- Status, plans and pledges

UNIBE-LHEP TIER 2 REPORT

Gianfranco Sciacca

AEC - Laboratory for High Energy Physics, University of Bern, Switzerland

CHIPP-CSCS face 2 face - 04 October 2017

UNIBE-LHEP TIER 2 REPORT

Site installation - UNIBE-LHEP

2660 cores, 24601 HS06 (15000 pledged), 600 TB SE, 180 TB scratch file systems

VOs: ATLAS, t2k.org, fermilab/uboone

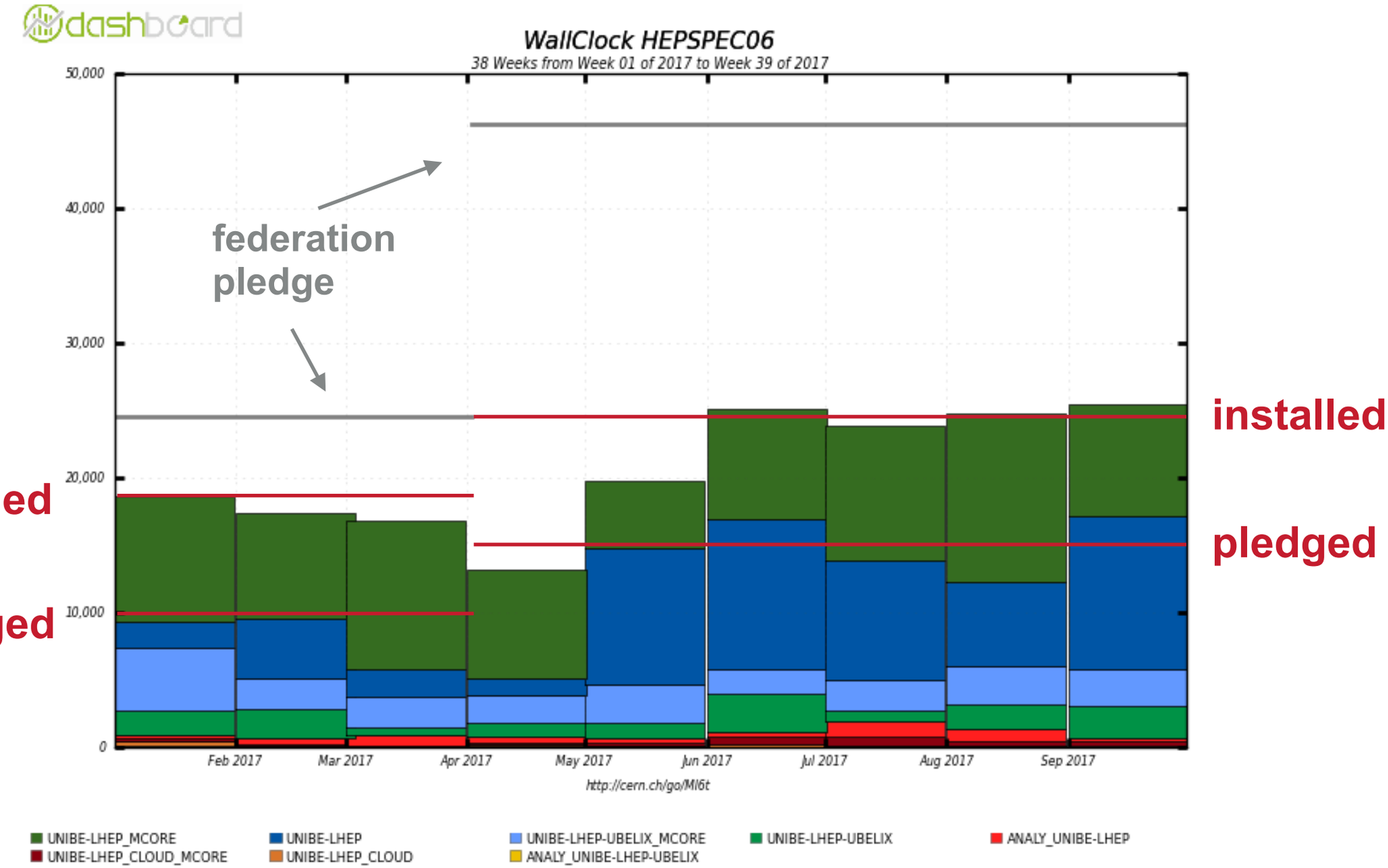
- **Two in house clusters, two ARC CEs, *2360 logical cores***
  - ce01.lhep.unibe.ch : 15422 HS06 - *1624 logical cores - Lustre 120 TB*
  - ce02.lhep.unibe.ch : 5875 HS06 - *736 cores - Lustre 60 TB*
- **Share on Ubelix, one ARC CE, *300 logical cores***
  - nordugrid.unibe.ch : 3304 HS06 (*) (*opportunistic usage up to twice as much*)
- **SwitchEngine cluster dismissed, *300 logical cores***
  - ce04.lhep.unibe.ch : this was about 3300 HS06 (*)
- **DPM Storage Element - *600 TB***
  - ATLASDATADISK *441 TB*
  - ATLASSCRATCHDISK *15 TB*
  - ATLASLOCALGROUPDISK *72 TB*
  - UBOONEPRODDISK *45 TB*

(*) *this is 300 cores @ 11.03HS06 (average coefficient calculated by Rebus). The real Ubelix HS06 share would be 4650 HS06*

UNIBE-LHEP TIER 2 REPORT

WallClock HS06 - UNIBE-LHEP

- **UNIBE-LHEP** *HS06 ATLAS* ***installed (**): 24601 - pledged: 15000*** (2017)
- **UNIBE-LHEP** *HS06 ATLAS* ***installed (**): 18642 - pledged: 10000*** (2016)



(**) *- also serving t2k.org and fermilab/uboone*
- some opportunistic usage on Ubelix

UNIBE-LHEP TIER 2 REPORT

Hardware status and pledges - UNIBE-LHEP

- **After a few years hiatus, we secured some very limited funds for hardware**
  - replaced some ailing hardware, and increased the HS06 capacity
  - some of the old hardware still going strong
  - about half of the ce01 cluster capacity (or ~1/3 of the total in house) is now new hardware
  - added 150 TB to the SE, 100 TB to *ATLASDATADISK* and 50 TB for *fermilab/uboone*
  - could bring the storage pledge to ATLAS above the threshold indicated for decommissioning
- **A bit hard to predict how we will evolve in the future years**
  - we have a *tentative* roadmap for hardware replacement and 20% increase year-on-year over 4 years
  - many factors might influence this (hopefully in the right direction!)
- **Pledges**

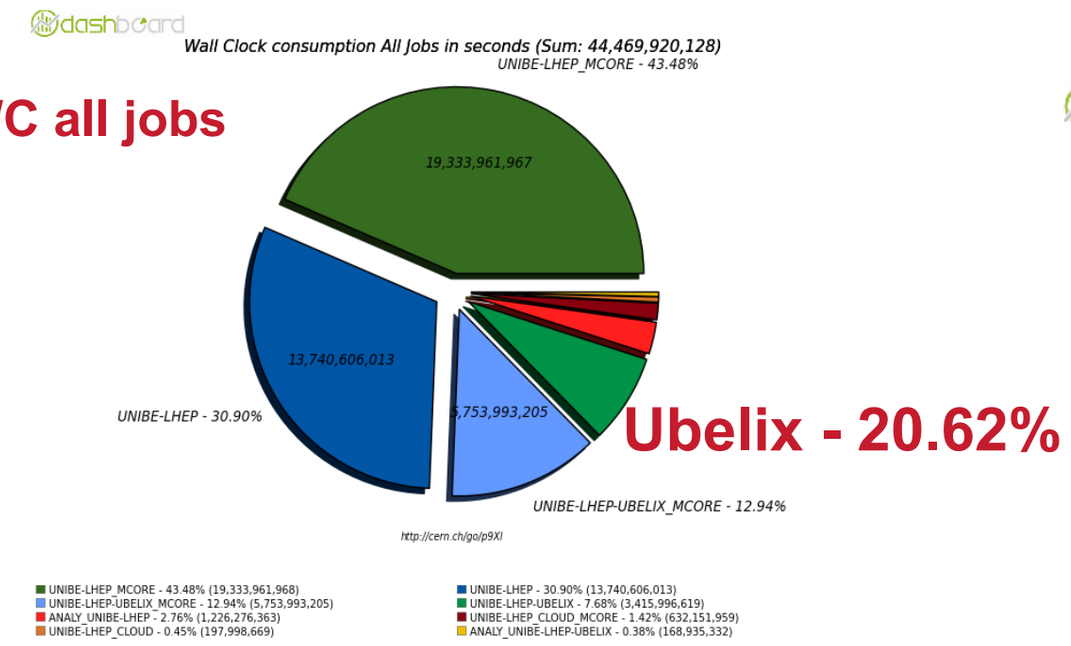
	HS06	TB
2016	10000	350
2017	15000	400
2018	18000	500

UNIBE-LHEP TIER 2 REPORT

Processing shares

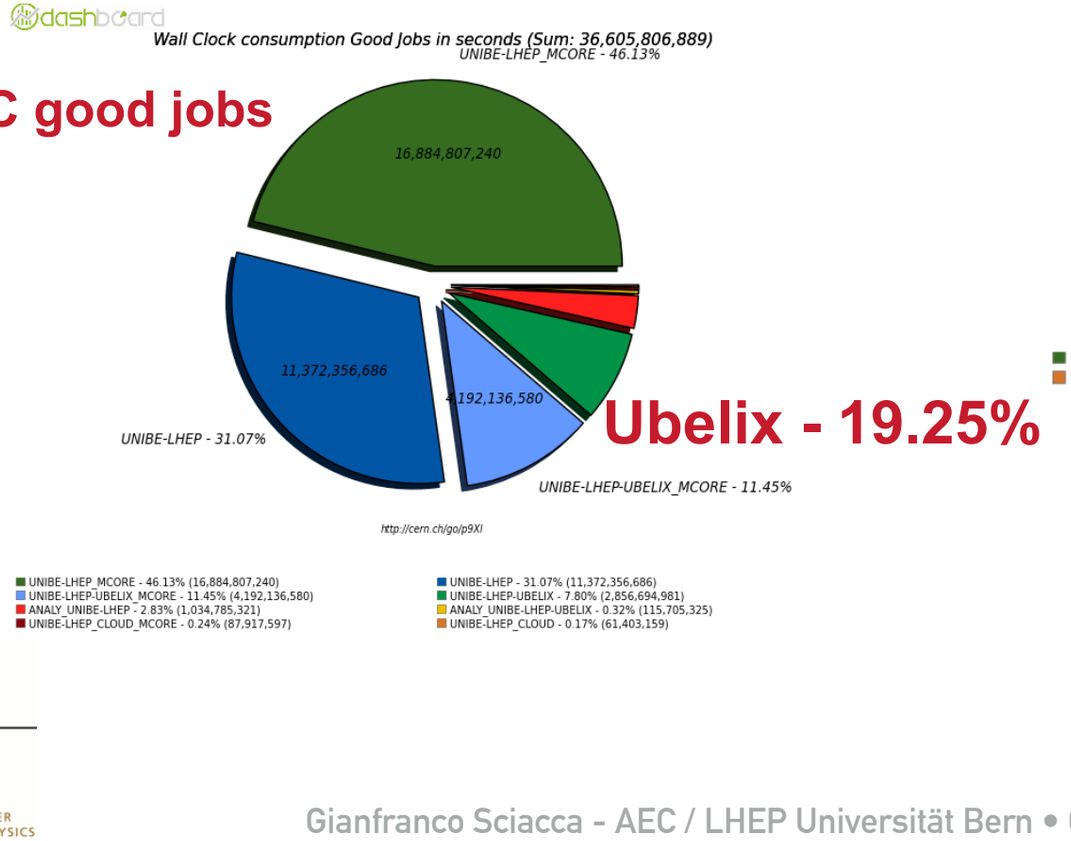
- **Ubelix estimated** in between 12% and 21% ***of the total installed capacity (2017)***

WC all jobs

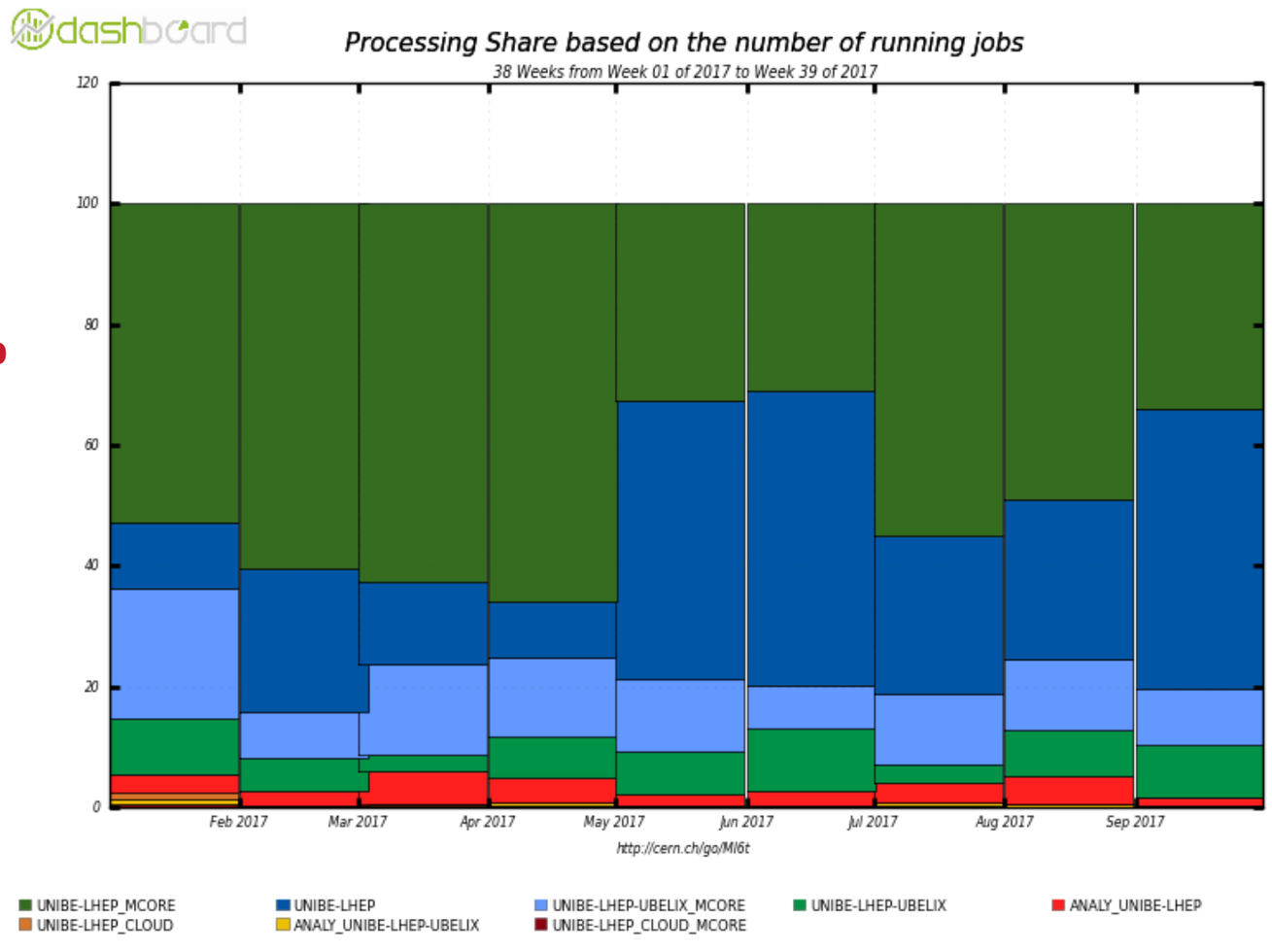


Ubelix - 20.62%

WC good jobs

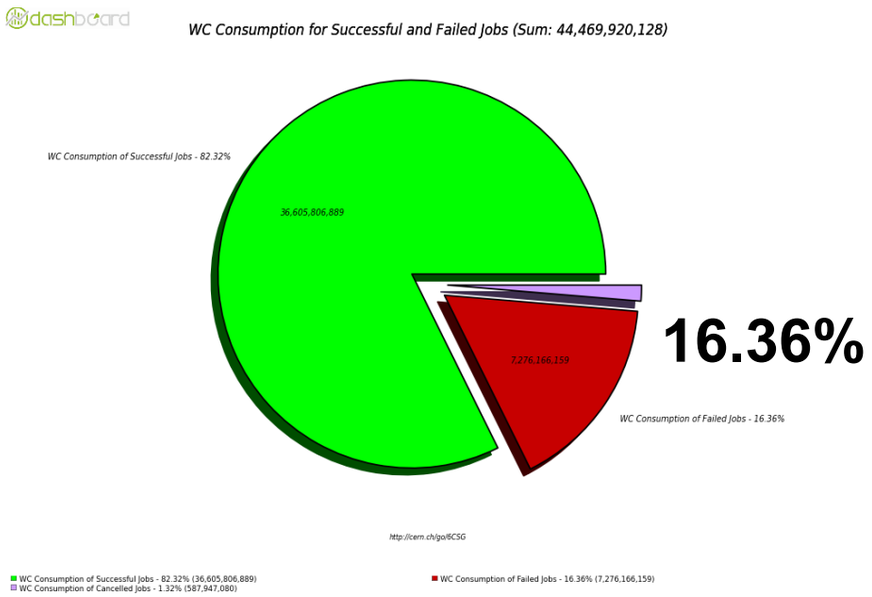


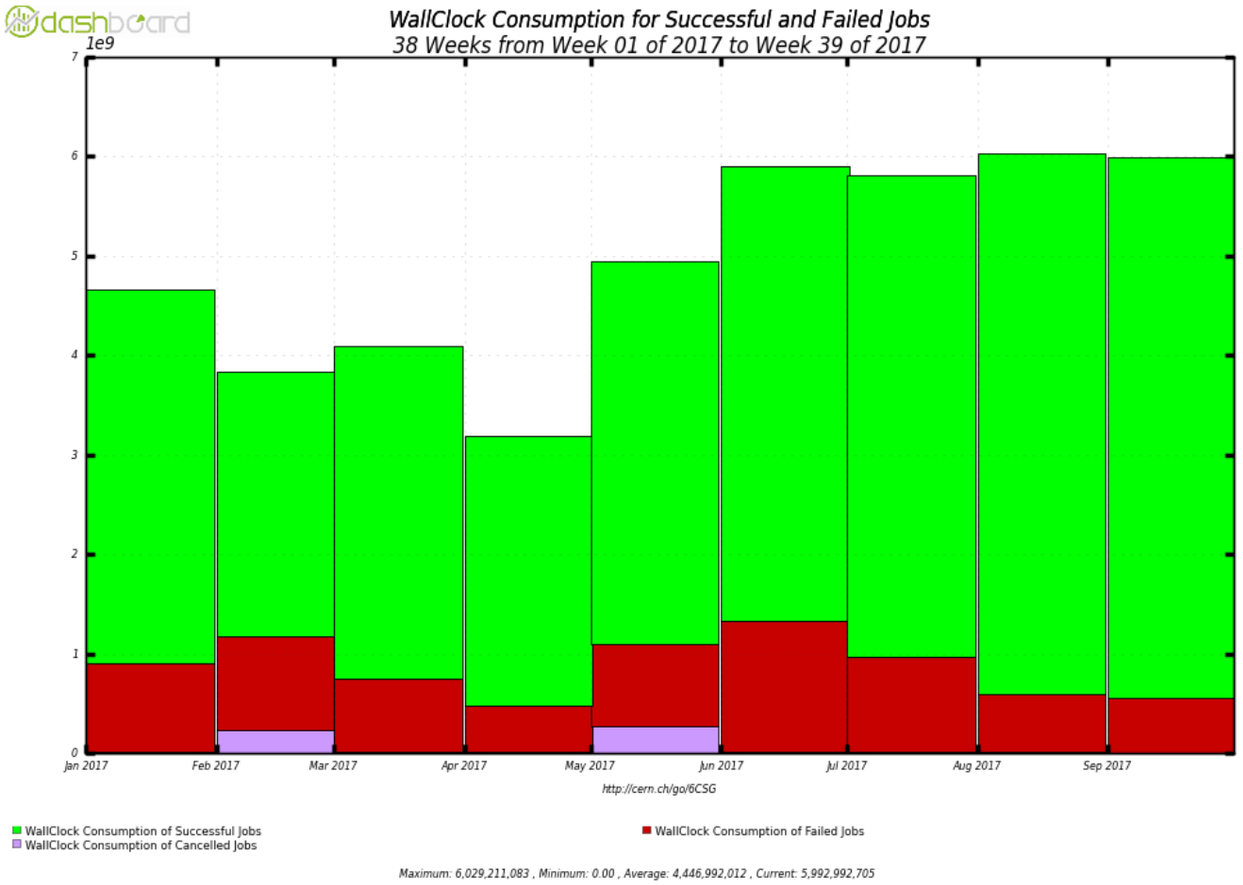
Ubelix - 19.25%

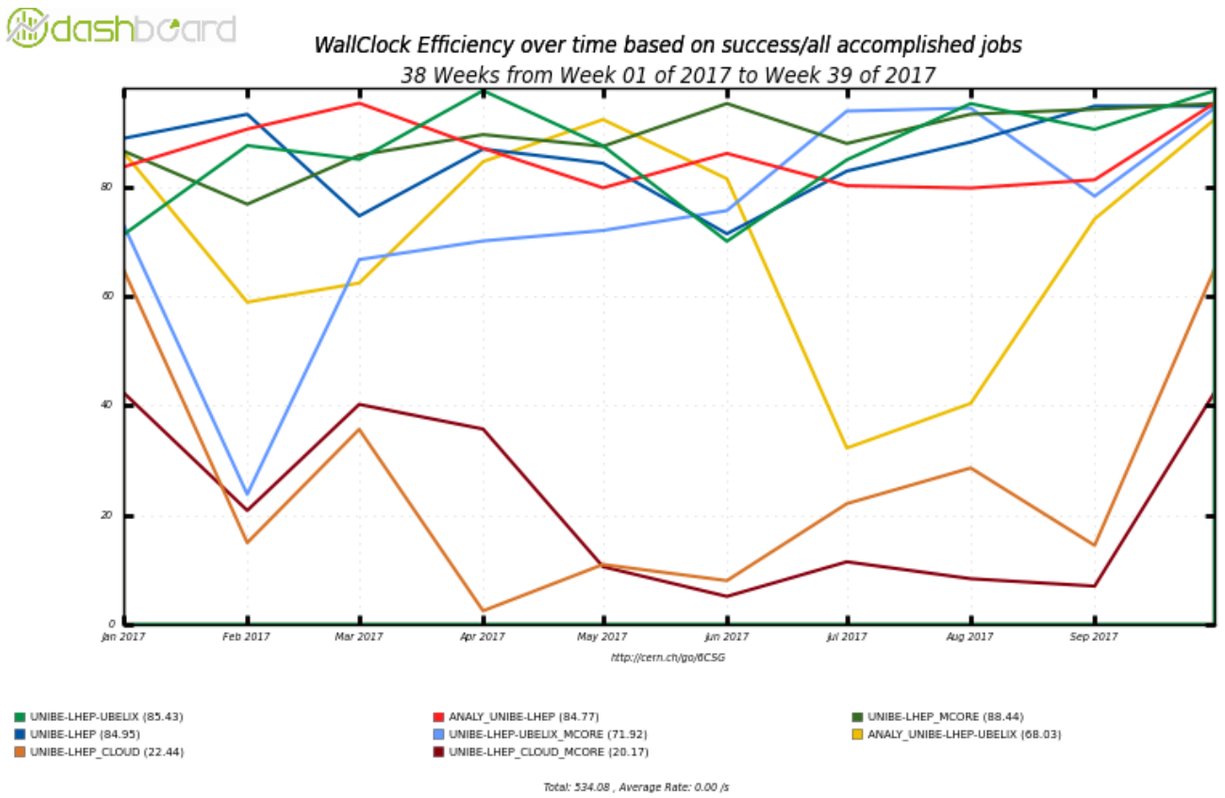


UNIBE-LHEP TIER 2 REPORT

Success vs fail WallClock efficiency

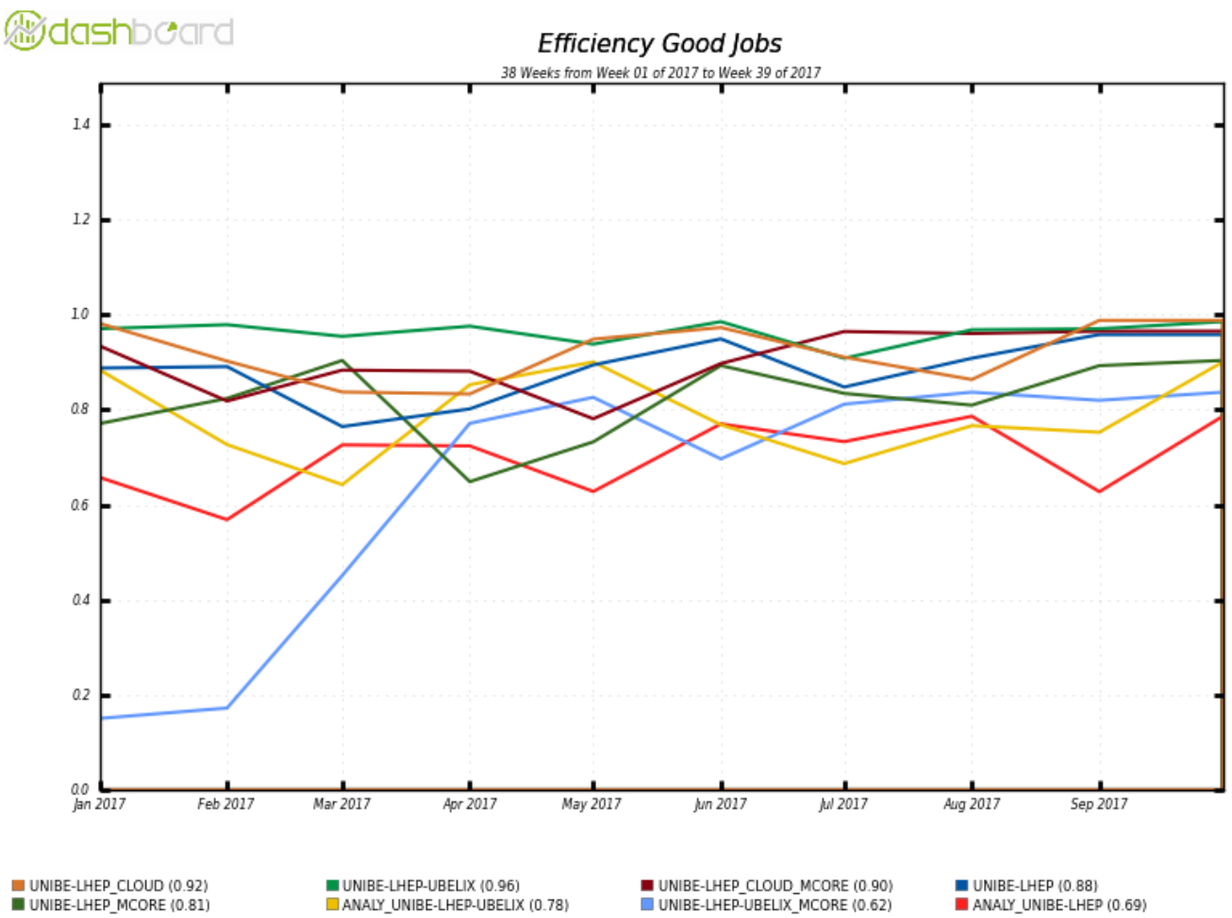


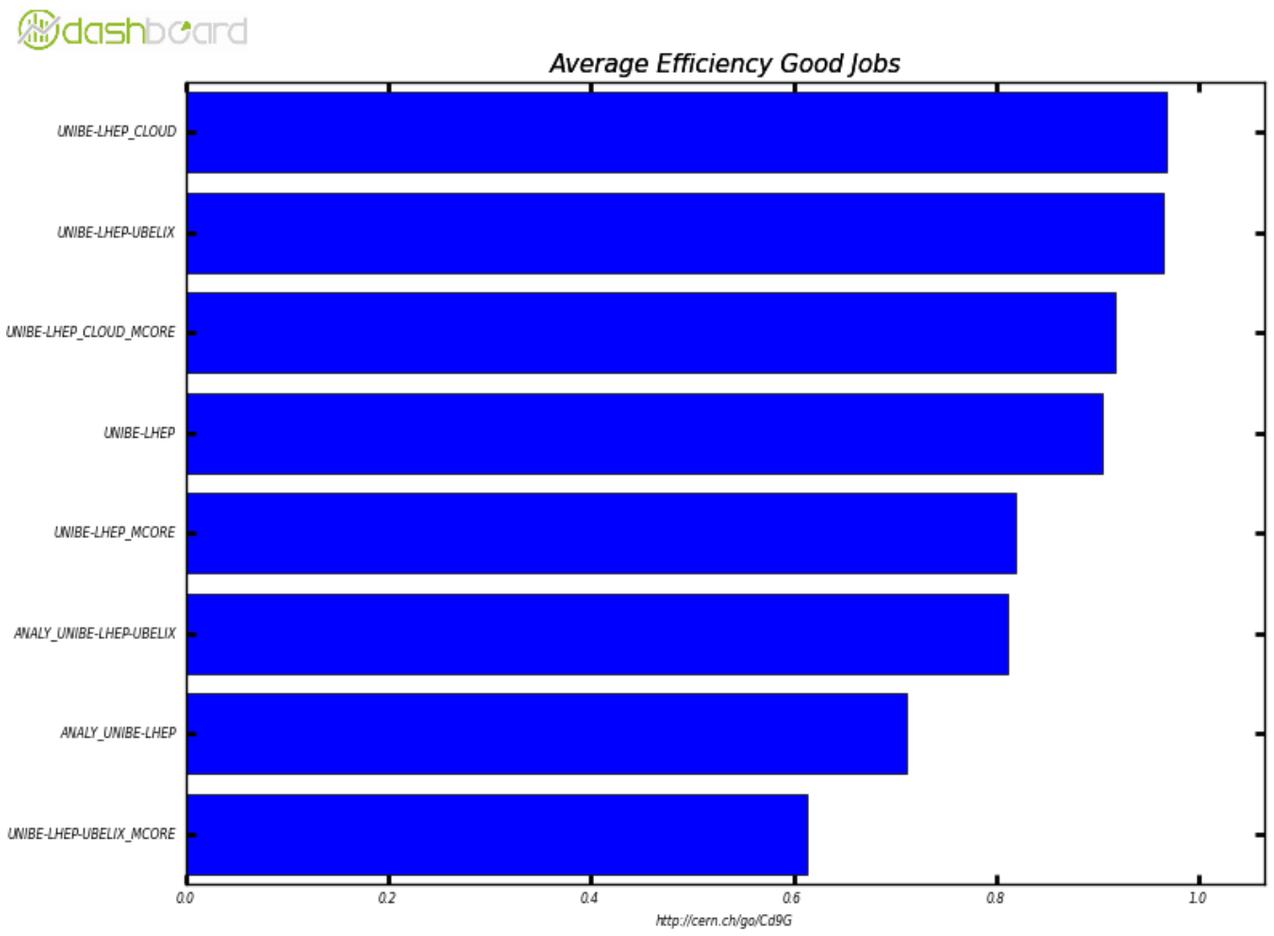




UNIBE-LHEP TIER 2 REPORT

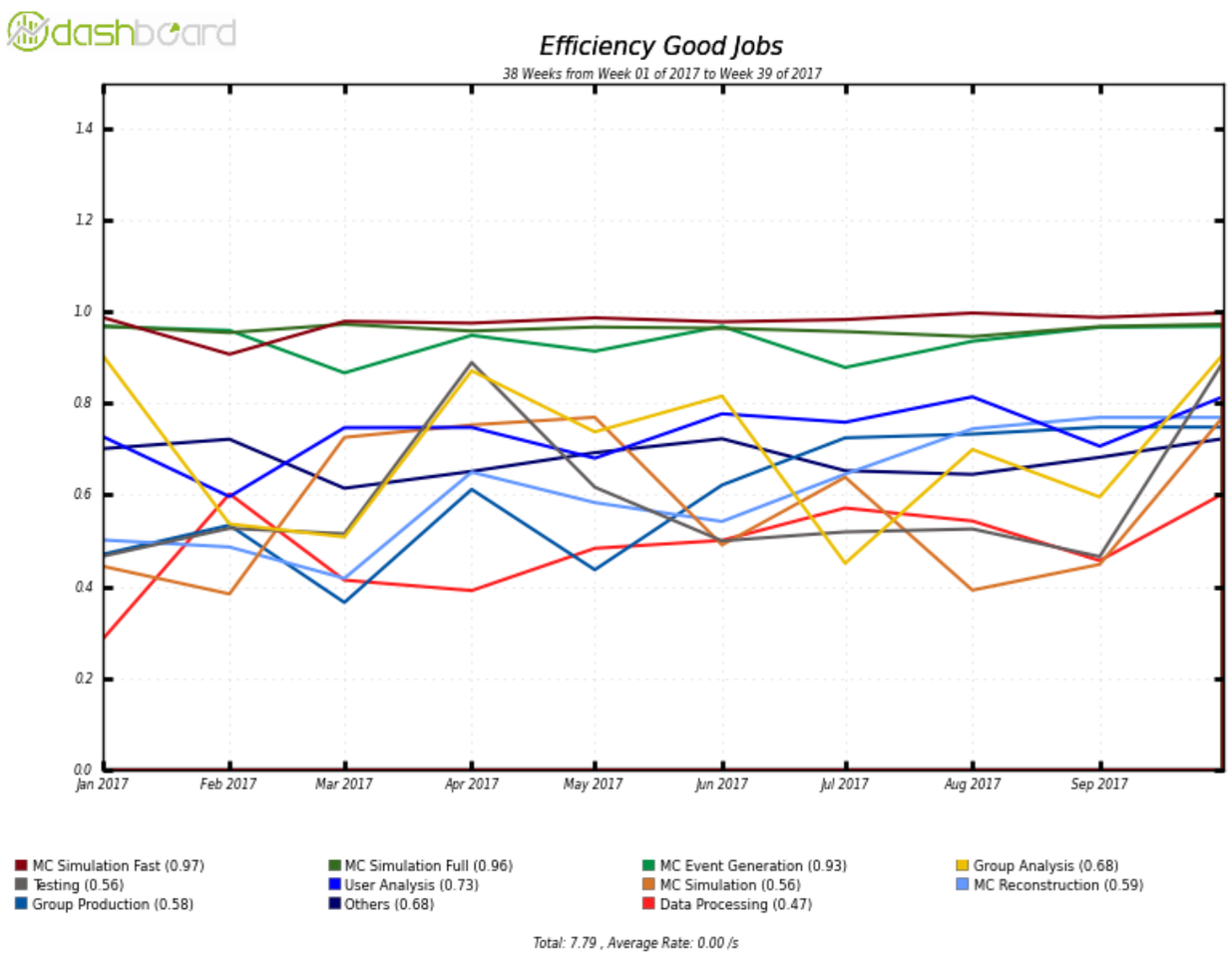
CPU / WallClock efficiency

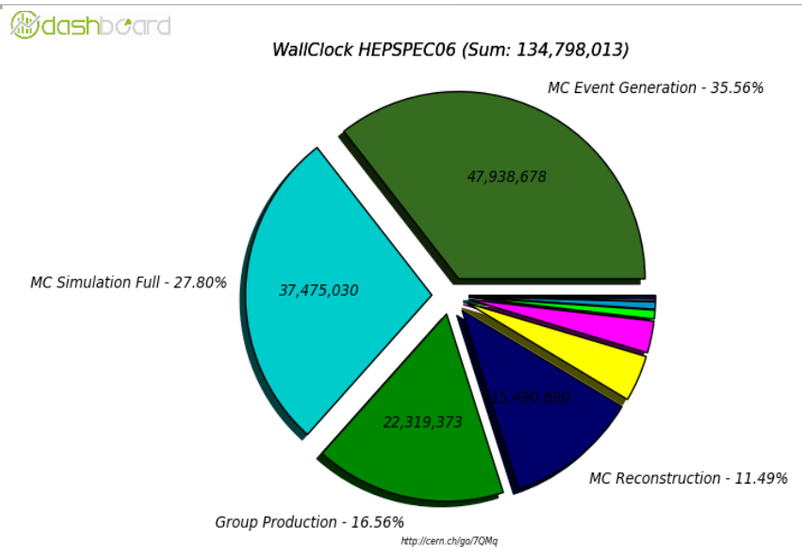


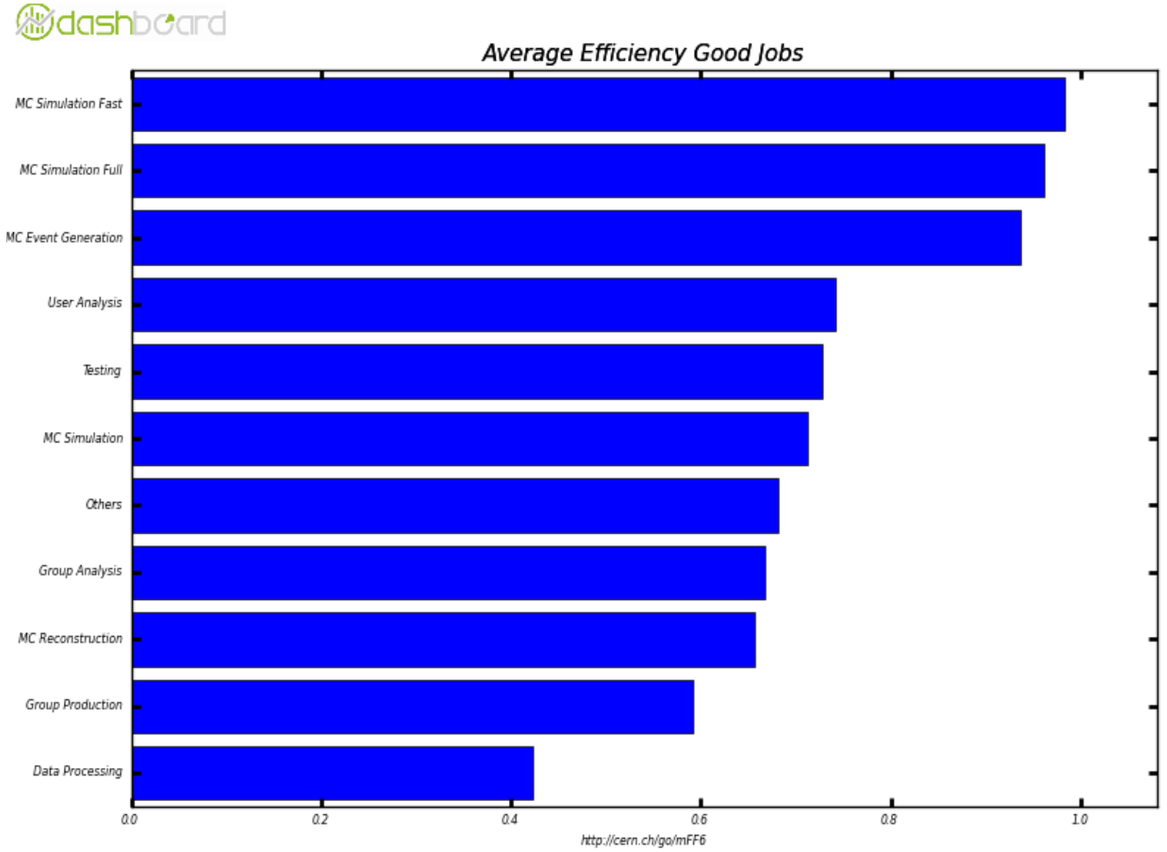


UNIBE-LHEP TIER 2 REPORT

CPU / WallClock efficiency by type of job

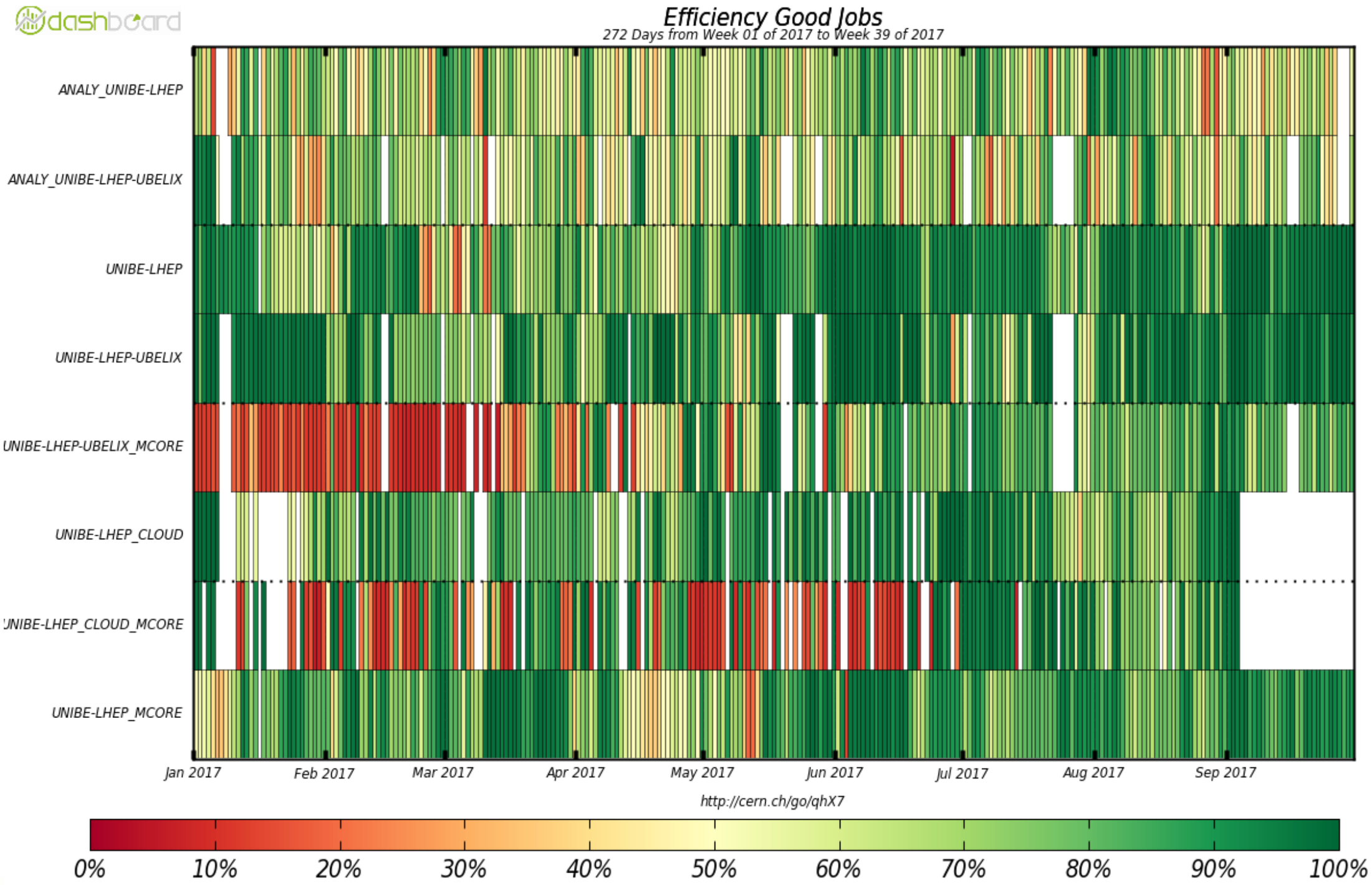






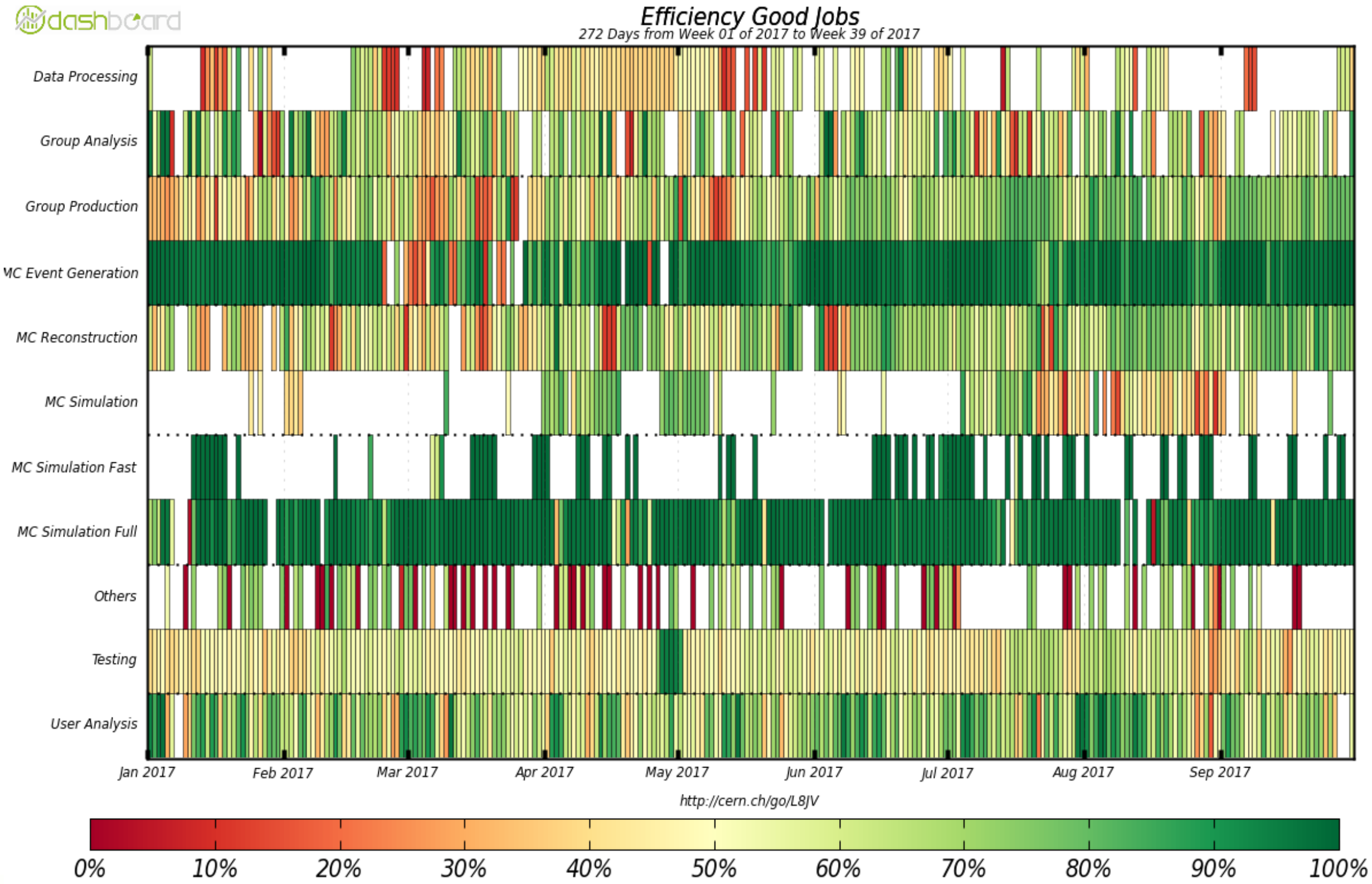
UNIBE-LHEP TIER 2 REPORT

CPU / WallClock efficiency by PanDA queue



UNIBE-LHEP TIER 2 REPORT

CPU / WallClock efficiency by ADC activity



UNIBE-LHEP TIER 2 REPORT

Recent and/or recurring issues

- **ARC middleware instabilities**
  - Several ARC bugs have afflicted us in the first part of the year. These were uncovered due to the new, heavier patterns of I/O and data staging
  - Eventually rooted out by end of May, thanks to the ARC sites in NDGF that helped as guinea pigs in the production environment
  - Version *5.3.1-1* turned out to bring back stability
  - Making extensive use of *gangliarc* plugins to detect issues early

- **LAN bottleneck on the ce02 cluster**
  - Also due to hugely increased staging and I/O activity
  - Upgraded the gateway to the LAN to 10GbE (and added lustre OSTs)

- **Lustre MDS occasionally lock up (mostly on ce02)**
  - Improved monitoring detects the failure early, so I can remedy quickly with reduced downtime for the service

- **SLURM Node Health Check a bit too eager**
  - Better safe than sorry. Still tuning it, but it has brought a large improvement (reflecting in reduced failure rates)

- **The SE stopped working for FTS very recently (failed x509 negotiation)**
  - due to *openssl* and *globus-gsi* libs on the head node (SLC5) no longer matching
  - prompted an upgrade of the head node to SLC6 in 2 days (after postponing it for 2 years : -)

UNIBE-LHEP TIER 2 REPORT

Summary and outlook

- **Contributing ~40% of the CH-ATLAS Tier-2 HS06 delivery**
- **Thanks to some long overdue bits of funding, we could stabilise operations a bit since spring**
- **Better reliability**
  - better monitoring, improved w.r.t. job failure rates
  - very little need of manual intervention
- **Availability maximised thanks to the redundant cluster setup**
  - can service or remedy one cluster, while the other still crunches data
- ***There is always room for improvement!***
- **We aim at a moderate growth over the next few years**
  - ~20% year-on-year, bar disasters
  - always on the look for opportunistic resources, since we have the tools and experience for making proficient use of them with relatively little effort