Tier-2 program, our successes have been given a strong external endorsement
  - US T2 sites acknowledged as a strategic strength for physics
  - It's all because of the excellent work being done at the sites, in both operations and development
- We are entering a period where we must focus on acquiring resources in preparation for the next run, and on being able to use all resources as flexibly as possible
  - There are many ways for sites to help out with this
- It continues to be a pleasure for me to work with all of you