Tier 2 site report: CSCS

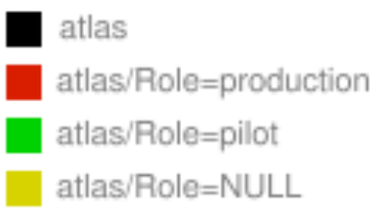
Marc Goulette (Geneva)
Gianfranco Sciacca (Bern)

CSCS report (March)

CPU: https://mon.lcg.cscs.ch/pbsplots/pbsplots.py? (inc. WLCG and NorduGrid)

	Nr of jobs	Walltime (h)	CPU time (h)
Total	**209376** (106313)	**280419** (352811)	(297207)
PROD	(52856)	(196696)	(178447)
PILOT	(44782)	(154656)	(118498)
User	(8675)	(1460)	(262)

February figures in brackets



full figures will be given in the next few days

- **CPU Efficiency: 74.2** (84.3 February)
- Compared to previous month, considerable less usage (see slides 4, 5)
- Fraction of total jobs: cream 71.1%, ARC 5.9%, lcg-CE 23%

CSCS report (March)

Disk: http://bourricot.cern.ch/dq2/accounting/t2_reports/FZKSITES/

Report for FZKSITES (UTC 2011-04-08 14:15:03.498342)

Site	Used(G)	Free(G)	Total(G)	%
CSCS-LCG2_DATADISK	156952	275047	431999	36
CSCS-LCG2_GROUPDISK	17095	32905	50000	34
CSCS-LCG2_HOTDISK	847	2153	3000	28
CSCS-LCG2_LOCALGROUPDISK	9533	467	10000	95
CSCS-LCG2_PRODDISK	1011	8989	10000	10
CSCS-LCG2_SCRATCHDISK	13433	14566	27999	47
Totals (TB)	**198.9**	**334.1**	**533**	**37.3**
(previous month)	(195.1)	(218.8)	(414)	

- 120TB disk came online, assigned to tokens
  https://twiki.cern.ch/twiki/bin/view/Atlas/StorageSetUp#Disk_Pool_Size_For_a_Tier_2
- Restored GROUPDISK to 50TB, added to HOTDISK, SCRATCHDISK
- All remaining to DATADISK
- Online: 533TB / Pledged: 468.5TB
- ~18TB to come in the next few weeks to DATADISK (evt. 450TB)

CSCS report (March)

General news

- April 2011 pledged hardware online now (next day to last month's meeting):

  2000 HEP-SPEC06 (~20% increase in capacity)
  Total capacity is now 13488 HEP-SPEC06 with fair shares atlas=40:cms=40:lhcb=20
  (AMD Opteron 6172 @ 2.1GHz with 3GB of memory per job slot)

  Online: 5395 (but not a hard number) / Pledge: 5420

- New 88TB GPFS shared scratch file system (60 disks, 2.5GB/s write, 1.5GB/s read)
  (in addition to the 115TB Lustre scratch area - 384 disks, 4GB/s operational, 7.5GB/s peak)

- Upgrade of both CREAM instances to 3.2.10-0 completed

- lcg-CEs set to drain, will be turned off next week

- Attempt to setup shared swap area for diskless nodes with no swap failed (see last month report)

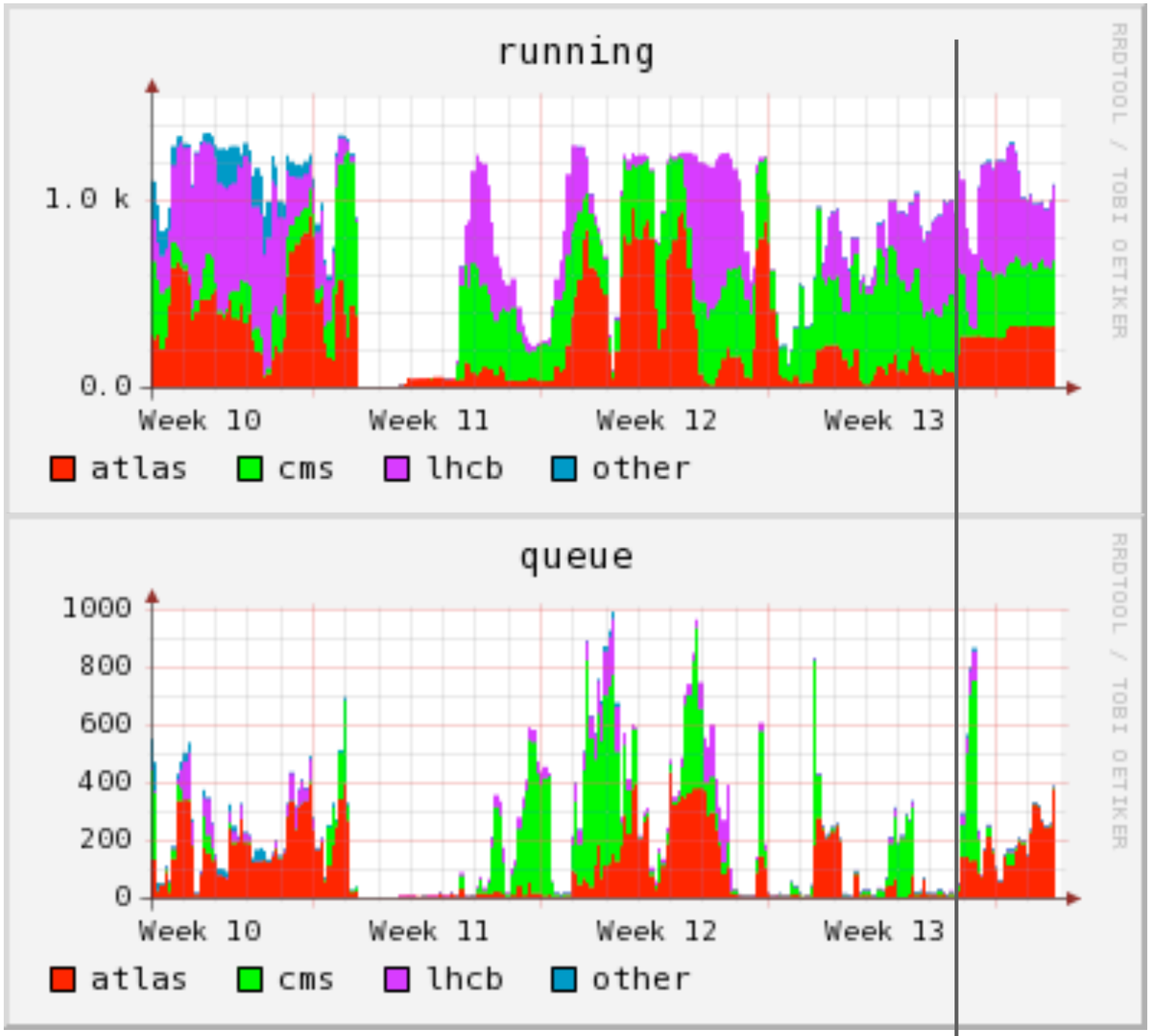
  Setup himem queues for memory hungry jobs (only):

```
1760    381     1           0     1   cream01.lcg.cscs.ch:8443/cream-pbs-atlashimem
1760    381     1           0     1   cream02.lcg.cscs.ch:8443/cream-pbs-atlashimem

default memory allocated on these queues is 4GB (2GB on all other queues). +20% tolerance, beyond which job is killed
```

  Each job on these will effectively eat up 2 job slots, so these should only get (reconstruction?) jobs that need more memory (not all analysis jobs)

CSCS report (March)



Nominal fair-shares:

atlas=40:cms=40:lhcb=20

ATLAS node confinement in place (see next slide)

CSCS report (March)

General issues

- Not as smooth operation as wished in March

- Very high metadata load on the lustre scratch. Stressed lustre more than usual.
  (Probably) caused fatal timeouts to LHCb jobs

  => traced back to heavy usage of "find" by ATLAS jobs

  => Wrapped find command to gather statistics. Higher than expected usage of find (99.5% ATLAS)

  => "du" usage ~1% compared to "find".

  => Admins decided to confine ATLAS to nodes with scratch on GPFS (not affected nearly as such) in order to restore stability

  => Only ~50% of the CPU pledge met that way

  => (in April) agreed with all VOs to move from hard confinement to "preferential node allocation":
  setup Moab to preferentially schedule ATLAS on the GPFS nodes first, other VOs on the lustre nodes first

  => Eventually understood (Paul Nilsson) that the frequency of find performed by pilots is going to be reduced, and eventually its usage discontinued (nice, since apparently not really needed)

  => Further investigation in Geneva to find out whether the application side is also at fault

  => Will look forward to remove the preferential node allocation as soon as this issue is fixed