

**U.S. CMS
University Facilities
Project Status**

J. Letts, K. Lannon
March 23, 2015

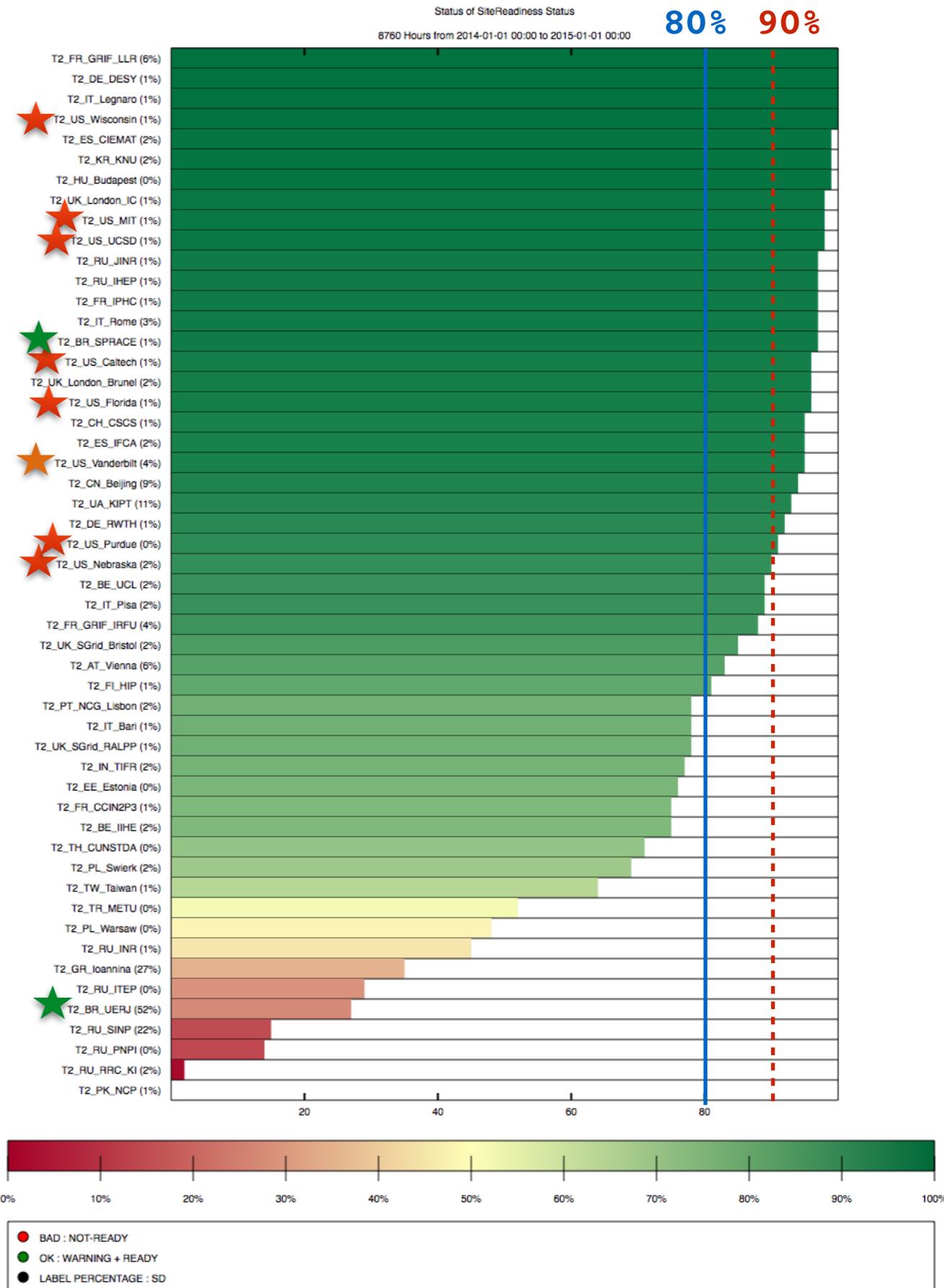
Outline

(in no particular order)

- **Success in 2014!**
- **Processing, Storage, and Networking**
- **CMS Resource Requests through 2016**
- **Budget Outlook (to meet those requests)**
- **Computing Model Evolution & Networking**
- **Program of the Day**

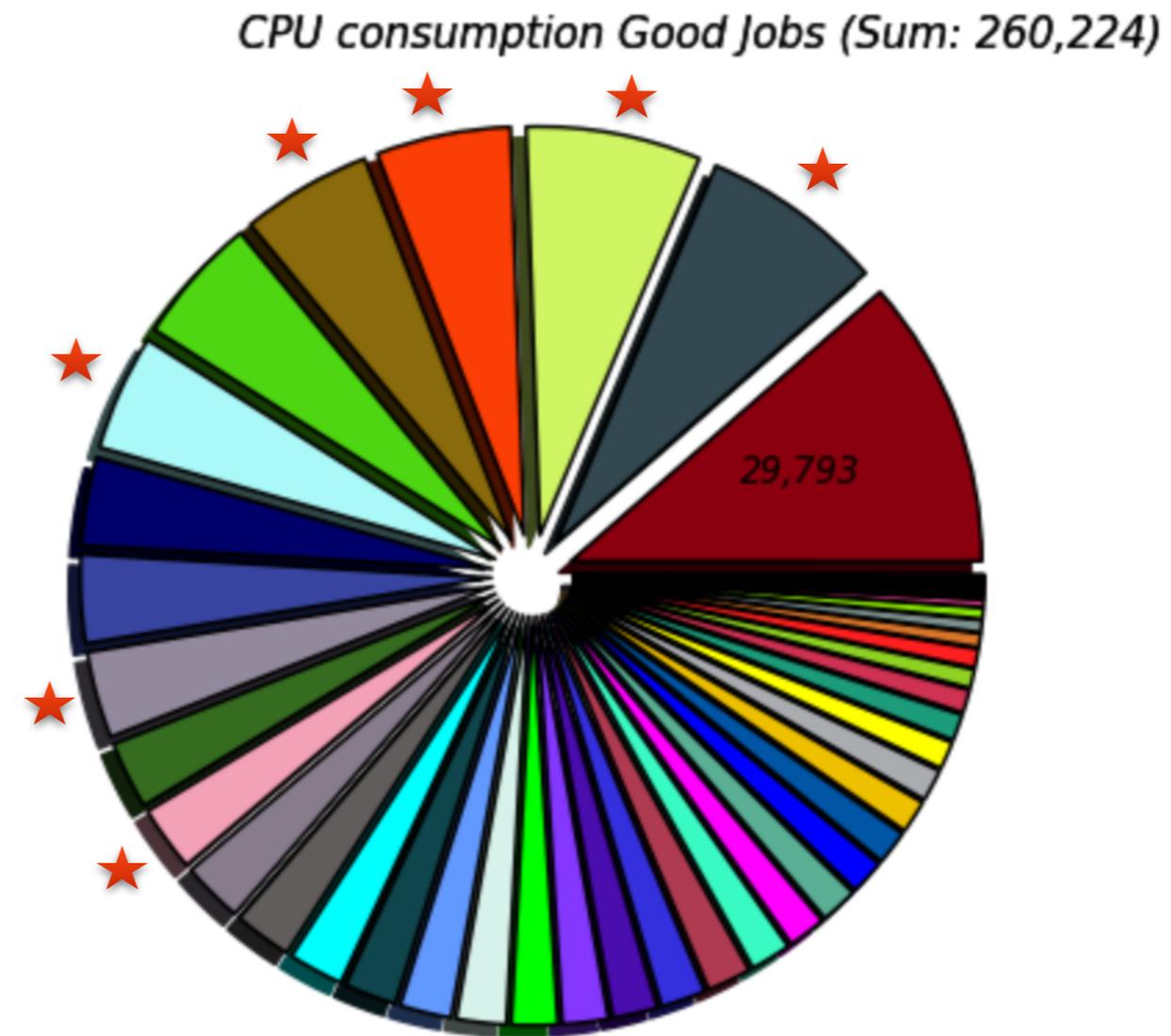
Success in 2014

- Site Readiness is one of the many baseline metrics that we keep an eye on.
- Tests functionality of various services at a site such as storage, Grid job submission, and data transfer.
- Allowances made for scheduled downtimes (%) and weekends.
- CMS goal is >80%. All 8 U.S. Tier-2 sites and SPRACE were >90% in 2014!



Success in 2014

- **The 7 HEP Tier-2 sites delivered 35% of good CPU time from all Tier-2 sites to CMS in 2014**
- **However, usage was light last year.**
- **Run II is coming!**



Talk by Oli on Operations during Run II this morning.

Storage

- Our storage pledge to CMS for 2015 is 1,040 TB/site.
- This is “usable” storage, which takes into account the need for resiliency. The raw disk needed is about double the amount usable.
- Currently almost every site deploys ~2PB of usable storage.

| | Usable Storage Space (TB) | Centrally Managed (TB) |
|-----------|---------------------------|------------------------|
| Caltech | 2,000 | 602 |
| Florida | 2,227 | 827 |
| MIT | 2,000 | 632 |
| Nebraska | 2,200 | 718 |
| Purdue | 2,150 | 892 |
| UCSD | 2,000 | 544 |
| Wisconsin | 2,250 | 794 |

Storage

- Bulk of 1PB pledge to CMS is used for centrally placed physics data samples (Dynamic Data Placement).
- The choice to over-provision storage was made deliberately to enable U.S. physicists. This space is available for their needs and their physics analysis groups.
- Storage was also not expected to get any cheaper, unlike CPU.

| | Usable Storage Space (TB) | Centrally Managed (TB) |
|-----------|---------------------------|------------------------|
| Caltech | 2,000 | 602 |
| Florida | 2,227 | 827 |
| MIT | 2,000 | 632 |
| Nebraska | 2,200 | 718 |
| Purdue | 2,150 | 892 |
| UCSD | 2,000 | 544 |
| Wisconsin | 2,250 | 794 |

Processing

- U.S. CMS is ~29% of total CMS headcount.
- Because we over-pledge Tier-1 resources (40%), we under-pledge Tier-2 resources (25%) to get our total computing commitment about right.
- For processing, this amounts to 18kHS06 per U.S. Tier-2 site in 2015, or about 1,800 batch slots.
- Our “build to cost” funding model allows us to deliver more to U.S. physicists, beyond our pledge to global CMS.

| Site | Batch Slots Deployed |
|-----------|----------------------|
| Caltech | 5,780 |
| Florida | 4,126 |
| MIT | 5,200 |
| Nebraska | 5,392 |
| Purdue | 11,360 |
| UCSD | 5,256 |
| Wisconsin | 6,730 |

So far we have not explicitly preferred U.S. users on above-pledge resources. We haven't needed to do so..

Talk by Brian Bockelman on Technology upgrades and Milestones this morning.

Processing

- Comparing kHS06 to batch slots, the average per batch slot is ~10.
- Some site's numbers look a bit low ...
- New benchmarking initiatives may be one of our goals for 2015.

| Site | kHS06 | Batch Slots Deployed | kHS06/Slot |
|----------------|-------|----------------------|------------|
| Caltech | 53 | 5,780 | 9.2 |
| Florida | 27 | 4,126 | 6.5 |
| MIT | 37 | 5,200 | 7.1 |
| Nebraska | 56 | 5,392 | 10.4 |
| Purdue | 128 | 11,360 | 11.3 |
| UCSD | 49 | 5,256 | 9.3 |
| Wisconsin | 67 | 6,730 | 10.0 |
| Total/ Avg. | 417 | 43,844 | 9.5 |

Talk by Brian Bockelman on Technology upgrades and Milestones this morning.

Networking

- **Most sites have upgraded to 100 Gbps WAN connections, or have plans to do so in 2015.**
- **Effectively leverage AAA**
- **We are organizing with ESNet the connection of all sites to the LHCONE VRF network this year.**
- **Nebraska and UCSD connected in February, Caltech this week?**

| Site | WAN (Gbps) Connection |
|-----------|-----------------------|
| Caltech | 100 |
| Florida | 100 |
| MIT | 10 |
| Nebraska | 100 |
| Purdue | 100 |
| UCSD | 80 |
| Wisconsin | 100 |

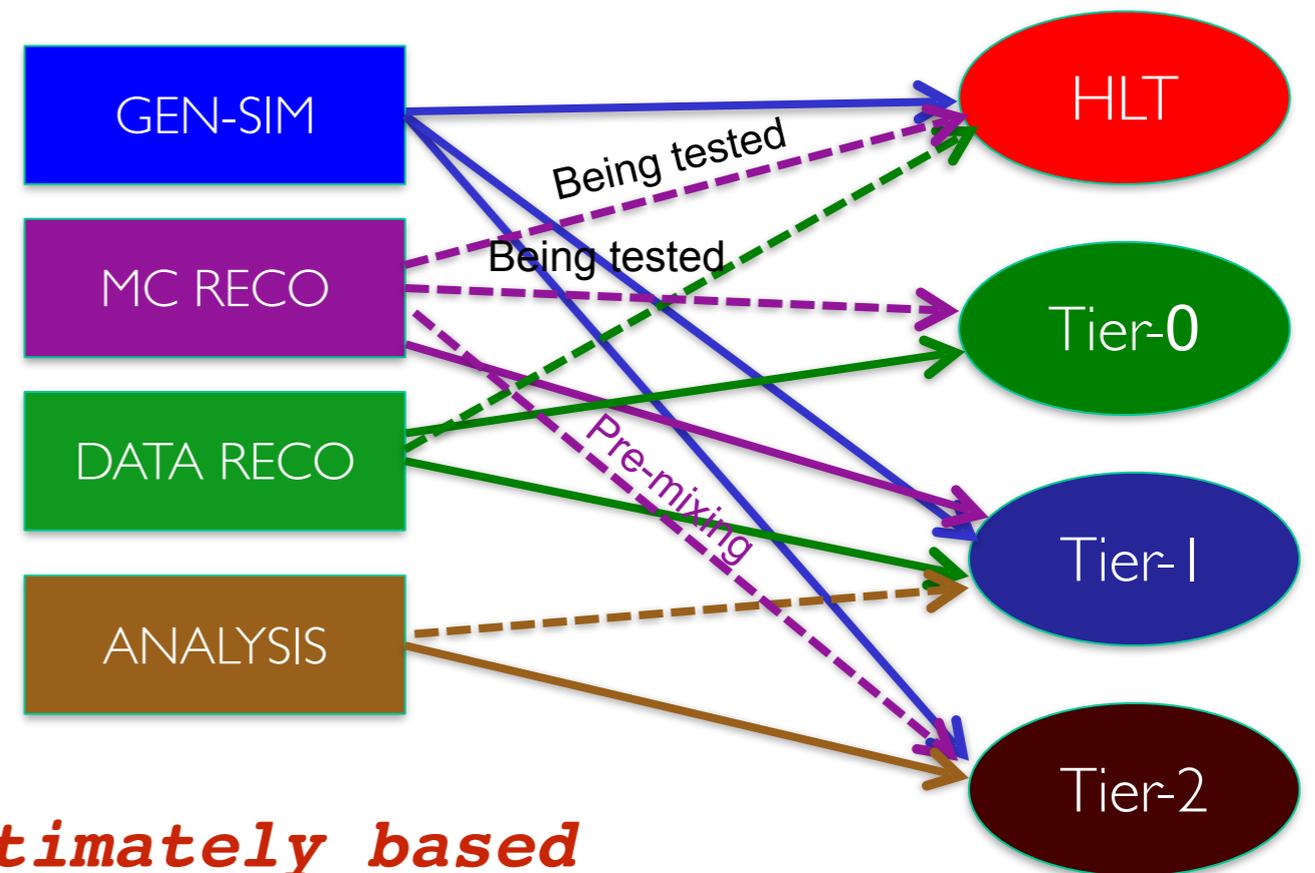
Talk by Jason Zurawski from ESNet in the joint session with ATLAS this afternoon.

Networking

- The way we exploit networking is evolving quickly!
- In Run II, resources are intended to work more like a coherent system than a collection of sites with specialized functions.



- Data Federation
- Single Global Pool for resource provisioning, including the HLT.



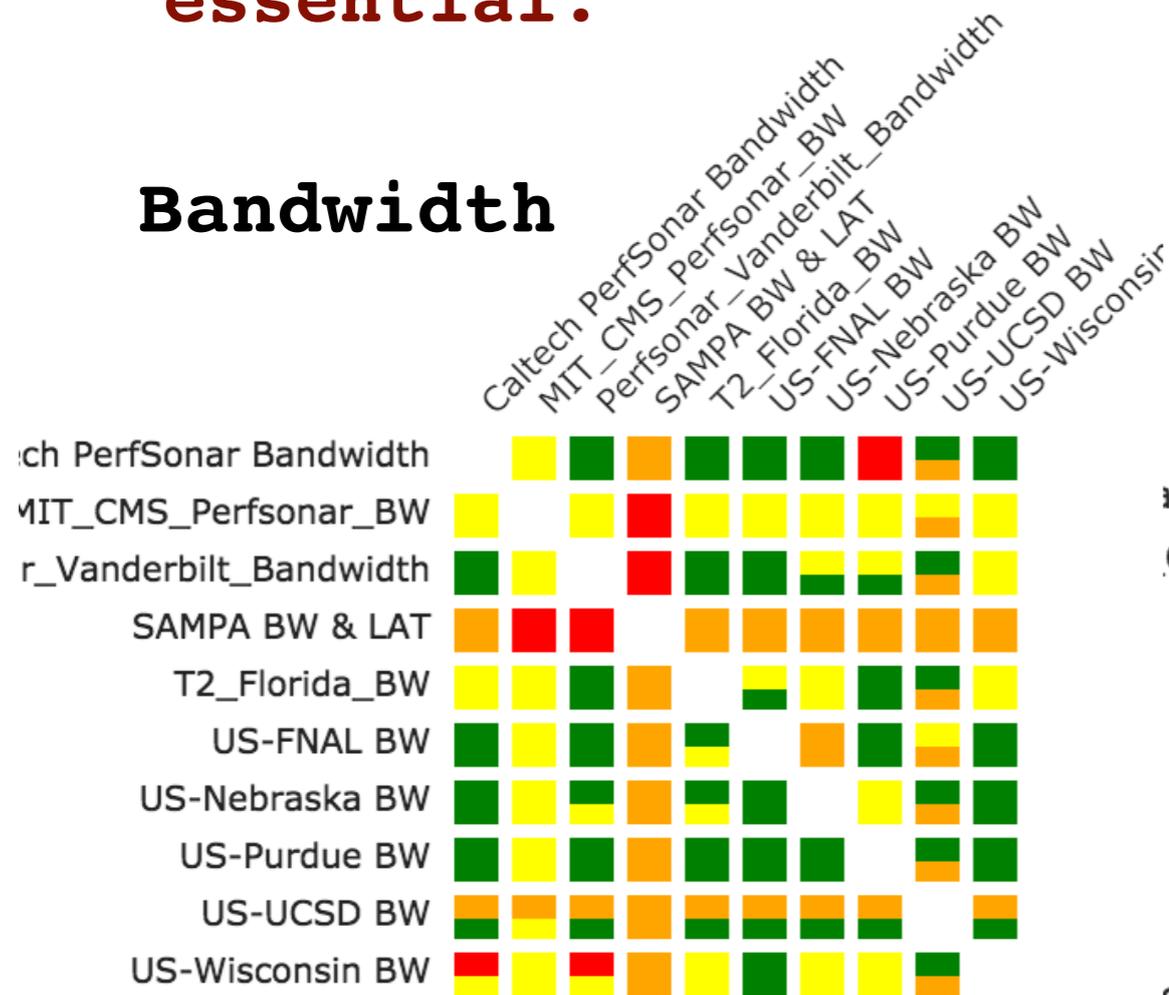
Since our funding model is ultimately based on the tiered computing model, will there be funding implications down the road?

Talk by Peter Elmer this afternoon on the Evolution of the CMS Computing Model.

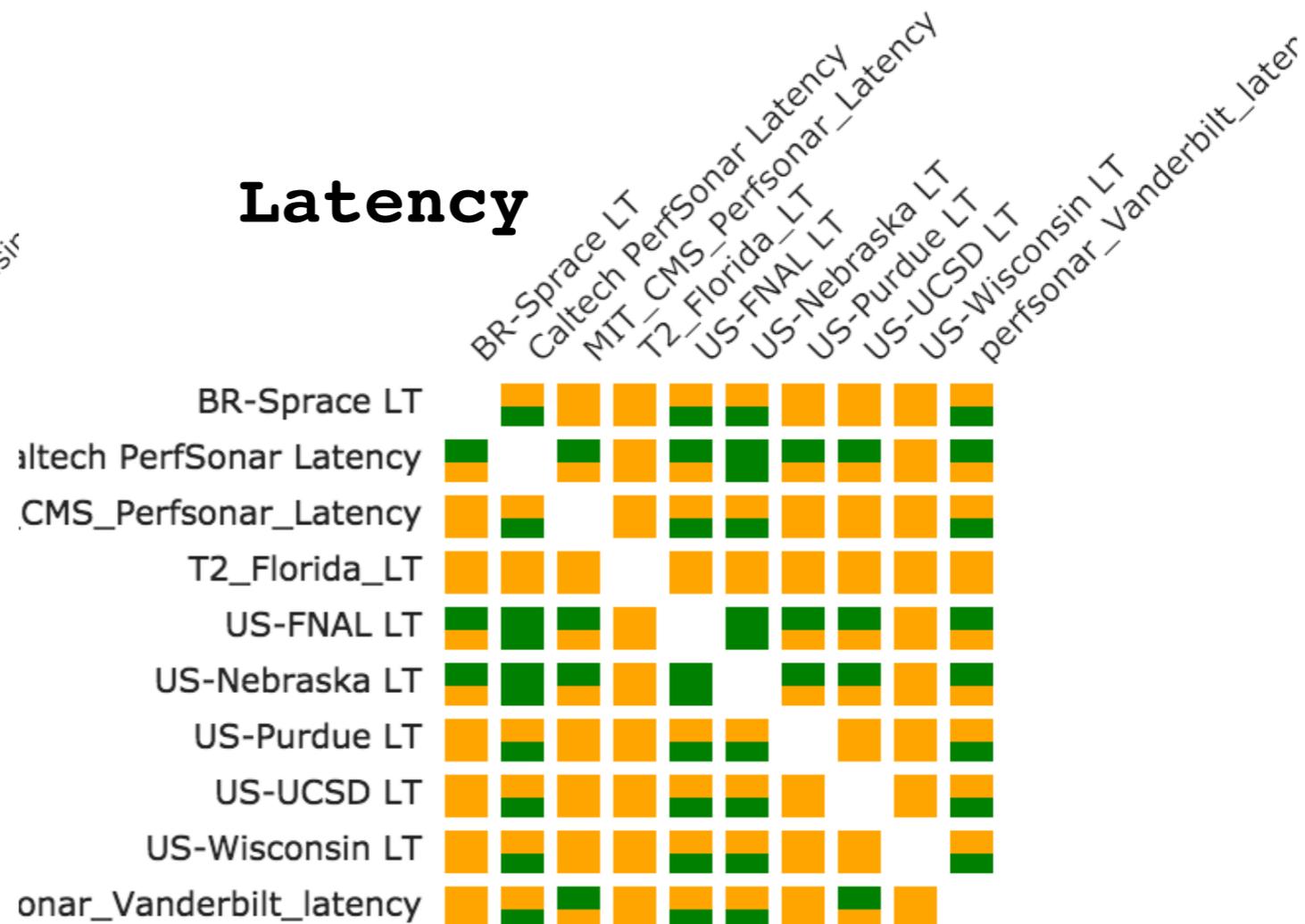
perfSONAR Performance

- **perfSONAR reliability has been poor, and degrading.**
- **As networking becomes more and more important, tools to identify and diagnose problems will be essential.**

Bandwidth



Latency



Talk by Shawn McKee this afternoon on Network Awareness and perfSONAR.

CMS Resource Requests

- **CMS Resource requests foreseen through 2016**
- **2015 pledges already met.**
- **Steeper increase projected in 2016.**

See Maria Girone's introductory talk in last week's Offline and Computing Week at CERN.

| Pledge per Site | 2014 | 2015 | 2016 |
|------------------------|-------------|-----------------------------|-----------------------------|
| Disk (PB) | 0.95 | 1.0 (+16%) | 1.4 (+37%) |
| CPU (kHS06) | 14 | 18 (+25%) | 25 (+40%) |

Hardware Budget Cuts

U.S. CMS
Operations
Program



Impact of the Computing Cuts

Expect less hardware money in 2016-2017 to pay for Phase 2 R&D

Profit from "buy ahead" to meet our pledges.

- Given the CMS computing resource estimates for Run 2, facility cost need to be up again this year and stay high during the run
 - Following a low period in LS1 of just maintaining, but not upgrading
- However, we are now planning for deep cuts into the facility budgets both on Tier-1 and on Tier-2, to fund R&D for Phase 2
 - This will have very significant impacts on our ability to provide the agreed/pledged disk storage resources to CMS, and to the level of support for U.S. physics analysis computing, data hosting and serving
 - Less computing and data access will be available for U.S. physicists
- Partial mitigation strategies
 - Will carefully review disk storage needs in the U.S., given improvements in computing model, event formats, trans-atlantic networking
 - Will look for ways of minimizing reliance on additional Tier-1 storage and instead develop cost effective approaches to hosting/serving the large disk-based data samples for CMS, while minimizing the needed investments in Tier-1 disk upgrades
 - Look for ways to save on Tier-2 hardware upgrades in 2016 and 2017, profiting from "buy-ahead" in 2013-15
- We will look at every possible cost savings across the program, to re-establish as much of the computing upgrades as feasible

See Lothar's talk in the March 13th U.S. CMS Meeting.

Hardware Budget Cuts

- **Baseline funding level for hardware is \$250K/site/year.**
- **Proposed cut in 2016 is about \$107K/site relative to the baseline.**
- **Cut in 2017 is less, about \$71K/site relative to the baseline.**
- **Funding levels return to baseline in 2018 once the Phase 2 R&D program has its own funding sources.**

Hardware Budget Cuts

- **Good news:**
 - **Budgets for personnel are not proposed to be cut.**
 - **We have bought ahead enough to meet our Tier-2 pledges for the next few years.**
- **Down side: Fewer above-pledge resources for U.S. physicists.**

Opportunistic

- One of the real “gold mines” of the Tier-2 program is opportunistic computing.
- Resources seamlessly accessible through the Tier-2's
- Includes university clusters and special allocations

Opportunistic

| Site | U.S. CMS Batch Slots | Opportunistic | (%) | Verified accessible through Global Pool |
|-----------|----------------------|----------------------|-----|-----------------------------------------|
| Caltech | 5,780 | 200 [*] | 3% | 5,676 |
| Florida | 4,126 | 4,000 | 97% | 5,771 ^{**} |
| MIT | 5,200 | 0 | 0% | 5,656 |
| Nebraska | 5,392 | 3,000 | 56% | 9,483 |
| Purdue | 11,360 | 9,200 | 81% | 20,518 |
| UCSD | 5,256 | 3,000 ^{***} | 57% | 4,597 |
| Wisconsin | 6,730 | 1,500 | 22% | 7,225 |

- Global Pool can “hit” most resources over time.
- In the past 2 weeks, most (or more) cores used at some time.

^{*} Increasing to 700 later in 2015.

^{**} Was verified to be much higher in an earlier test.

^{***} Special Allocation at Gordon

Changes in Management

- Thank you to Ken Bloom for his leadership of U.S. CMS Tier-2 Facilities Operations over the past several years.
- You have left us with a mature and well-functioning program.
- And good luck to him as the new leader of the Software & Computing Operations program.