

Wisconsin CMS T2 Site Report

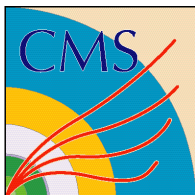


D. Bradley, S. Dasu, A. Mohapatra, C. Seys, C. Vuosalo
(**HEP Computing Group**)

Outline

- Infrastructure
- Resources
- Management & Operation
- Future Evolution
- Summary

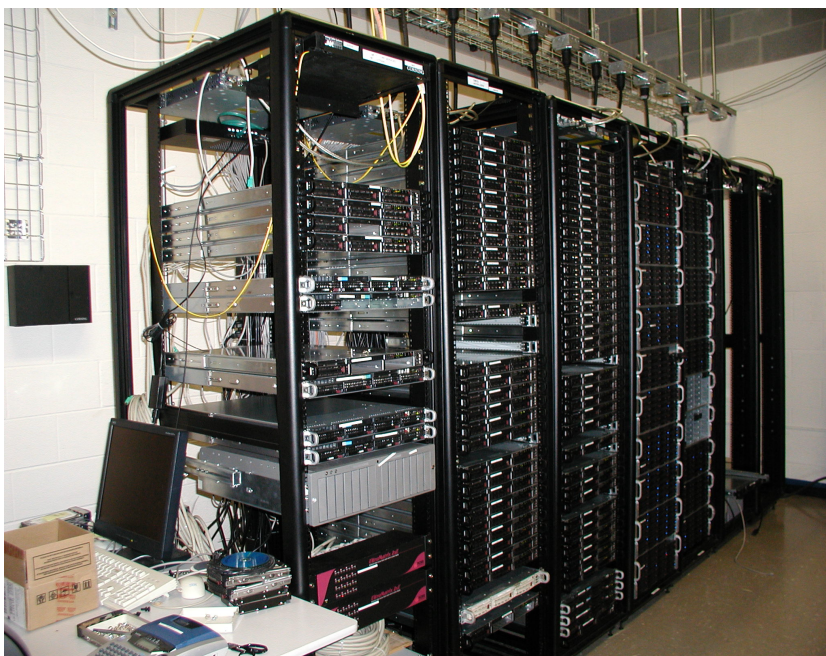


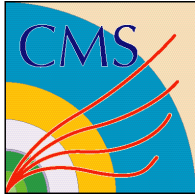


Infrastructure



- ✓ 2 machine rooms, 18 racks
- ✓ Power supply - 650 kW
- ✓ Cooling
 - Chilled water based air coolers and POD based hot aisles



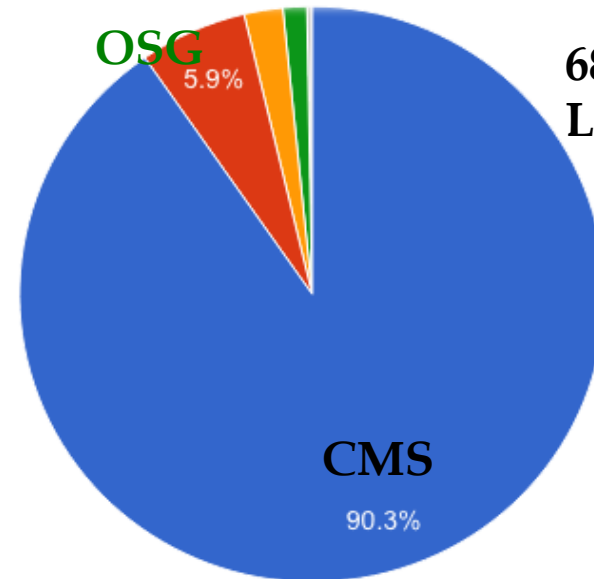


Compute / Storage Resources



✓ Compute (SL6 OS)

- T2 HEP Pool - @9800 cores now
- Added 2700 cores since last Oct
- Intel E5-2650/2660 chipsets
- Will add 1500 more cores in a few weeks
- Retired 400 oldest cores

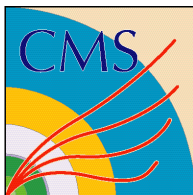


68 M Hrs
Last 1 Yr.

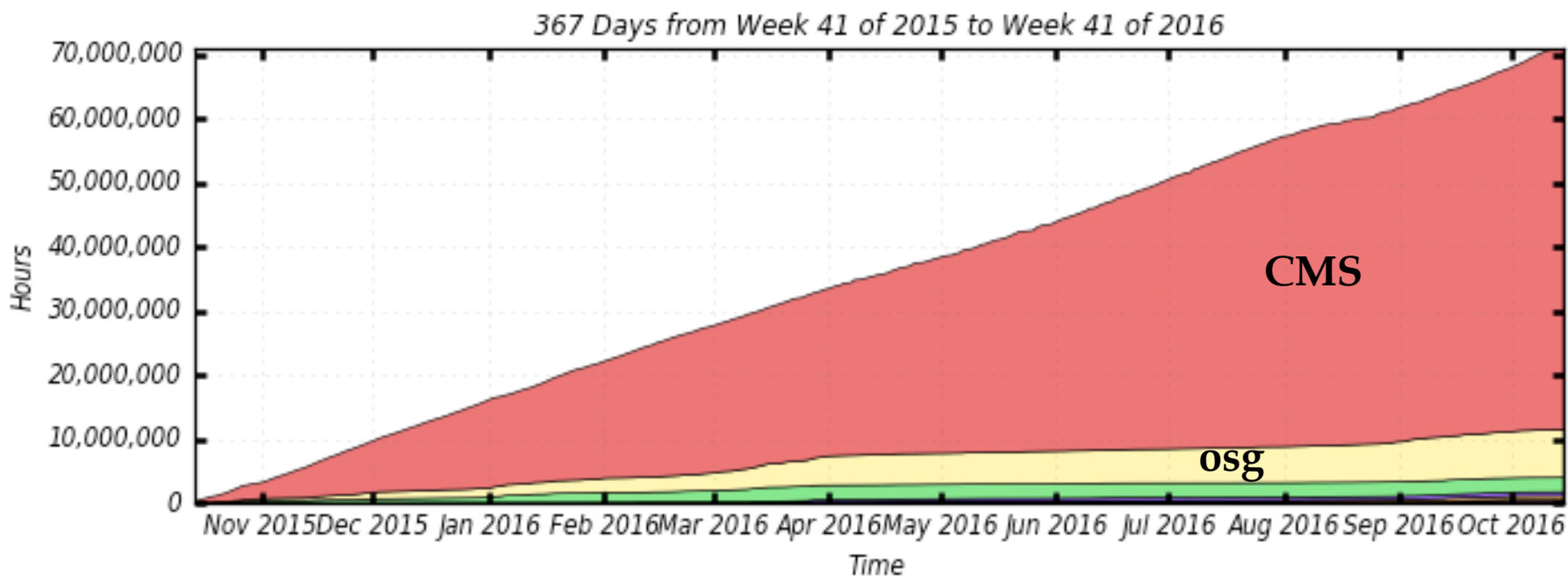
✓ Storage FS (Hadoop-2.0)

- In use for 5+ years
- Now @5.0 PB raw → 2.5 useable (2x replication)
- Will add 1PB in a few weeks



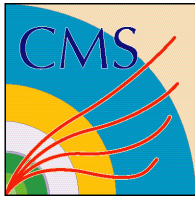


VO usage of GLOW resource



cms (59,558,297)	osg (7,483,121)	mu2e (2,261,914)	nova (617,361)	fermilab (577,621)
atlas (344,181)	hcc (157,817)	microboone (915.16)	darkside (795.63)	annie (676.98)
minos (668.12)	lariat (666.30)	argoneut (665.86)	seaqwest (663.85)	gm2 (661.14)
numix (659.43)	lar1nd (658.02)	minerva (657.77)	mis (57.62)	Other (6.71)

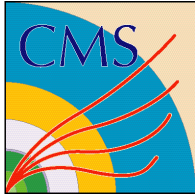
Total: 71,008,068 Hours, Average Rate: 2.24 Hours/s



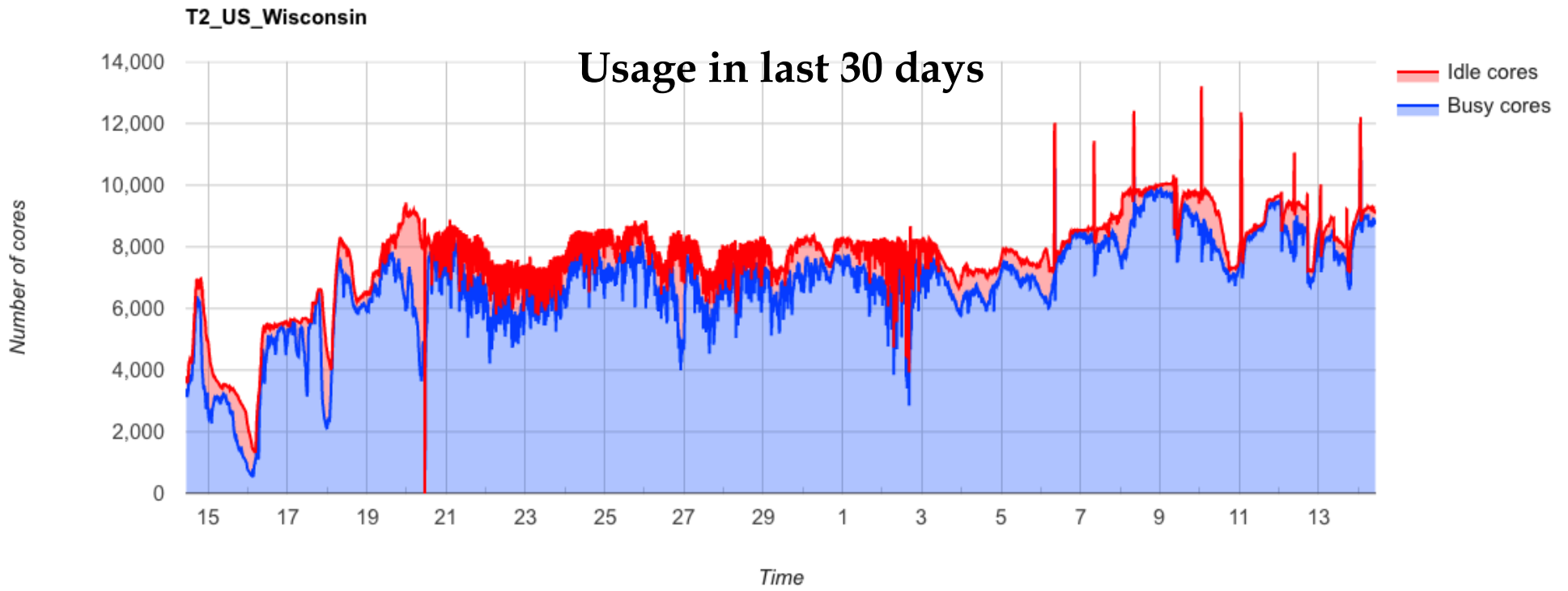
Job Scheduling and Multicore



- ✓ CEs : HTCondor-CE (3 CEs for redundancy)
- ✓ Job Batch Queue : HTCondor
- ✓ Running in full multicore mode for past 6 months.
- ✓ Supports “condor_ssh_to_job” (for debugging purpose).
- ✓ Opportunistic resources in the campus (CHTC) are used by CMS jobs regularly.
- ✓ Other VOs (osg, nova, fermilab etc.) use the T2 slots opportunistically.



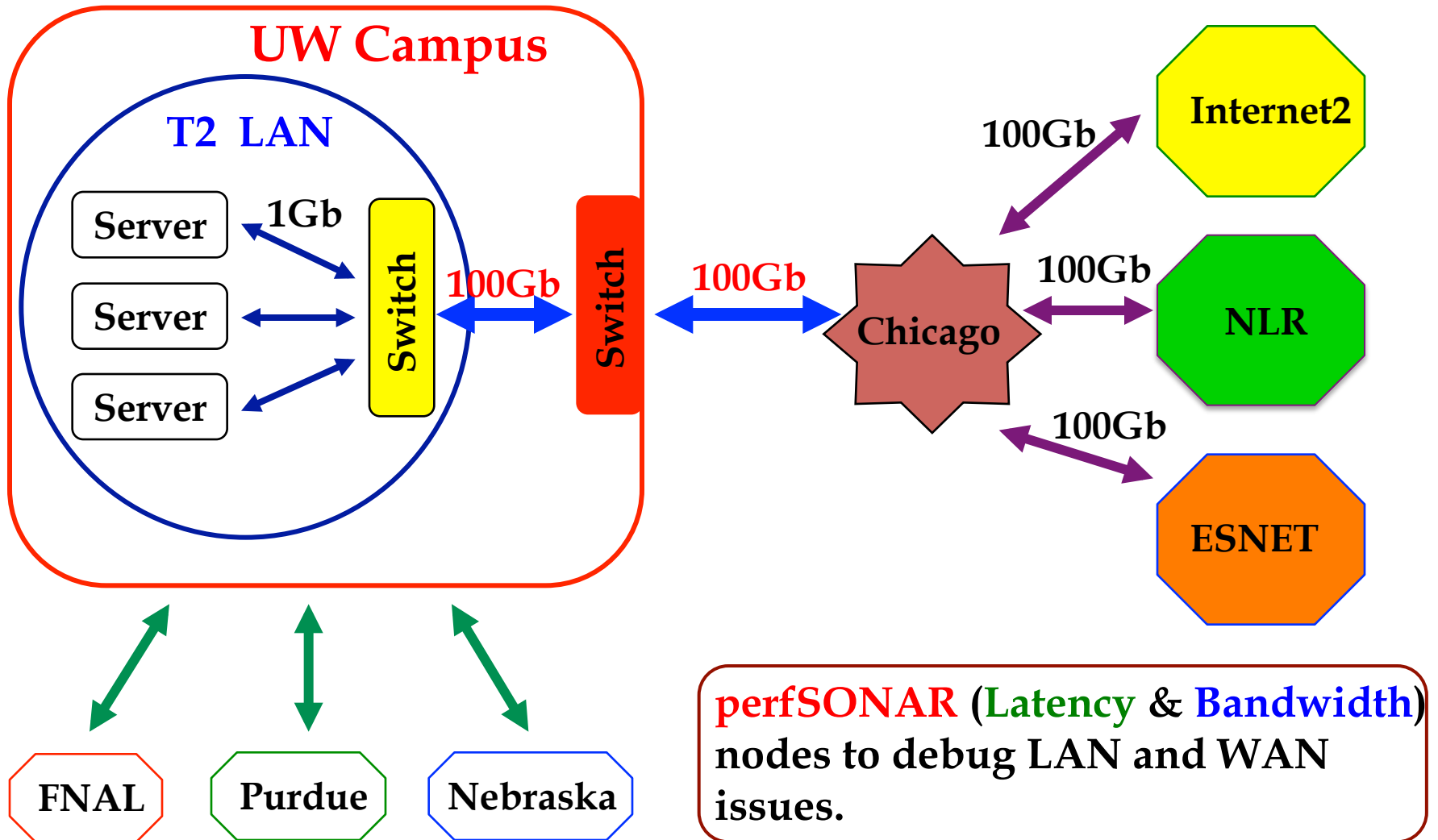
Multicore usage by CMS Pilots

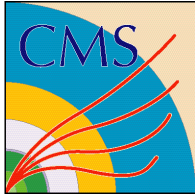


[avg, min, max]: [6806.0, 0, 11728] [852.0, 0, 4739]



Network Configuration



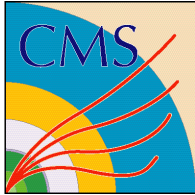


100Gb WAN with IPv4/IPv6



- ✓ WAN → 100Gb
- ✓ Server rooms interconnection → 80Gb
- ✓ Within server rooms → 1Tb switch backplane
- ✓ Dual Stack IPv4/v6 static configuration
- ✓ GridFTP (& SRM), Xrootd works with IPv4/v6
- ✓ Hadoop is still running with IPv4 only.

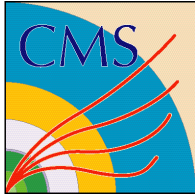
- ✓ Connected to LHCONE network



Software and Services



- ✓ **File systems, authentication, & proxy service**
 - AFS, NFS, CVMFS, Kerberos, Frontier/Squid
- ✓ **Job batch system**
 - HTCondor
- ✓ **OSG software stack**
 - HTCondor-CE, glexec, GUMS, SEs
- ✓ **Storage and Services**
 - Hadoop (hdfs), SRM, GridFTP, Xrootd
- ✓ **Cluster management & monitoring**
 - Puppet, Local Yum Repo(Ceph), Ganeti, Nagios/Icinga, Ganglia



Cluster Management & Monitoring



- ✓ **Puppet & Foreman**
 - Automatic software deployment and configuration
- ✓ **Ganeti (VM Manager/Debian)**
 - Handles a few dozen VMs using DRBD, KVM
- ✓ **Nagios/Ganglia**
 - Hardware, disks temp, services, memory, cpu/disk usage, I/O, network etc.
- ✓ **CMS and OSG dedicated tools**
 - SAM, Hammer Cloud, Dashboard, and RSV
- ✓ **Network latency/bandwidth (perfSONAR)**



Hadoop High Availability Test



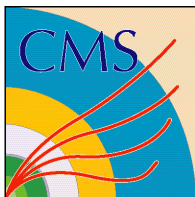
- ✓ **Non-HA hadoop : Primary name node is the single point failure.**
- ✓ **HA cluster basic requirements :**
 - ✓ **2 name nodes (active and standby).**
 - ✓ **At least 3 journal nodes (to share the file block info with both name nodes)**
 - ✓ **At least 3 zookeeper nodes (responsible for triggering the automatic failover mechanism)**
- ✓ **Functionality tested using local environment configuration in puppet.**
 - ✓ **Created test files in hdfs. Both active and standby nodes registered files correctly.**
 - ✓ **Failover worked as expected.**



Miscellaneous Information updates



- ✓ **HS06 benchmarking prior to each batch of hardware purchase.**
- ✓ **Experience with HTCondor-CE is positive except some bug bites occasionally.**
 - ✓ **Ex : Issue with multicore pre-emption (startd crash) in 8.4.2 forced us to keep Non CMS VOs OFF the cluster.**
- ✓ **Replacing Bestman2 with load balanced gridFTP service (using Linux Virtual Server).**
 - ✓ **Nebraska is doing very well with this. We are working on it.**
- ✓ **Planning to move away from using OSG BDII service.**



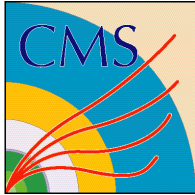
Summary



- ✓ Sufficient power and cooling capacity for next couple of years.
- ✓ The site is in good health and performing as expected.

T2_US_Wisconsin

LifeStatus:	✓																					
Site Readiness:	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Maintenance:	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up		
HammerCloud:	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	98%	100%	100%	100%		
SAM Availability:	100%	100%	100%	100%	100%	100%	100%	100%	100%	94%	98%	100%	100%	100%	100%	98%	100%	100%	100%	100%		
Good T2 links from T1s:	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16		
Good T2 links to T1s:	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16	16/16		
Active T2 links from T1s:	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
Active T2 links to T1s:	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	23	24	25	26	27	28	29	30	01	02	03	04	05	06	07	08	09	10	11	12	13	14
	Sep								Oct													



Thank You !



Questions / Comments ?