

## Status (progress?) since last meeting

➔ **Spent last quarter mostly chasing up problems on both clusters (ce01, ce02)**  
and patching up our ailing hardware

➔ **ARC upgraded to 4.2.0-1.e16** (smooth upgrade on live system)  
Forced by a critical bug that caused the ARC cache to run out of space rendering one cluster non-operable

➔ **Nagios server deployed**  
Basic tests and alarms so far

➔ **ce03 running well** (targeting Todi at CSCS)

## HammerCloud Gangarobot

History Legend

■ offline  
 ■ brokeroff  
 ■ online  
 ■ NoQueue  
 ■ test

### Historic view for "panda\_queues\_all" from 00:00 01.09.2014 to 00:00 27.01.2015

Show  entries Search:

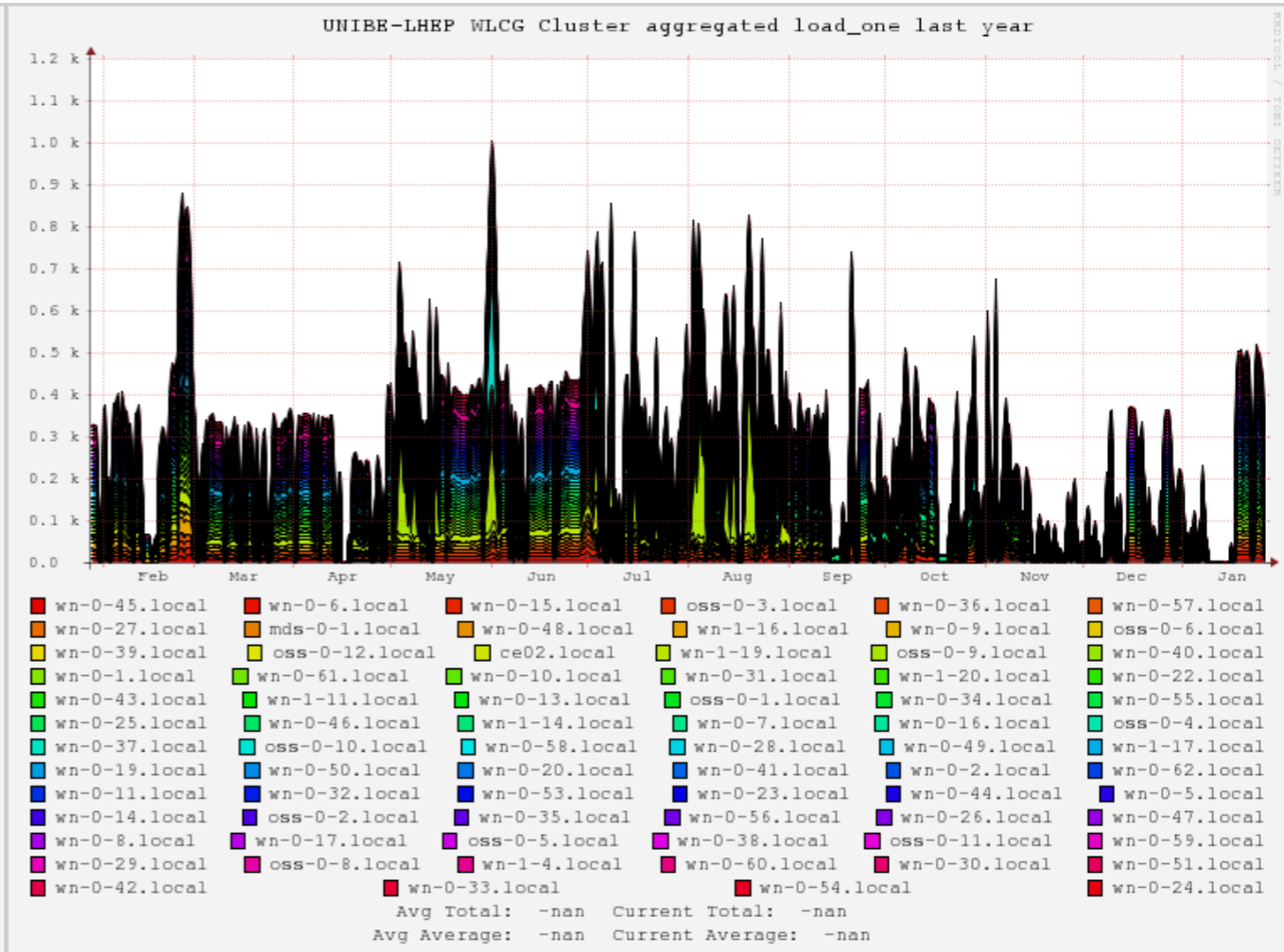
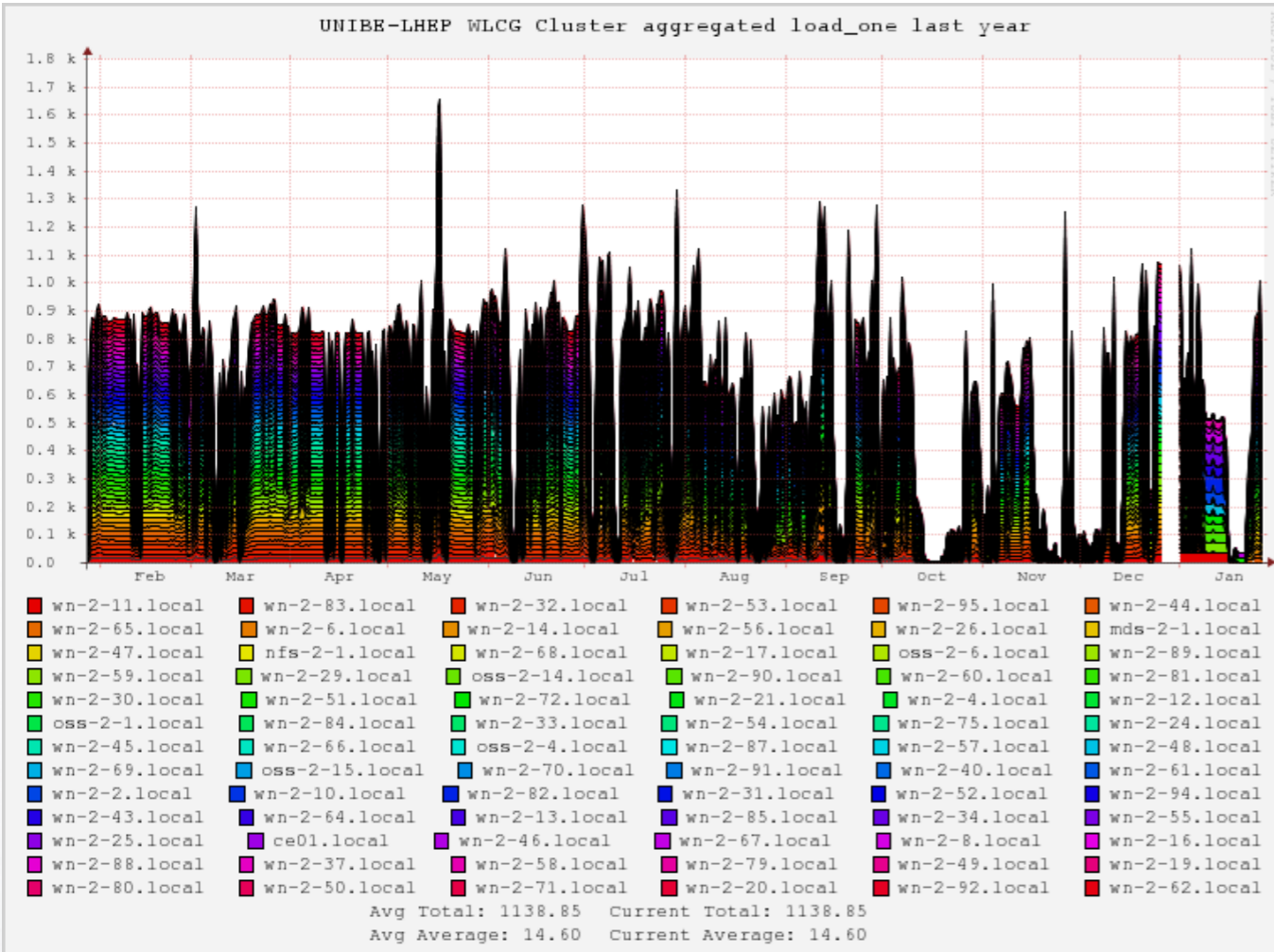
PANDA queue	SITE Name	TIER	CLOUD	History plot time bin = 296 hours	offline		brokeroff		online		NoQueue		test	
					%	count	%	count	%	count	%	count	%	count
ANALY_CSCS	CSCS-LCG2	T2D	DE		1.84	4	0.23	2	85.95	34	0	0	9.79	27
ANALY_CSCS_GLEXEC	CSCS-LCG2	T2D	DE		0	0	0	0	0	1	0	0	58.98	7
ANALY_UNIBE-LHEP	UNIBE-LHEP	T2	ND		3.35	1	0	0	78.83	24	0	0	15.64	24
UNIBE-LHEP	UNIBE-LHEP	T2	ND		3.35	1	0	0	75.79	10	0	0	18.67	10
UNIBE-LHEP-UBELIX	UNIBE-LHEP	T2	ND		0	0	0	0	97.81	10	0	0	0	0
UNIGE-DPNC	UNIGE-DPNC	T3	ND		0	0	0	0	96.34	13	0	0	1.47	3

Showing 1 to 6 of 6 entries First Previous 1 Next Last



## ce01

## ce02



## Issues and mitigations (1/3)

➔ **classic a-rex crashes marginally less frequent yet still there** (on average twice a month)

- Following Nagios alerts, restart service manually (lock/pid files left behind)

➔ **Nodes crashing due to memory starving** (long ongoing issue)

- complex issue related to the way memory is handled by 64-bit native applications.
- many tweaks in the gridengine configuration to mitigate (not enough)
- cgroups would be the best way to address this, but not available in the version of GE we run (GE-2011.11p1-1)
- already scaling the requested memory by a factor 2: increased to 2.5 has so far solved the issue
- tens/hundreds of k of jobs killed by gridengine lead to a request from ATLAS to drop the vmem limit
- then things magically stabilised as it happens ...

➔ **Re-deploying crashed nodes proved itself to be single largest and most time-consuming issue to deal with during most of the past months**

➔ **ARC cache full on ce01.lhep**

- cache cleanup thresholds are set in arc.conf (not honoured)
- tried manual cleaning (`/usr/libexec/arc/cache-clean`): no help
- tried brute force (`find /grid/lustre/cache/ -type f -mtime +45 -delete`) brought cache from 100% down to 66%
- however, advised about an ARC bug, so an upgrade to **4.2.0-1.e16** was needed



## Issues and mitigations (2/3)

### ➔ Hardware issues:

- **One of three RAID controllers failed on one of the DPM disk servers.**

With 20 controllers in total, the failure took a while to become evident. Eventually replaced the controller, but the server needed re-installation and configuration (OS + DPM stack), as it crashed at boot after service

- **Main switch upset affecting all services**

This is the switch the links up to the campus network (10Gb). Hiccup/intermittent operation between all ports. Power-cycling needed

- **Network down on the LAN of ce01 (31st December!!!!)**

This 1GB NIC serving the private ethernet network of the ce01.lhep cluster locked up. Every attempt of reviving it were vane. A clean re-boot was out of question (thousands of open files on lustre). Power-cycle eventually. All came back up clean.

- **One more lustre OSS locked up**

Just removed it from the system. Deployment of new hardware still pending

## Issues and mitigations (3/3)

### ➔ Miscellaneous:

- **Jura accounting issues on ce01.lhep (only)**

There's a limit of 1000 jobs per summary sent from Jura to APEL (`jobreport_options="urbatch:1000` in `arc.conf`)  
 Yet often, summaries including up to a few k jobs are created. These are not sent and remain in the "outgoing"  
 directory ( `/var/spool/arc/ssm/mq.afroditi.hellasgrid.gr/outgoing/00000000/` )

Wrote some scripts to split those in smaller summaries with <1k jobs each, and re-publish. Limited success. Ongoing

- **Security issues**

=> **CVE-2014-6271 (bash code injection)**

Easily patched

=> **CVE-2014-9322 (kernel - 18 Dec 2014)**

UI's patched straight away

Clusters patched in January, but quite problematic.

- Rebuilt 2.5.3 lustre clients but these did not work out of the box as usual
- Tried many tweaks and also a number of different lustre versions
- Eventually settled on:

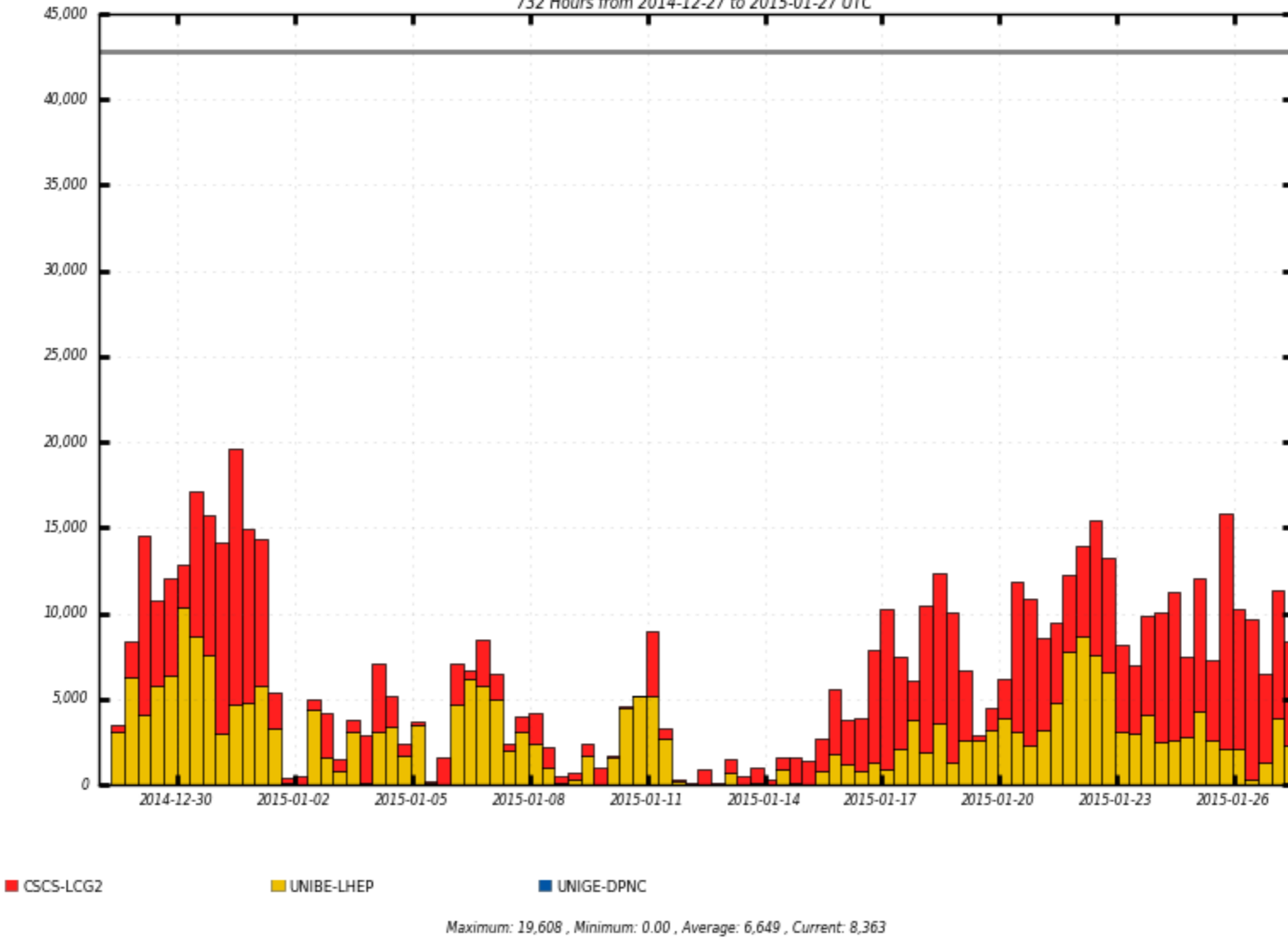
ce02: kernel 2.6.32-504.3.3.el6 + lustre-client-2.5.1-2.6.32\_431.5.1.el6 (pre-built, weak-modules)

ce01: kernel 2.6.32-504.3.3.el6 + lustre-client-2.5.3-2.6.32\_504.3.3 (built with a few tweaked options)



## CPU HEPSPEC06 Hours

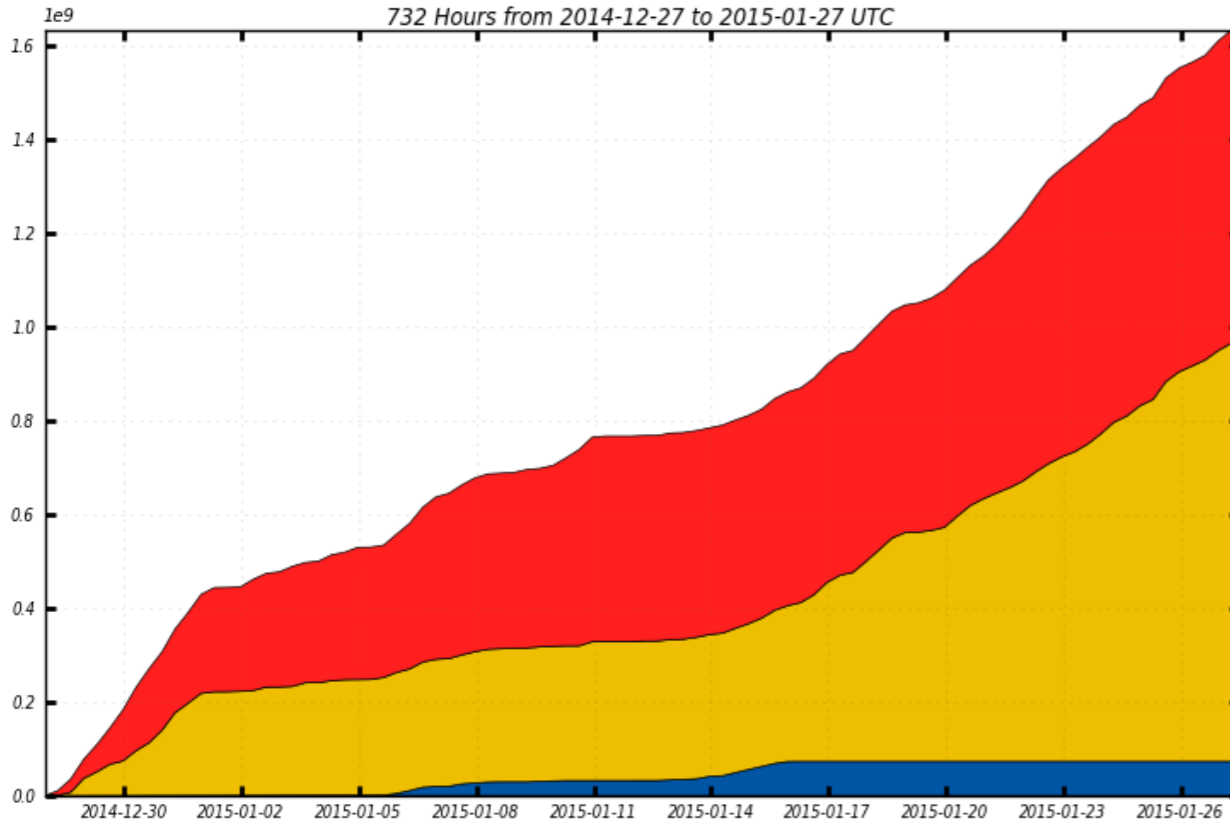
732 Hours from 2014-12-27 to 2015-01-27 UTC





dashboard

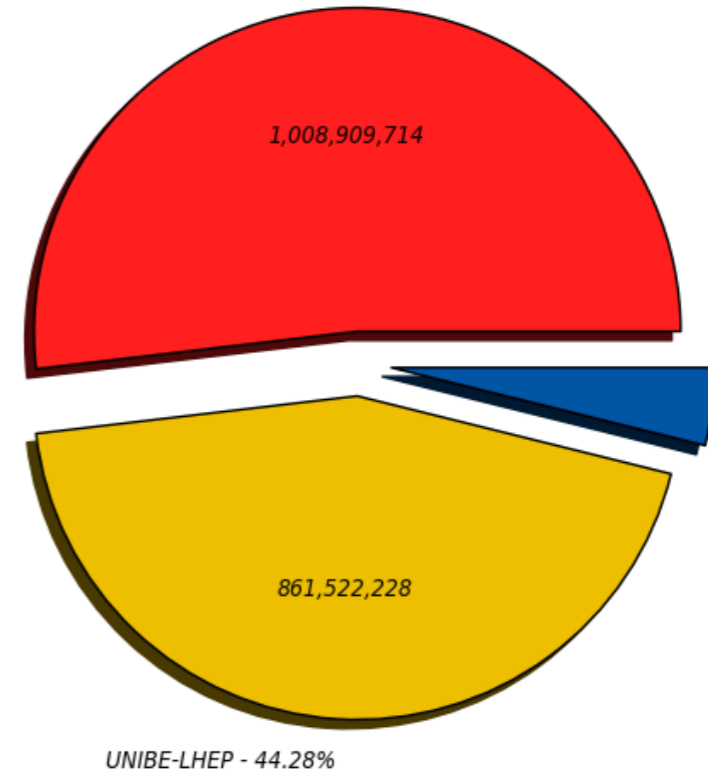
CPU consumption Good Jobs in seconds  
 732 Hours from 2014-12-27 to 2015-01-27 UTC



■ CSCS-LCG2 (891,993,663)   ■ UNIBE-LHEP (667,720,230)   ■ UNIGE-DPNC (73,241,505)  
 Total: 1,632,955,398 , Average Rate: 619.16 /s

dashboard

CPU consumption All Jobs in seconds (Sum: 1,945,715,890)  
 CSCS-LCG2 - 51.85%

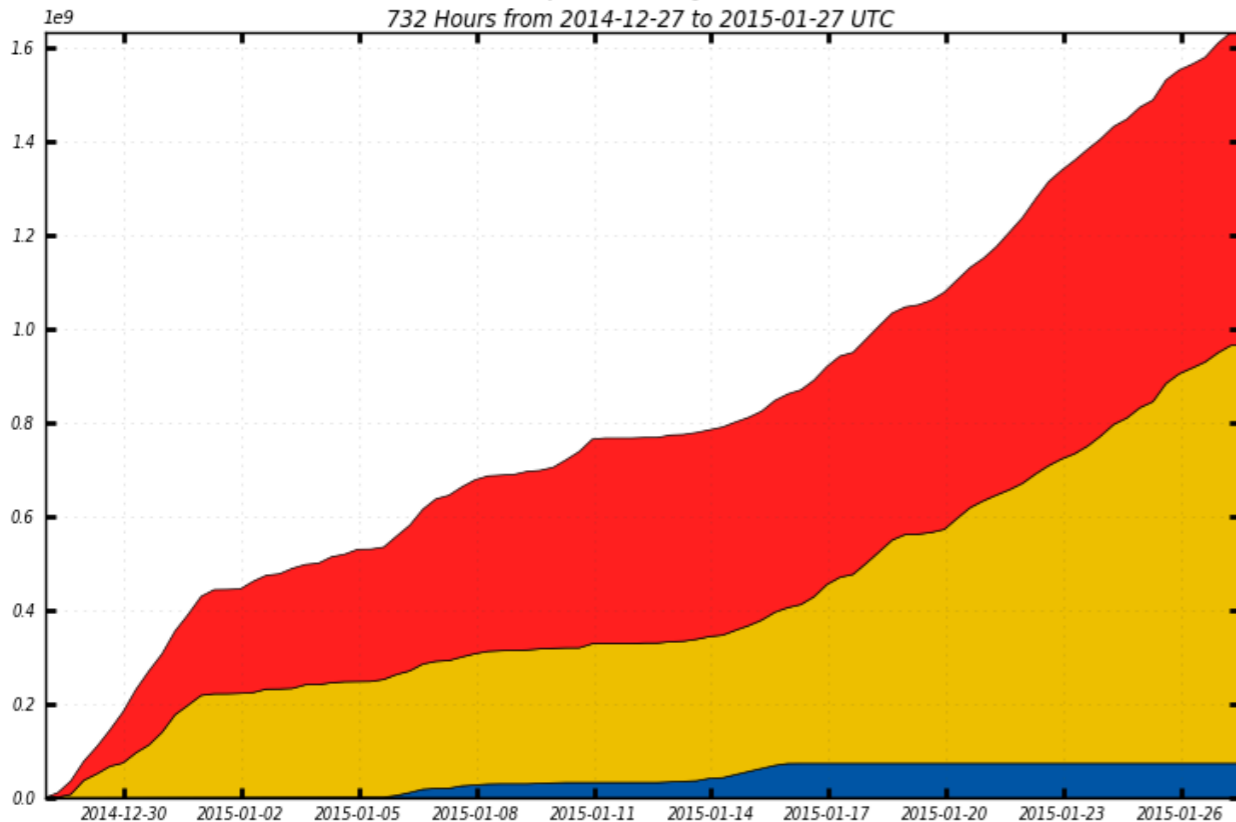


■ CSCS-LCG2 - 51.85% (1,008,909,715)   ■ UNIBE-LHEP - 44.28% (861,522,228)   ■ UNIGE-DPNC - 3.87% (75,283,947)



dashboard

CPU consumption Good Jobs in seconds  
732 Hours from 2014-12-27 to 2015-01-27 UTC

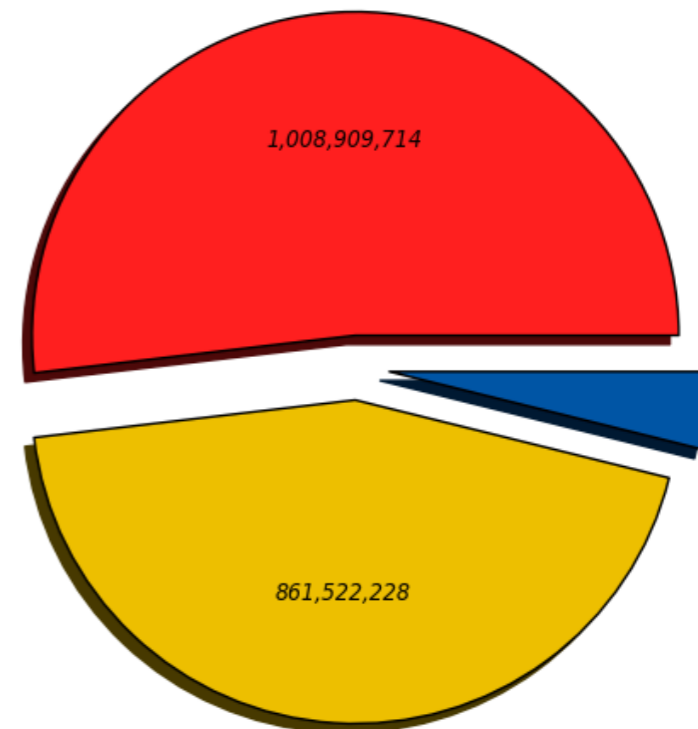


■ CSCS-LCG2 (891,993,663)    ■ UNIBE-LHEP (667,720,230)    ■ UNIGE-DPNC (73,241,505)

Total: 1,632,955,398 , Average Rate: 619.16 /s

dashboard

CPU consumption All Jobs in seconds (Sum: 1,945,715,890)  
CSCS-LCG2 - 51.85%



UNIBE-LHEP - 44.28%

■ CSCS-LCG2 - 51.85% (1,008,909,715)    ■ UNIBE-LHEP - 44.28% (861,522,228)    ■ UNIGE-DPNC - 3.87% (75,283,947)

Normalised CPU time [units HEPSPC06.Hours] by SITE and DATE

SITE	Jan 15	Total	%
CSCS-LCG2	2,397,536	2,397,536	56.83%
UNIBE-LHEP	1,821,040	1,821,040	43.17%
<b>Total</b>	<b>4,218,576</b>	<b>4,218,576</b>	
<b>Percentage</b>	<b>100.00%</b>		

[Click here for a CSV dump of this table](#)

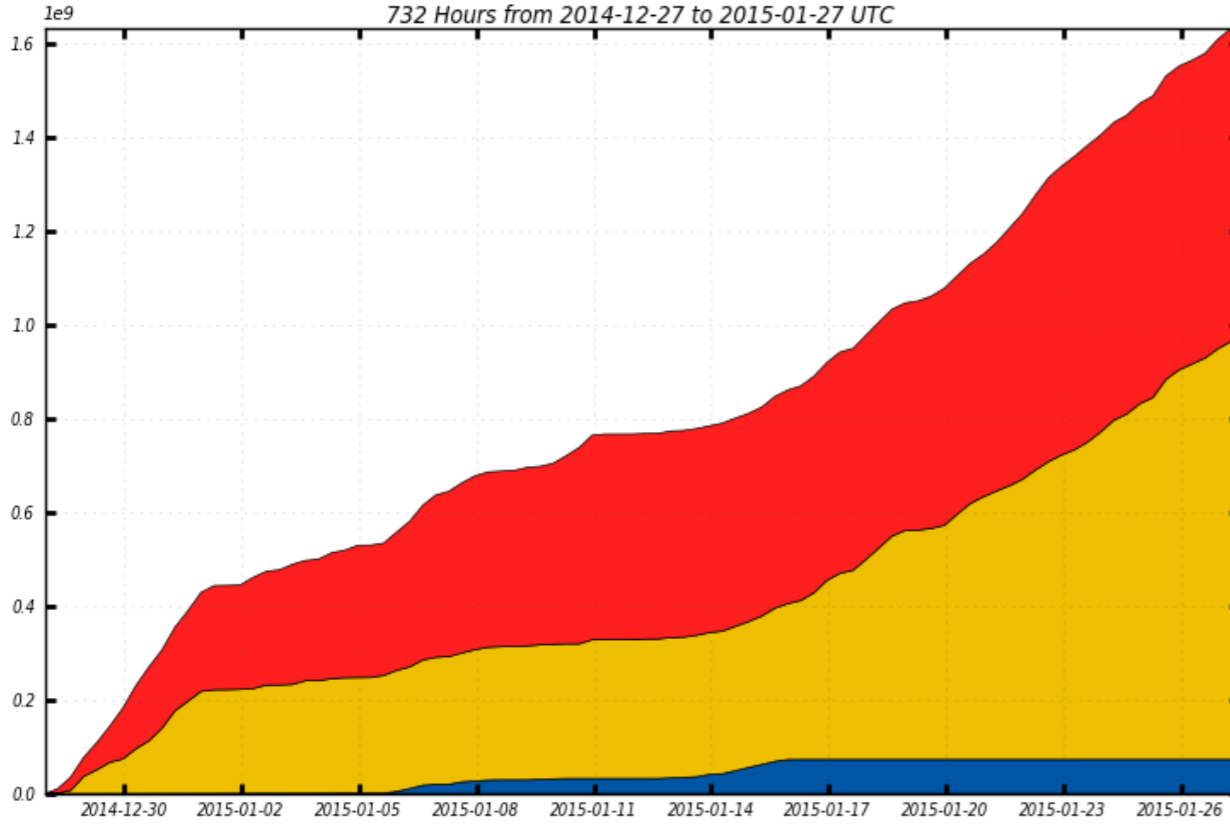
[Click here for a Extended CSV dump of this table](#)

[Click here for XML encoded data](#)



dashboard

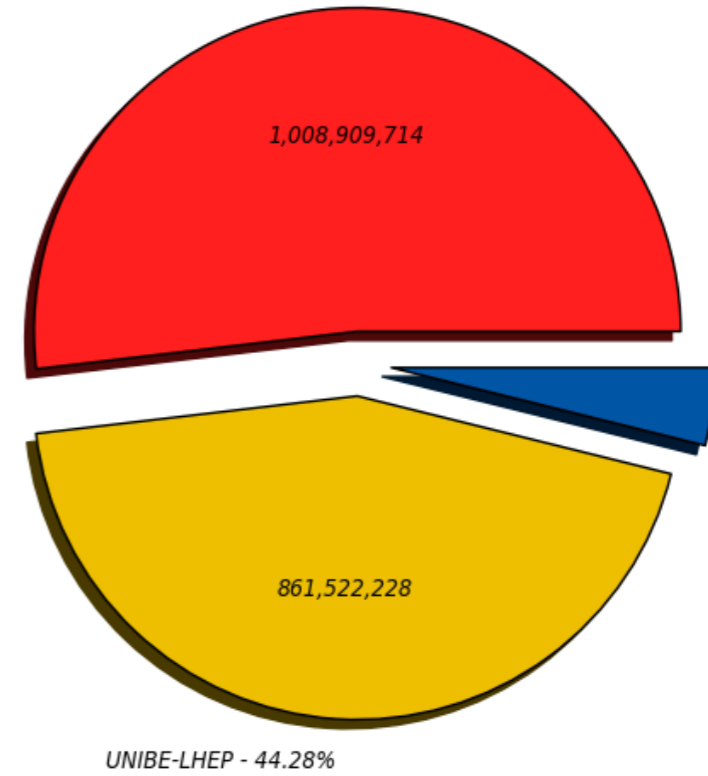
CPU consumption Good Jobs in seconds  
732 Hours from 2014-12-27 to 2015-01-27 UTC



■ CSCS-LCG2 (891,993,663) ■ UNIBE-LHEP (667,720,230) ■ UNIGE-DPNC (73,241,505)  
Total: 1,632,955,398 , Average Rate: 619.16 /s

dashboard

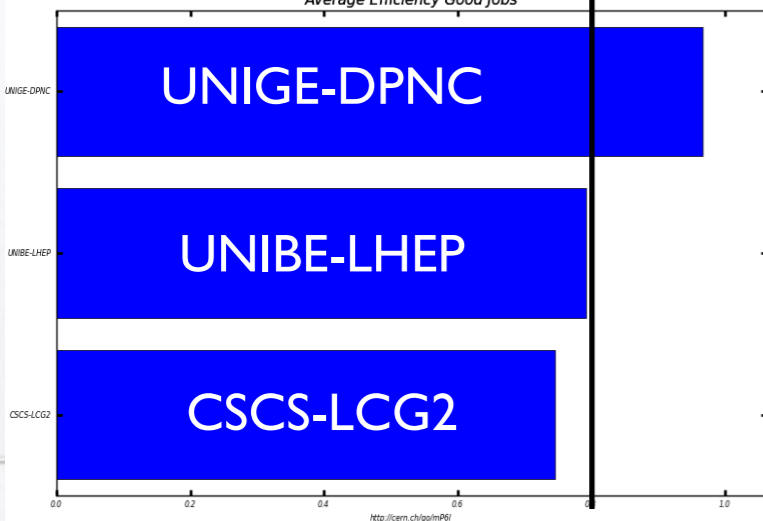
CPU consumption All Jobs in seconds (Sum: 1,945,715,890)  
CSCS-LCG2 - 51.85%



■ CSCS-LCG2 - 51.85% (1,008,909,715) ■ UNIBE-LHEP - 44.28% (861,522,228) ■ UNIGE-DPNC - 3.87% (75,283,947)

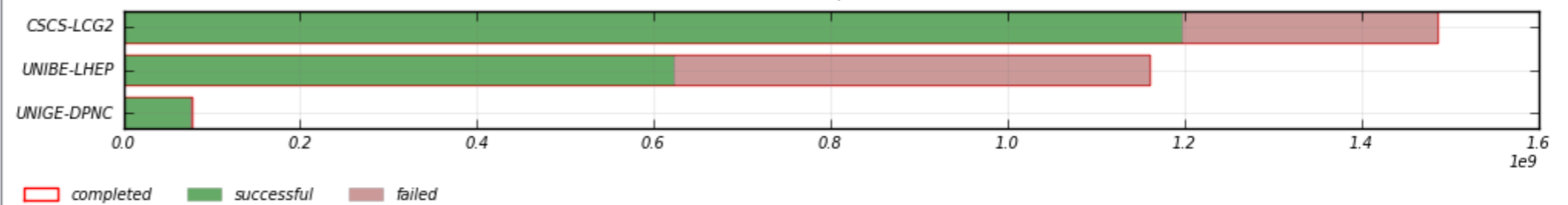
dashboard

Average Efficiency Good Jobs



0.8

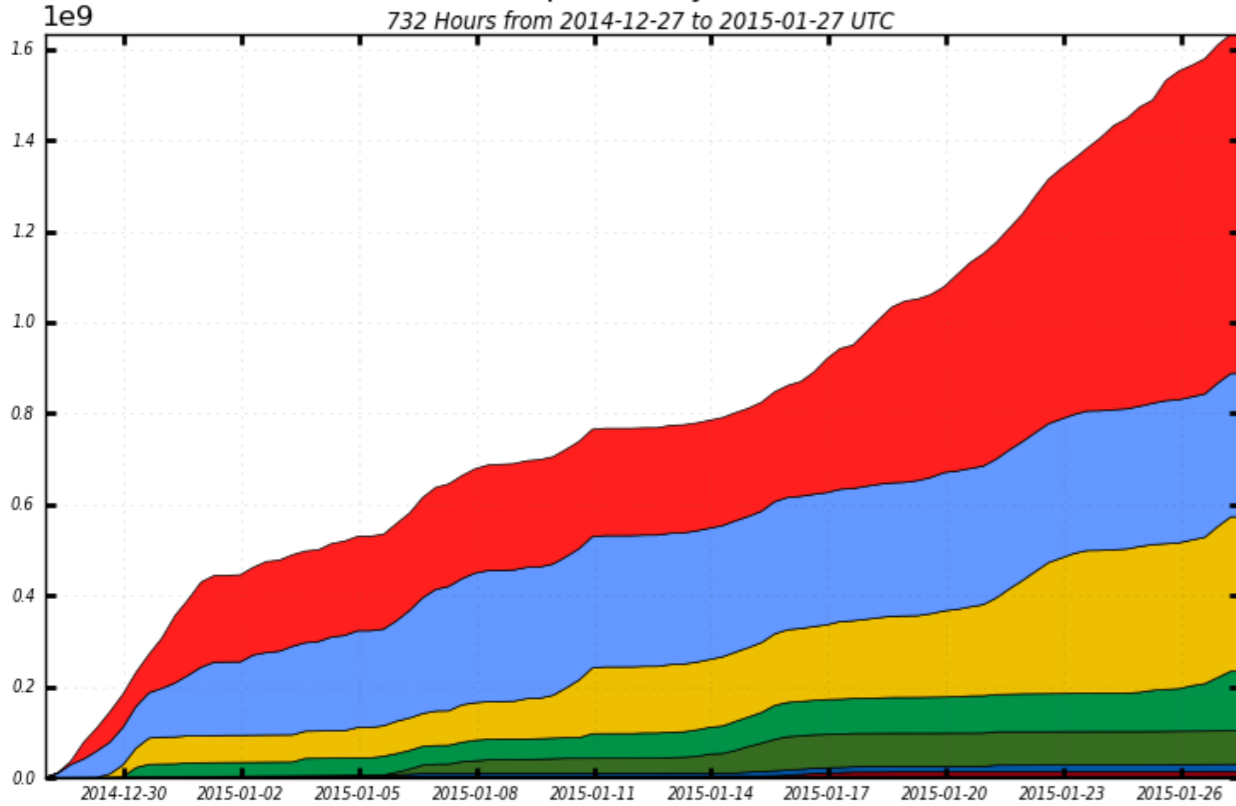
WallClock consumption in seconds







CPU consumption Good Jobs in seconds  
732 Hours from 2014-12-27 to 2015-01-27 UTC

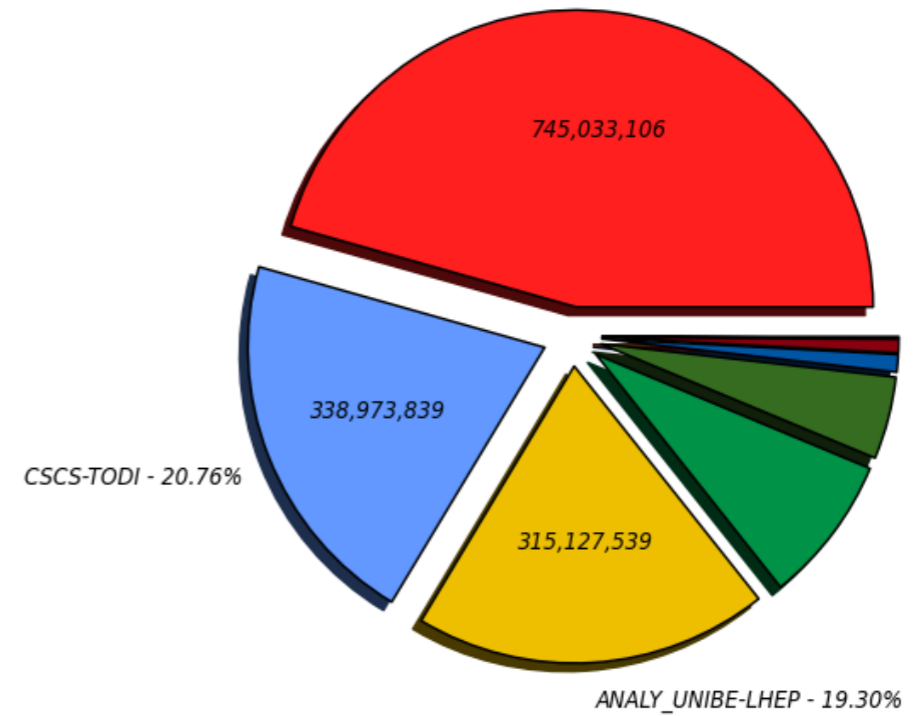


- ANALY\_CSCS (745,033,106)
- CSCS-TODI (338,973,839)
- ANALY\_UNIBE-LHEP (315,127,539)
- CSCS-LCG2\_MCORE (131,382,668)
- UNIGE-DPNC (73,241,505)
- CSCS-LCG2 (15,577,889)
- UNIBE-LHEP-UBELIX (11,911,037)
- UNIBE-LHEP (1,707,815)

Total: 1,632,955,398 , Average Rate: 619.06 /s



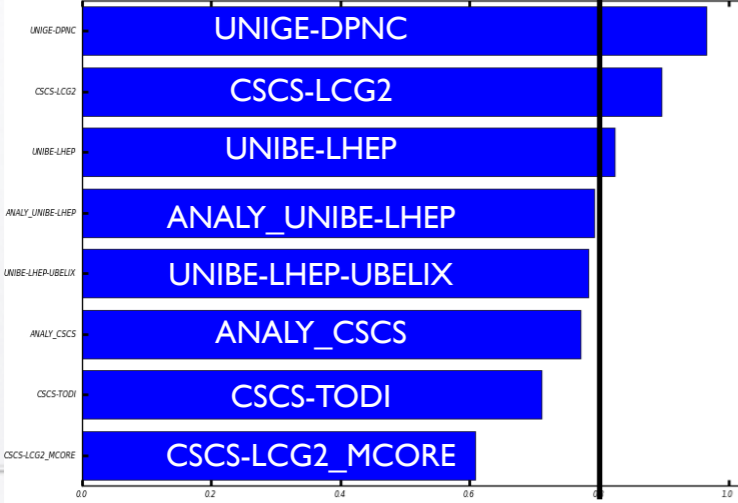
CPU consumption Good Jobs in seconds (Sum: 1,632,955,398)  
ANALY\_CSCS - 45.62%



- ANALY\_CSCS - 45.62% (745,033,106)
- ANALY\_UNIBE-LHEP - 19.30% (315,127,539)
- CSCS-TODI - 20.76% (338,973,839)
- CSCS-LCG2\_MCORE - 8.05% (131,382,668)
- UNIGE-DPNC - 4.49% (73,241,505)
- CSCS-LCG2 - 0.95% (15,577,889)
- UNIBE-LHEP-UBELIX - 0.73% (11,911,037)
- UNIBE-LHEP - 0.10% (1,707,815)

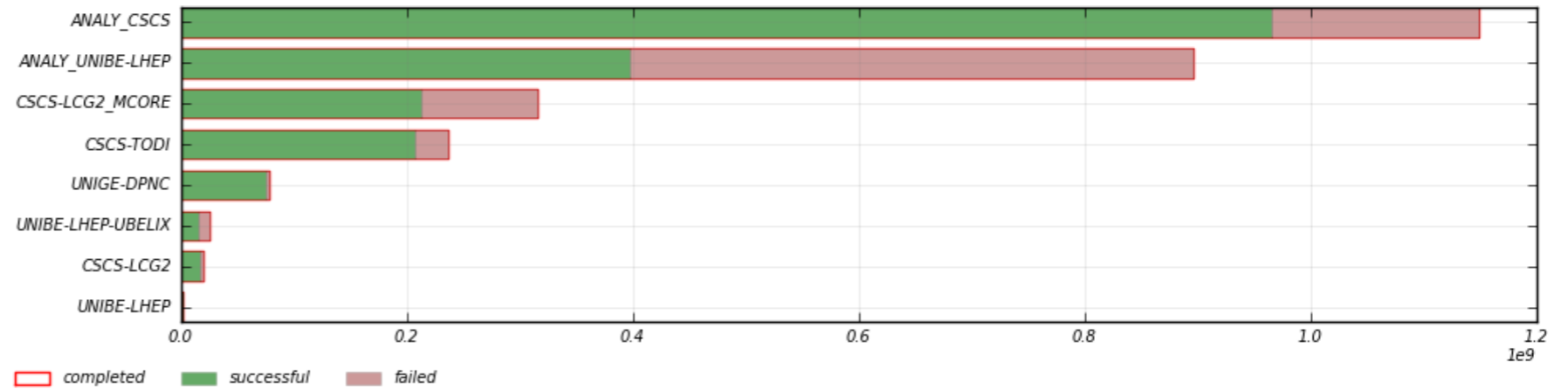


Average Efficiency Good Jobs



0.8

WallClock consumption in seconds



completed successful failed

