

# PSI CMS T3 Status & '1[6-7-8-9] HW Plan

March '16  
Fabio Martinelli

# CPU, VMs, dCache Storage, ZFS NFSv4 Storage ( **news** )

WNs/UIs	CPU	Cores/Node	HS06/node	HS06/core	Tot cores	Tot HS06
20 * WN SL6 disposed soon	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

dCache Storage	TB Net per System
3 * SUN x4500 ( <b>Read-Only</b> )	15
5 * SUN x4540 ( <b>Read-Only</b> )	31
1 * SGI IS5500 ( Read-Write )	270
1 * NetApp E5400 ( Read-Write )	270
Files are replicated on the Read-Only pools to improve the storage bandwidth	Tot. ~200 ( <b>Read-Only</b> ) Tot. ~550 ( Read-Write )

VMs
Sun Grid Engine master + MySQL DB
Site BDII, dCache SRM, dCache PostgreSQL
Ganglia Web, LDAP Server , Nagios 4
CMS Frontier ( Squid ), CMS PhEDEx
OSSEC, SALTSTACK
CVMFS ( Squid )

ZFS NFSv4 Storage	TB Net per System
1 * HP G9 ( NFSv4 server )	~10 ( 24*600GB 15K SAS disks )
1 * HP G9 ( backup server )	~23 ( 12*3TB 7.2K SATA disks )

## The new T3 shared /home

- 1 x HP DL380 G9 featuring :
  - **24** x SAS 600GB 15k disks
  - 2 x SAS 146GB 15k disks ( mdadm RAID1 )
  - 1 x HBA Controller P440 for the **24**+2 disks
  - 256GB RAM
  - 2 x CPU E5-2660v3 ( Tot 40 cores )
  - 8 x 1Gbit/s Ethernet to be replaced in Spring by 2 x 10Gbit/s BASE-T
  - Cost ~25KCHF



## The new T3 shared /home

- 1 x HP DL380 G9 featuring ( backup server ) :
  - 12 x SATA 3TB disks
  - 2 x SAS 146GB 15k disks ( mdadm RAID1 )
  - 1 x HBA Controller P440 for the 12+2 disks
  - 64GB RAM
  - 2 x CPU E5-2660v3 ( Tot 40 cores )
  - 8 x 1Gbit/s Ethernet to be replaced in Spring by 2 x 10Gbit/s BASE-T
  - Cost ~10KCHF



```
[root@nfs4-server ~]# zfs list -t filesystem -o name,used,mountpoint,compressratio
```

NAME	USED	MOUNTPOINT	LZ4 RATIO	
data01	4.86T	/zfs/data01	1.35x	← SETTINGS ARE INHERITED BY DESCENDENTS
data01/shome	4.34T	/zfs/data01/shome	1.33x	
data01/shome/amarini	43.2G	/zfs/data01/shome/amarini	1.12x	
data01/shome/aspiezia	6.27G	/zfs/data01/shome/aspiezia	2.14x	
data01/shome/bbilin	1.61M	/zfs/data01/shome/bbilin	1.00x	
data01/shome/bianchi	74.1G	/zfs/data01/shome/bianchi	1.14x	
data01/shome/caber	19.1G	/zfs/data01/shome/caber	1.19x	
data01/shome/casal	123G	/zfs/data01/shome/casal	1.32x	

...

```
[root@wn ~]# grep nfs4 /proc/mounts ← THE WN MOUNTs/UMOUNTs A USER FS WHEN HIS JOB RUNs/ENDs
```

```
t3nfs01:/zfs/data01/ /mnt nfs4
```

```
rw,relatime,vers=4,rsize=1048576,wsiz=1048576,namlen=255,hard,proto=tcp,port=0,timeo=600,retrans=2,sec=sys,clientaddr=192.33.123.90,minorversion=0,local_lock=none,addr=192.33.123.71 0 0
```

```
t3nfs01:/zfs/data01/shome /mnt/shome nfs4
```

```
rw,relatime,vers=4,rsize=1048576,wsiz=1048576,namlen=255,hard,proto=tcp,port=0,timeo=600,retrans=2,sec=sys,clientaddr=192.33.123.90,minorversion=0,local_lock=none,addr=192.33.123.71 0 0
```

```
t3nfs01:/zfs/data01/shome/amarini /mnt/shome/amarini nfs4
```

```
rw,relatime,vers=4,rsize=1048576,wsiz=1048576,namlen=255,hard,proto=tcp,port=0,timeo=600,retrans=2,sec=sys,clientaddr=192.33.123.90,minorversion=0,local_lock=none,addr=192.33.123.71 0 0
```

...

## Strategy 1 : Slight storage increase, double CPU cores

### T3 resources Dec '15

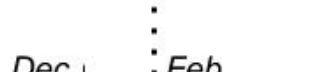
Reliable /pnfs size in TB = 550  
 t3wn\* Kilo HS06 = 6.2  
 t3wn\* cores = 464  
 /pnfs MB/s per core = 5.39

### T3 resources Jan '18

Reliable /pnfs size in TB = 880  
 t3wn\* Kilo HS06 = 11.1  
 t3wn\* cores = 960  
 /pnfs MB/s per core = 7.81



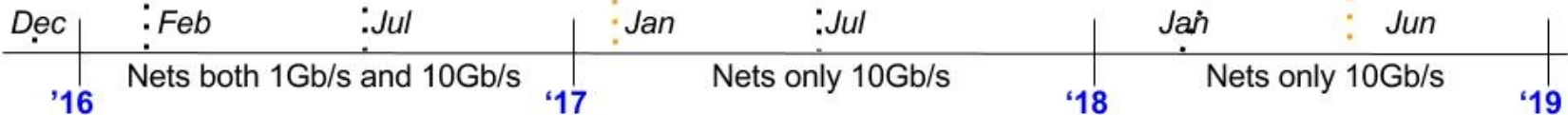
New 270TB in /pnfs :  
 ~50K CHF , Not Scalable  
 To fulfill the PSI storage request



Old 270TB in /pnfs running out of warranty in Jan '17  
 ~15K CHF to prolong the warranty till Jun '18



We've to replace 550TB in /pnfs  
 Probably only ~65K CHF in 2018 !





# Questions ?