UNIBE-LHEP TIER 2 REPORT

Gianfranco Sciacca

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

CHIPP-CSCS face 2 face - 13 September 2019

TIER-2 COMPUTE IN BERN

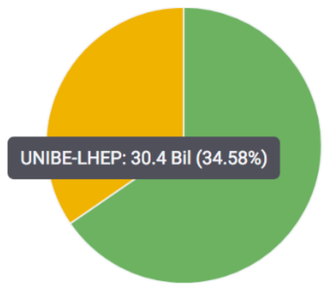
OVERVIEW

- **Resources at LHEP (T2 / T3) and on the shared University cluster Ubelix**
  - Pursued since years the T2 / T3 co-location strategy to maximise return for the investment:
    - Ability to absorb chaotic / urgent Tier-3 usage peaks
    - Never idle CPU cycles
  - T3 users: ATLAS local, uboone, t2k.org (probably DUNE in the near future)
- **Q1/2 2019**
  - delivered 35% of the Swiss Tier-2 ATLAS wallclock
  - processed 46% of the total Swiss Tier-2 events
- **Very good CPU efficiency (81%)**

TIER-2 COMPUTE IN BERN

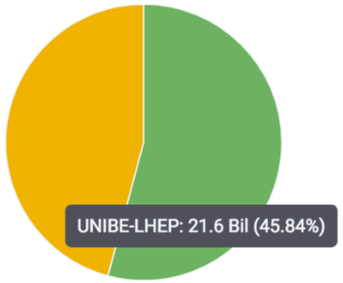
Q1/2 STATISTICS

Wallclock Consumption: All jobs in Seconds



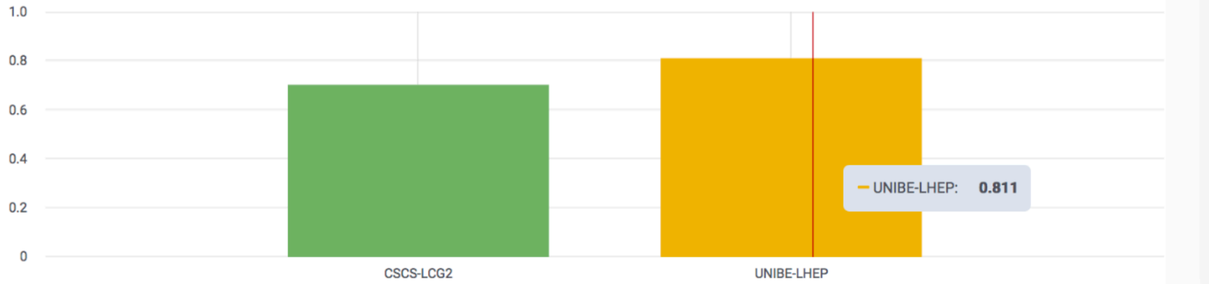
	total	percentage
CSCS-LCG2	57.5 Bil	65%
UNIBE-LHEP	30.4 Bil	35%

NEvents Processed - in percentage



	current	percentage
CSCS-LCG2	25.5 Bil	54%
UNIBE-LHEP	21.6 Bil	46%

Average CPU Efficiency Good jobs

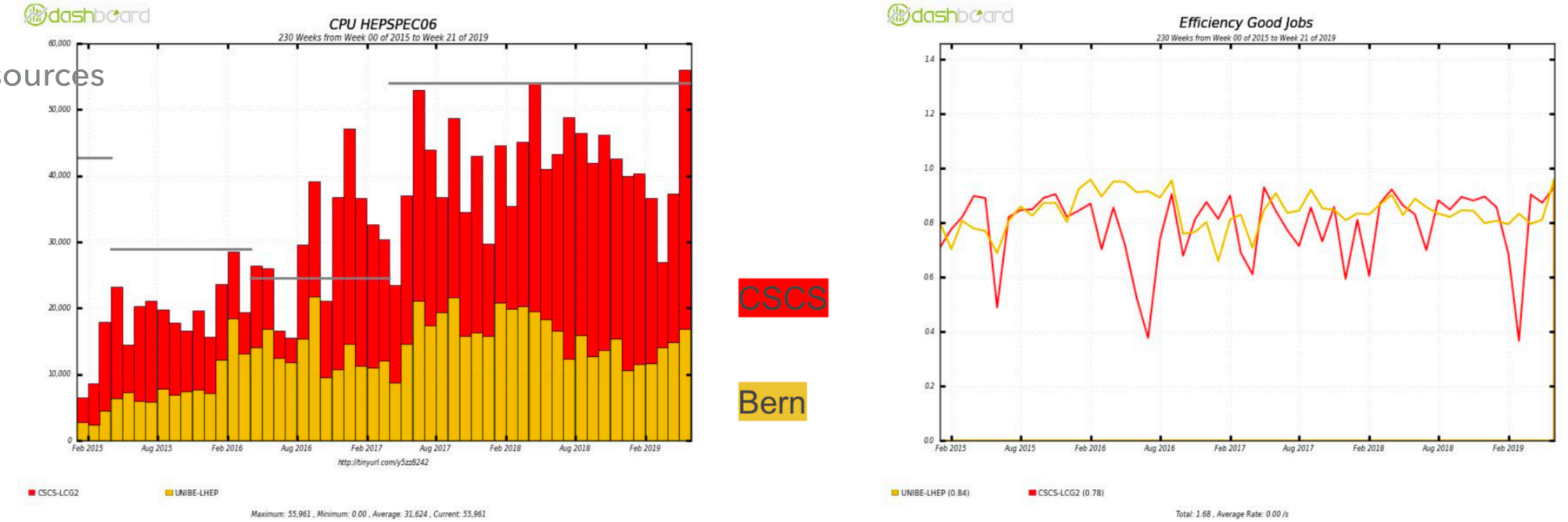


	min	max	avg	total
UNIBE-LHEP	0	1.377	0.811	3.40605679 K
CSCS-LCG2	0	0.971	0.703	2.95130253 K

TIER-2 COMPUTE IN BERN

OVERVIEW

- **Continuity and stability of Tier-2 CPU delivery despite modest investments**
- **In-house ability to deploy and/or exploit over the years:**
  - own hardware
  - CHIPP hardware retired at CSCS
  - Ubelix cluster
  - cloud resources
  - opportunistic resources



4.5 year historical view for the Tier-2 up to Q2 2019 (NOTE: Tier-3 cycles not included)

TIER-2 COMPUTE IN BERN

OVERVIEW

- **Global trend in WLCG is to extend hardware lifetime to gain (considerably) in resources within flat budgets**
- **Favourable conditions in Bern:**
  - Local expertise and deep involvement with the ATLAS computing community at high level
  - Infrastructure on premises and in-kind contribution from the University and LHEP:
    - Server room, cooling, electricity, LHEP workshops (electric, mechanical)
- **We have been extending CHIPP FLARE funded hardware lifetime <u>since 10 years</u>**
- **With the transition to HPC for the CSCS Tier-2, this will be lost**
  - The impact on resource availability for the future years is considerable and must be properly factored in the ***<u>future computing strategy for the HL-LHC</u>***

TIER-2 COMPUTE IN BERN

CURRENT STATUS - LHEP

- **Some of the decommissioned Phoenix nodes moved to Bern (with some delay)**
- **We had given a heads up at the f2f 1 year ago to give us details as soon as possible and explicitely asked "**for details *_now_*"** **at the Jan 2019 f2f**
  - In order to prepare the infrastructure on premises
  - We got the first details on 20th May
  - **This meant a considerable delay w.r.t. the pledge period**
    - We could not prepare the infrastructure as needed and been forced to improvise to get back online in some form
    - Brute fact: **we are underdelivering by large (~35%)**
    - The tentative plan is to make up for it up to Q4 2019 (more later)

TIER-2 COMPUTE IN BERN

CURRENT STATUS - LHEP

- **Upgraded ARC CE to Centos 7 (still on ARC 5.4)**
- **Upgraded to ROCKS 7, Centos 7, SLURM 18.08.8, Lustre 2.12.2**
- **Redeployed all ~recently purchased WNs and the Phonenix WNs**
  - ~15% could not yet be installed (various problems) => 624 cores
  - ~10% installed but still off pending power infrastructure works => 320 cores
  - 4032 online cores (736 to be rescued following Lustre and IB upgrades)

  - **WiP** => we should reach just short of 4.7k cores => **53.9 kHS06**
    pledged => **42 kHS06**

TIER-2 COMPUTE IN BERN

CURRENT STATUS - LHEP

- **Many concurrent issues to deal with during and after re-deployment**
  - **Slurm tuning** (job timeouts, OOM, cgroups, etc)
  - **Slow ARC infosys** (verbose slapd logging)
  - **a-rex periodically dying** (ongoing)
  - **Downtime for RAM upgrade on Lustre OSS servers and IB upgrade**
    - Problematic draining, plenty of stray jobs running for days
    - Brute force killing them broke most nodes (some yet to be recovered)
  - **Switch to pilot2** (early bugs)
  - **Switch to singularity** (buggy binary in cvmfs cached on nodes)
  - **aCT lost track of all jobs** (issue with slow eos at CERN)
  - **…**

TIER-2 COMPUTE IN BERN

CURRENT STATUS - UBELIX

- **Upgraded ARC CE to Centos 7 (still on ARC 5.4)**
- **up to 600 cores + up to 1200 pre-emptable** (was 600 up to 7th July)

  ⬇

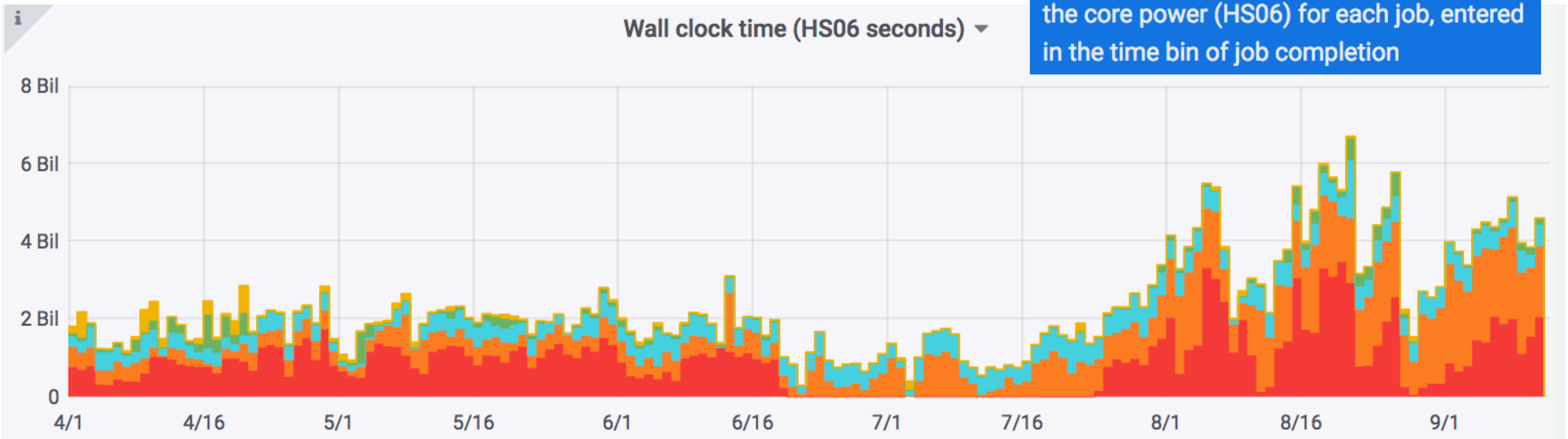
  9.3 kHS06

- **Cooling upgrade in server room about to begin**
  - Upgrade to water cooling -> limited capacity during the works
  - 4-day shutdown starting 23rd Sep
  - Then ~2160 running cores (out of 7408) => ~30% until 8th Nov (tbc)
  - ATLAS can expect a proportional reduction of the number of cores

TIER-2 COMPUTE IN BERN

Q2/3 (PARTIAL) STATISTICS: WC HS06

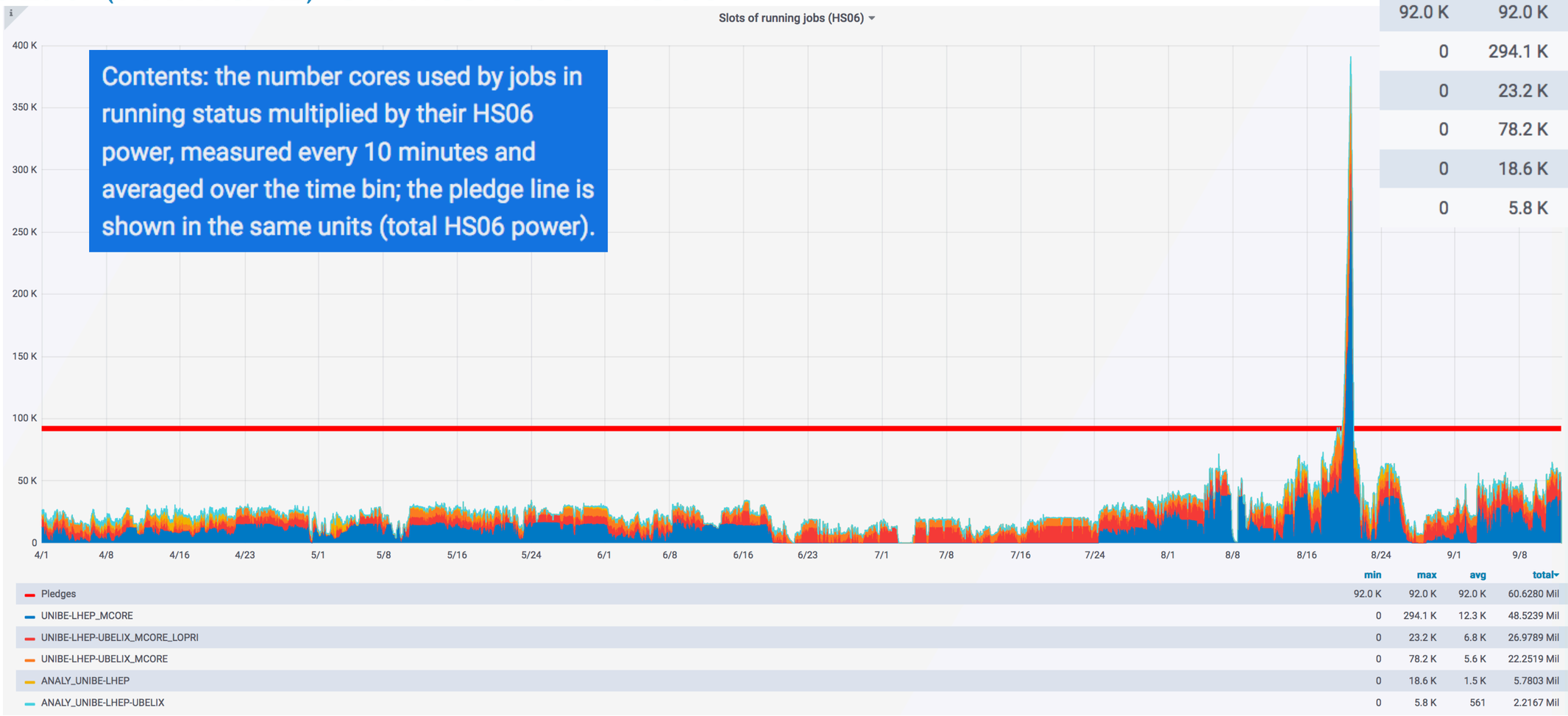
Contents: total time in running status multiplied by the number of used cores and the core power (HS06) for each job, entered in the time bin of job completion



UNIBE-LHEP_MCORE	0	3.469 Bil	885 Mil	146.061 Bil
UNIBE-LHEP-UBELIX_MCORE_LOPRI	0	2.562 Bil	807 Mil	133.123 Bil
UNIBE-LHEP-UBELIX_MCORE	0	1.501 Bil	454 Mil	74.831 Bil
ANALY_UNIBE-LHEP	0	869 Mil	97 Mil	15.992 Bil

TIER-2 COMPUTE IN BERN

Q2/3 (PARTIAL) STATISTICS: HS06



min	max	avg	total
92.0 K	92.0 K	92.0 K	60.6280 Mil
0	294.1 K	12.3 K	48.5239 Mil
0	23.2 K	6.8 K	26.9789 Mil
0	78.2 K	5.6 K	22.2519 Mil
0	18.6 K	1.5 K	5.7803 Mil
0	5.8 K	561	2.2167 Mil

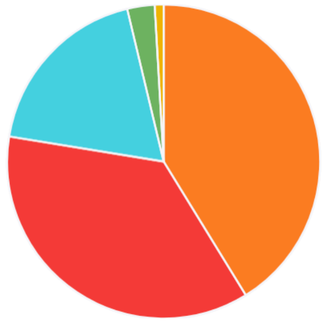
27 K

Pledge: 42 K

TIER-2 COMPUTE IN BERN

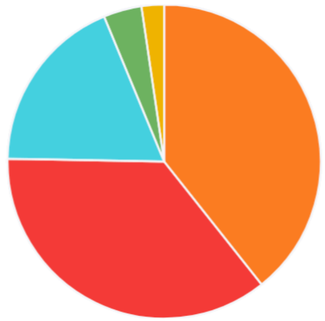
Q2/3 (PARTIAL) STATISTICS: PROCESSING SHARES

Wallclock Consumption: Successful jobs in Seconds



	current	percentage
UNIBE-LHEP-UBELIX_MCORE_LOPRI	11.64 Bil	41%
UNIBE-LHEP_MCORE	10.27 Bil	36%
UNIBE-LHEP-UBELIX_MCORE	5.28 Bil	19%
ANALY_UNIBE-LHEP	799 Mil	3%
ANALY_UNIBE-LHEP-UBELIX	261 Mil	1%

Wallclock Consumption: All jobs in Seconds

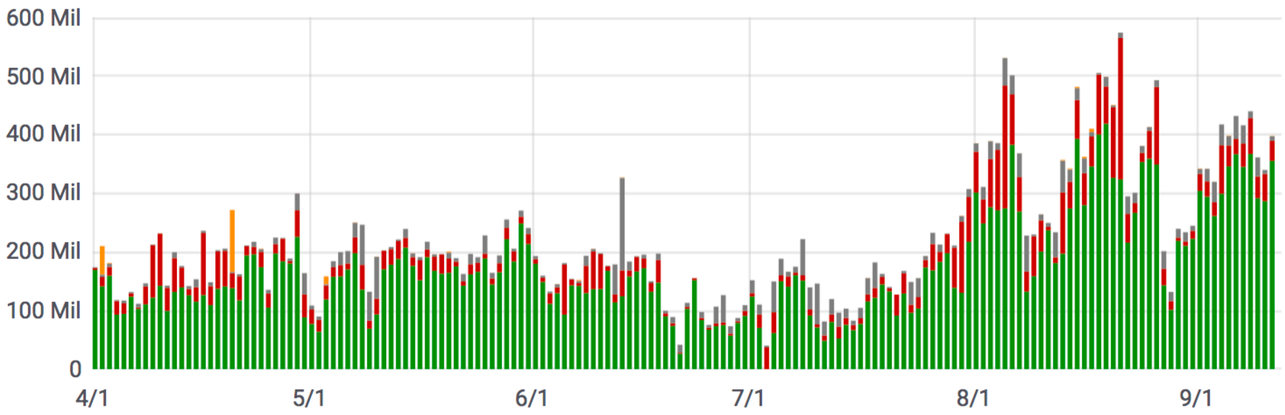


	total	percentage
UNIBE-LHEP-UBELIX_MCORE_LOPRI	14.34 Bil	39%
UNIBE-LHEP_MCORE	13.13 Bil	36%
UNIBE-LHEP-UBELIX_MCORE	6.73 Bil	18%
ANALY_UNIBE-LHEP	1.438 Bil	4%
ANALY_UNIBE-LHEP-UBELIX	849 Mil	2%

TIER-2 COMPUTE IN BERN

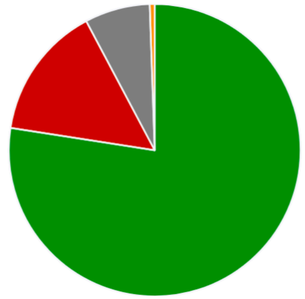
Q2/3 (PARTIAL) STATISTICS: SUCCESS VS FAIL WC EFFICIENCY

WallClock Consumption of Successful and Failed Jobs - Time Stacked Bar Graph



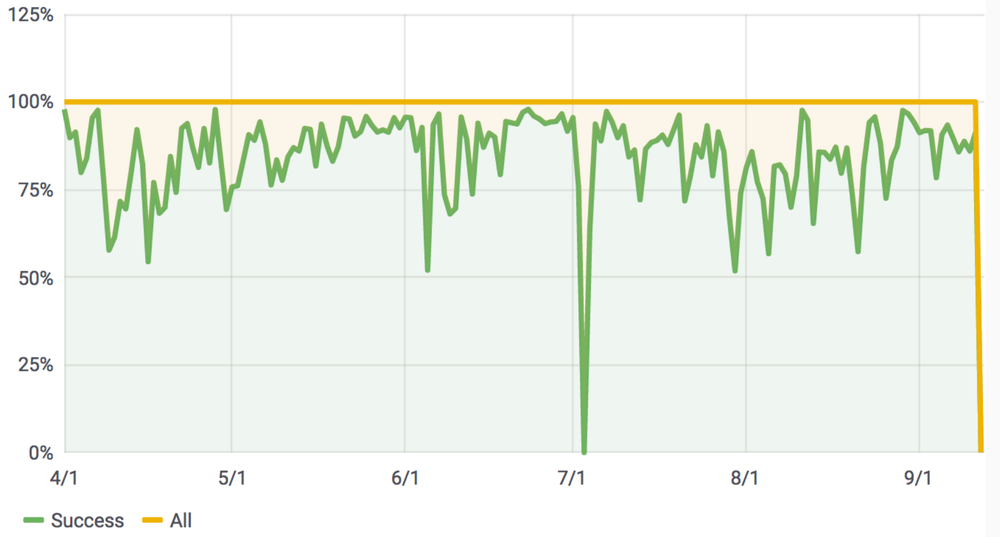
	min	max	avg	total
finished	0	419 Mil	170 Mil	28.120 Bil
failed	0	241 Mil	33 Mil	5.365 Bil
closed	0	158 Mil	16 Mil	2.645 Bil

WallClock Consumption of Successful and Failed Jobs - Pie Graph



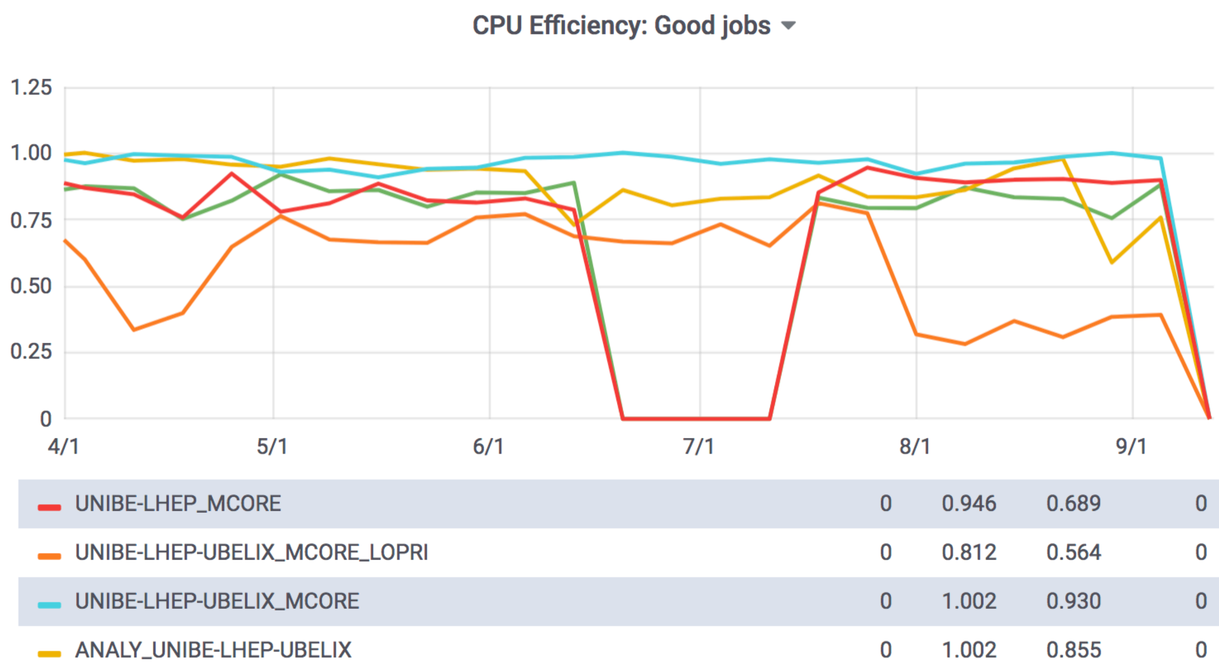
	current	percentage
finished	28.3 Bil	77%
failed	5.40 Bil	15%
closed	2.641 Bil	7%
cancelled	200 Mil	1%

WallClock Efficiency over time based on success/all accomplished jobs

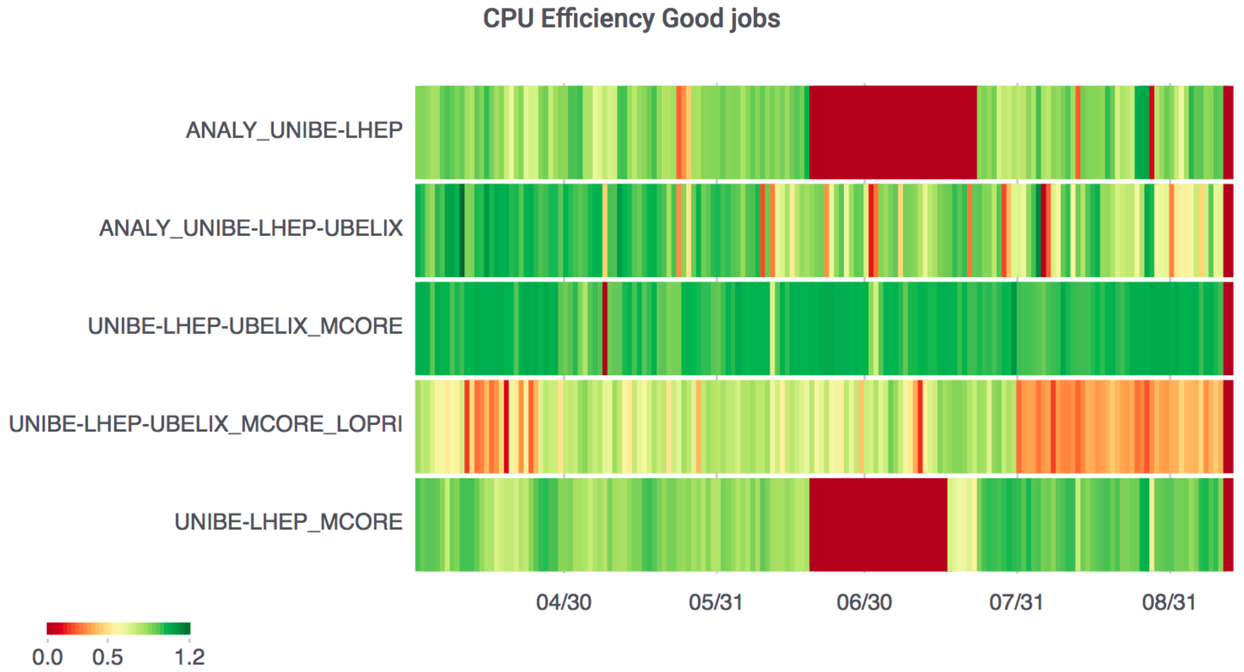


TIER-2 COMPUTE IN BERN

Q2/3 (PARTIAL) STATISTICS: WC EFFICIENCY FOR GOOD JOBS



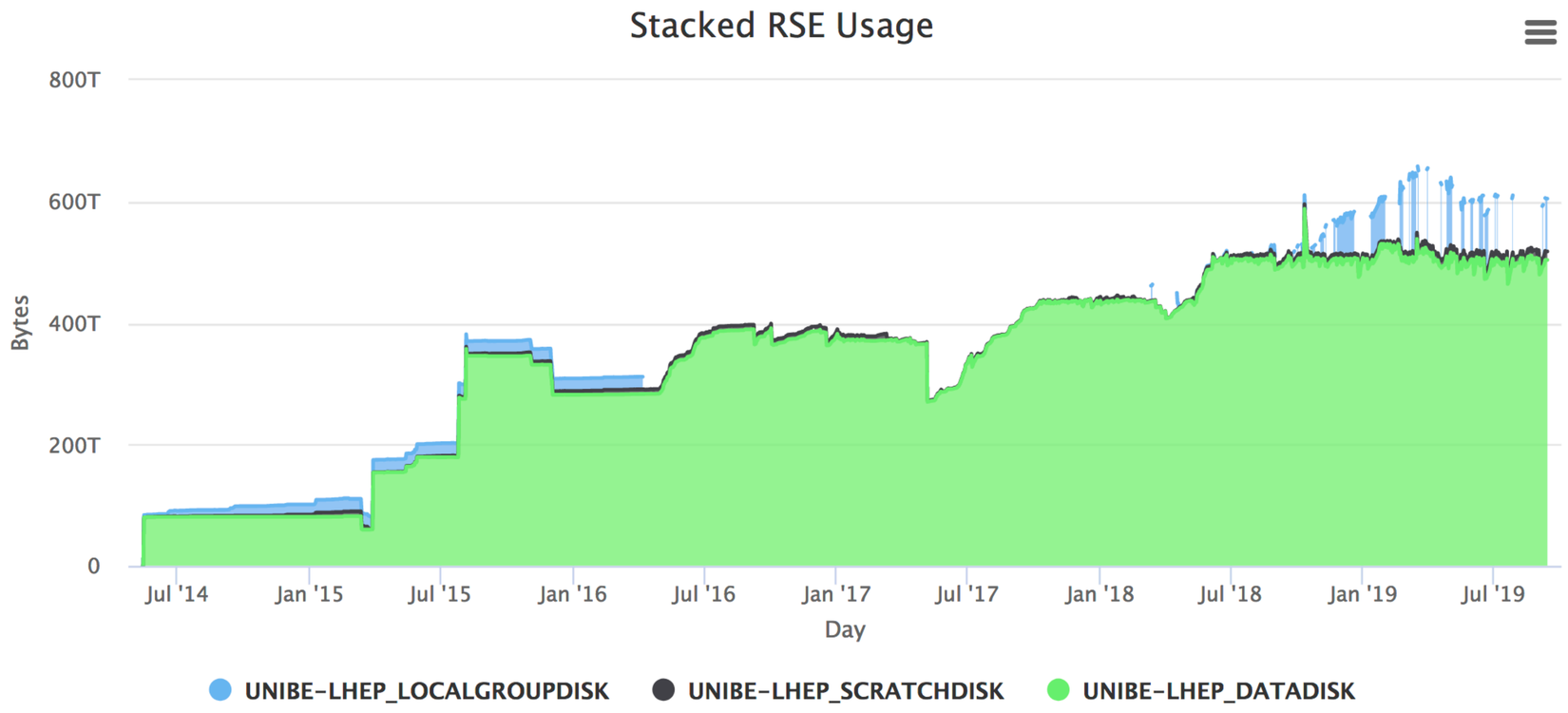
UNIBE-LHEP_MCORE	0	0.946	0.689	0
UNIBE-LHEP-UBELIX_MCORE_LOPRI	0	0.812	0.564	0
UNIBE-LHEP-UBELIX_MCORE	0	1.002	0.930	0
ANALY_UNIBE-LHEP-UBELIX	0	1.002	0.855	0



TIER-2 STORAGE IN BERN

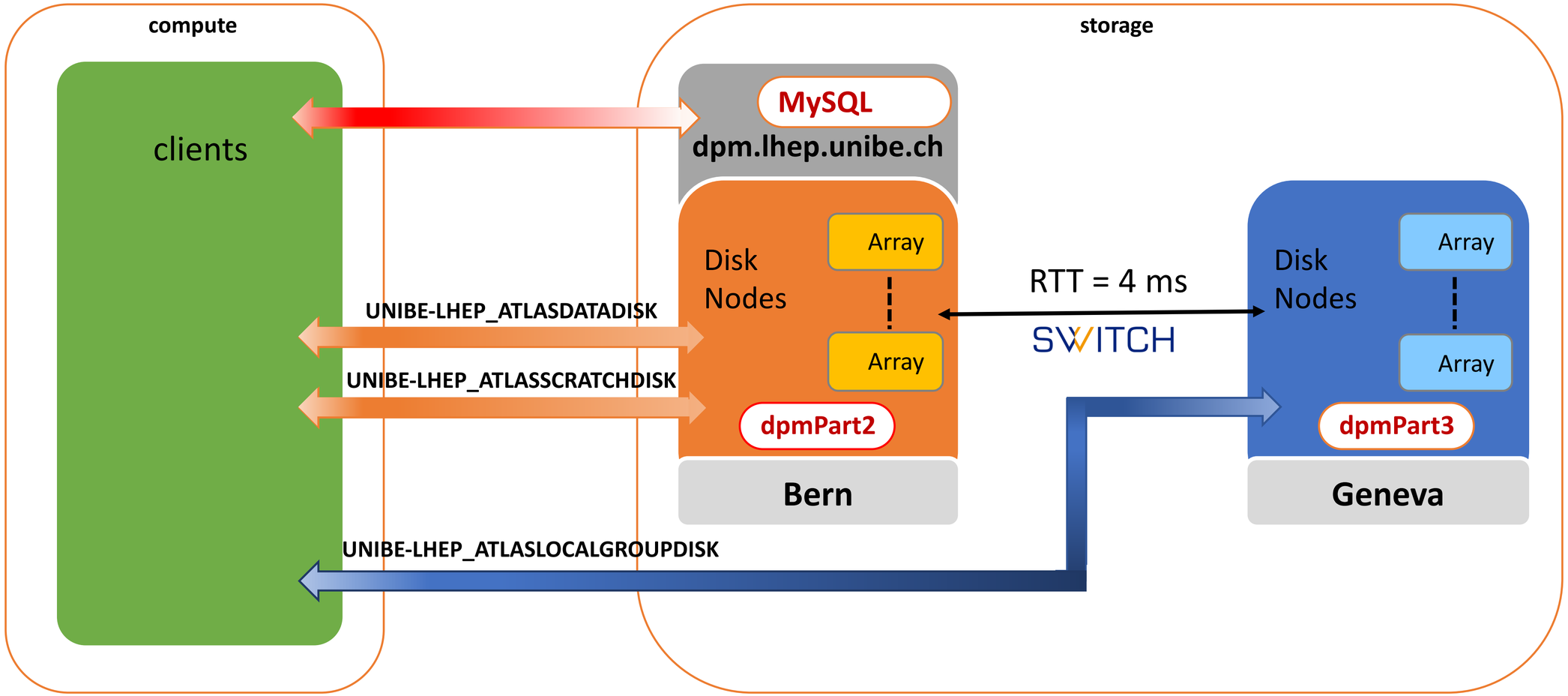
STATUS

- **1.2 PB, 973 TB allocated to tokens in DPM**
- **Distributed between Bern and Geneva**
  - Seamless operation so far



TIER-2 STORAGE IN BERN

LAYOUT



TIER-3 COMPUTE AND STORAGE IN BERN

CURRENT STATUS

- **LHEP clusters and Ubelix available to Tier-3 users (co-located Tier-3)**
  - ATLAS, T2K, MicroBooNE (DUNE coming soon)

- **Grid DPM Storage Element**
  - ATLASLOCALGROUPDISK (T2 / T3): reserved to ATLAS Swiss users => no burden on CSCS
  - SE also used by MicroBooNE

- **Additional Tier-3 only resources:**
  - Interactive processing servers at LHEP (with local storage for faster processing)
  - ~400TB of local NFS storage
  - Grid clients and software distribution via CVMFS
  - Resources heavily used

- **Moving in the direction of harmonising resources and sharing them between the different experiments**

BACKUP

ATLAS ONLY CPU and Disk at CSCS and Bern

Total installed unless marked as pledged — Pledges are pledge aspirations

	2017	2018	2019	2020	2021	2022	2023	2024
CPU in kHS06								
CSCS Phoenix	27.6	27.6	0.0	0.0	0.0	0.0	0.0	0.0
CSCS HPC	8.0	16.9	55.6	69.5	90.4	117.5	152.8	198.6
Bern	20.0	20.0	42.0	42.0	38.0	48.0	44.0	44.0
TOTAL ATLAS	55.6	64.5	97.6	111.5	128.4	165.5	196.8	242.6
TARGET PLEDGE	50.0	62.6	78.2	97.7	122.2	152.7	190.9	238.6
Disk in PB								
CSCS	1.6	1.9	2.3	2.8	3.6	4.7	6.1	7.9
Bern	0.4	0.5	0.6	0.8	1.0	1.3	1.7	2.1
TOTAL ATLAS	2.0	2.4	2.9	3.6	4.6	6.0	7.8	10.0
TARGET PLEDGE	2.3	2.7	3.2	3.9	4.8	5.9	7.4	9.6