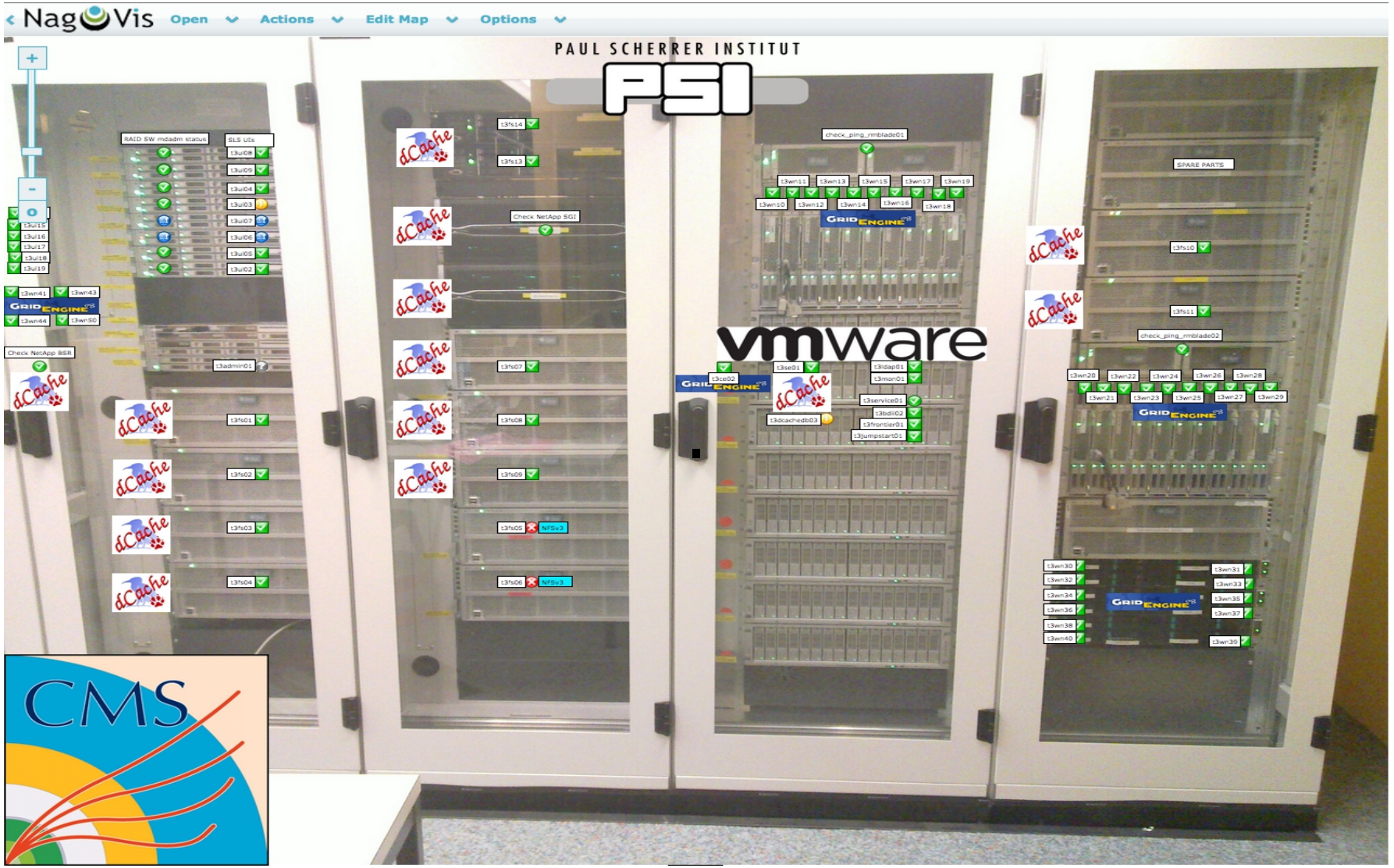
PSI T3 Status – Jan '15



NagVis picture from my recent talk: Monitoring the CMS T3 Cluster by Nagios

PSI T3 Status – Jan '15

CPUs, Storage and Virtual Services: news in red

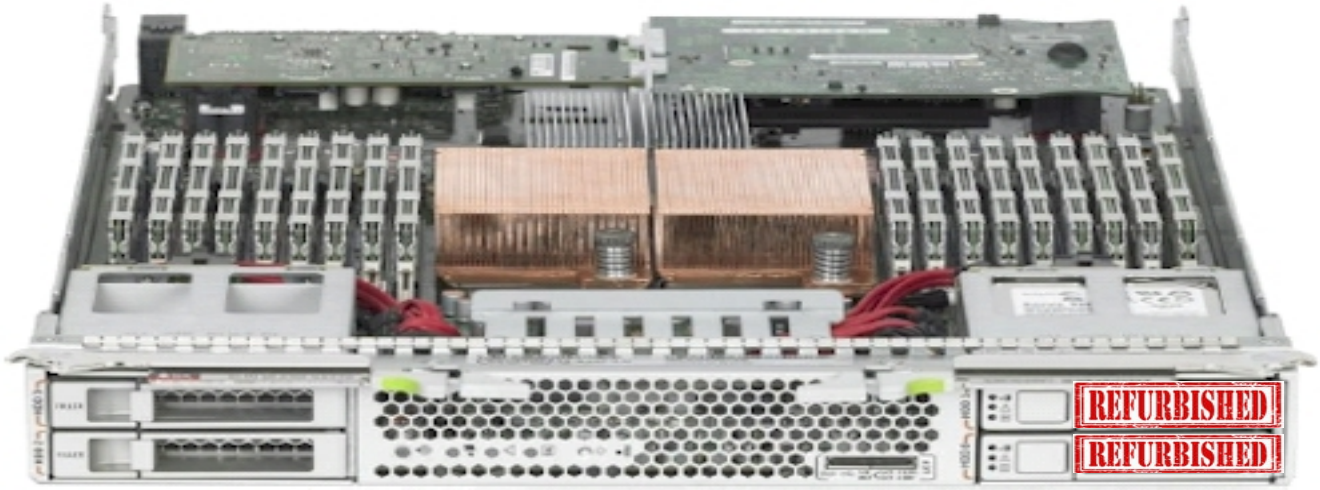
WNs/UIs	Processors	Cores/Node	HS06/node	HS06/core	Tot cores	Tot HS06
20 * WN SL6	X5560	8	117.53	14.69	160	2350
11 * WN SL6	E5-2670	16	263	16.44	176	2893
4 * WN SL6	AMD 6272	32	241	7.53	128	964
Tot. 36					Tot. ~ 460	Tot. ~ 6200
6 * UI SL6	AMD 6272	32	241	7.53	192	1446

Systems	TB Net per System
4 * SUN x4500	16
5 * SUN x4540	35
SGI IS5500	260
NetApp E5400	260
	Tot. ~ 760

Virtual Services
Sun Grid Engine master + MySQL DB
Site BDII, dCache SRM, dCache PostgreSQL
Ganglia Web, LDAP Server , Nagios
CMS Frontier (Squid), CMS PhEDEx
CVMFS (Squid)

About the 20 * WN SL6

- We got ~30 * 146GB 2.5” disks from CSCS (thank you again!)
- Used to bring from 2 to 4 the disks per Sun Blade
- The 4 disks allowed us to build a reasonably fast but safe **mdadm** RAID5 with 4 disks to deliver a **280GB XFS /scratch**
- We recommend **mdadm** ; it's free, easy to setup and rebuild, flexible, integrated in Kickstart, monitorable both by mdmonitor and by Nagios (indeed we use both of them)



T3 Steering Board held in Dec '14

- Yearly meeting among the PSI, ETHZ, UniZ representatives and the T3 admins to report the T3 status and plan its evolutions; topics are $$ / HW / SW

- Outcome, both in 2015 and 2016 we'll have less SNF $$

- Grid HW ; next Spring we're buying: 1* NetApp E5500 60*4TB disks SAS connected to 1* HP G9 server
- General HW ; 1* 48ports Cisco switch + 1* management server + 2* Oracle NFS servers (< 10TB each)
- About their details just ask ; if we'll change idea we'll inform you.
- Because of few $$ we won't be able to replace our oldest WNs, i.e. the Sun Blades I've reinstalled with 4 disks, so we ask to CSCS to notify us about their next decommissioned WNs ; consider this a permanent recurring request , thanks !!

SW , a comment about Puppet

- Like probably everybody we started with a Puppet Master and tens of Puppet Slaves
- It works, but since the PSI Puppet Master is a VM it's both very slow, > 1' for a Puppet run, and error prone, now and then some Puppet Slaves got their connections lost during the run.
- When we deploy and re-deploy a testing VM is boring to delete its previous certificate from the PSI Puppet Master
- We might install our own Puppet Master => yet another service

- We use the standalone Puppet apply method instead ; it's simpler to debug, much faster (< 1') because it exploits the local CPU, no lost connections Master → Slaves, the Puppet recipes are simply placed on ***/afs***, */nfs*, */gpfs …* and globally available.
- We use AFS keytabs to protect the sensible dirs in ***/afs***
- Without a good reason we won't reuse the Puppet Master.

SW , Planned Evolutions

- Our dCache 2.6 is next to its natural end-of-life so we'll migrate to dCache 2.10 by the end of Spring '15.

- A critical issue is understanding how to get the PSI T3 integrated in the new CMS Job Submission Framework ; **we might be forced to install our first CE !** We'd like to avoid it to don't raise the T3 complexity to the T2 level. To be studied well..

- We'll evaluate the Son of Grid Engine as a natural replacement of Sun Grid Engine 6.2u5 ; migrating to a complete new batch system like SLURM or Condor would just generate more effort .