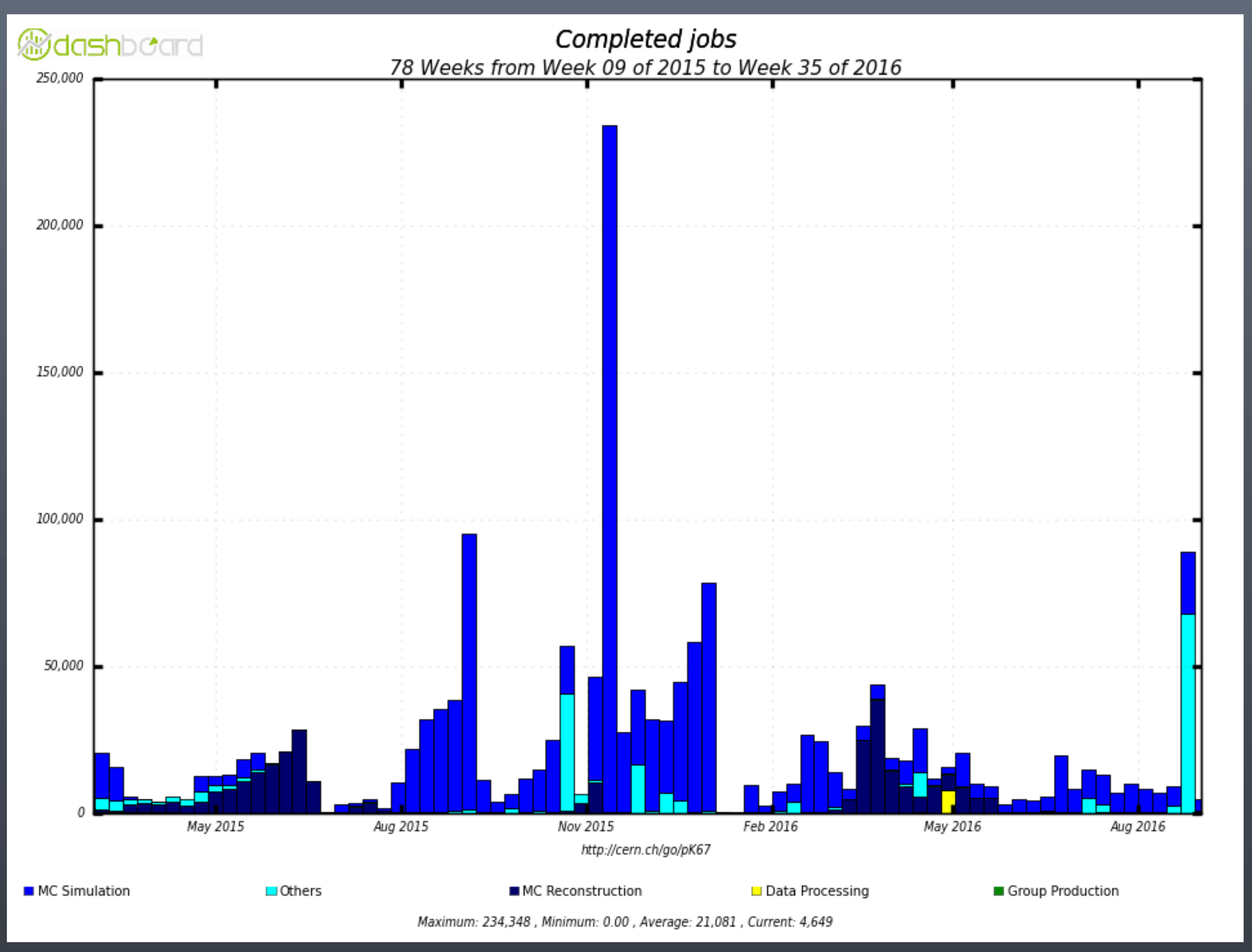
Sim@P1 project – CHEP2016

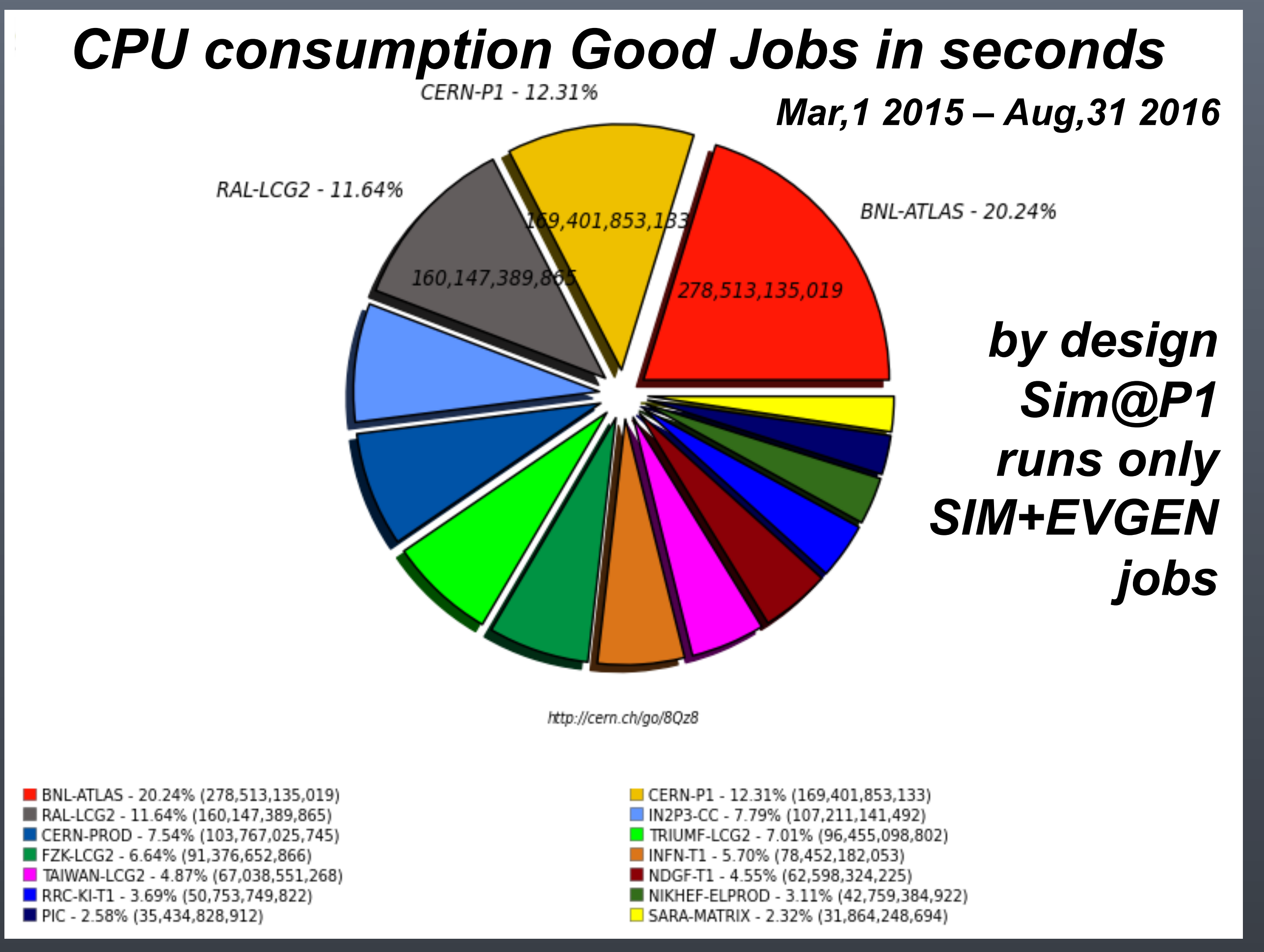
Evolution and experience with the ATLAS simulation at Point1 project

S. Ballestrero [1], F. Brasolin [3], D. Fazio [2], A. Di Girolamo [2], T. Kouba [5], C.J. Lee [2,9],
D.A. Scannicchio [4], J. Schovancová [2], M.S. Twomey [6], F. Wang [7], A. Zaytsev [8]

The simulation in Point1 project, based on OpenStack, uses in an opportunistic way the resources of the TDAQ High Level Trigger (HLT) farm of the ATLAS experiment. More than 2300 compute nodes (CNs) running up to 70k cores are exploited for running event generation and Monte Carlo production jobs, mostly CPU and not I/O bound, for a maximum of 5k 8-core parallel running jobs



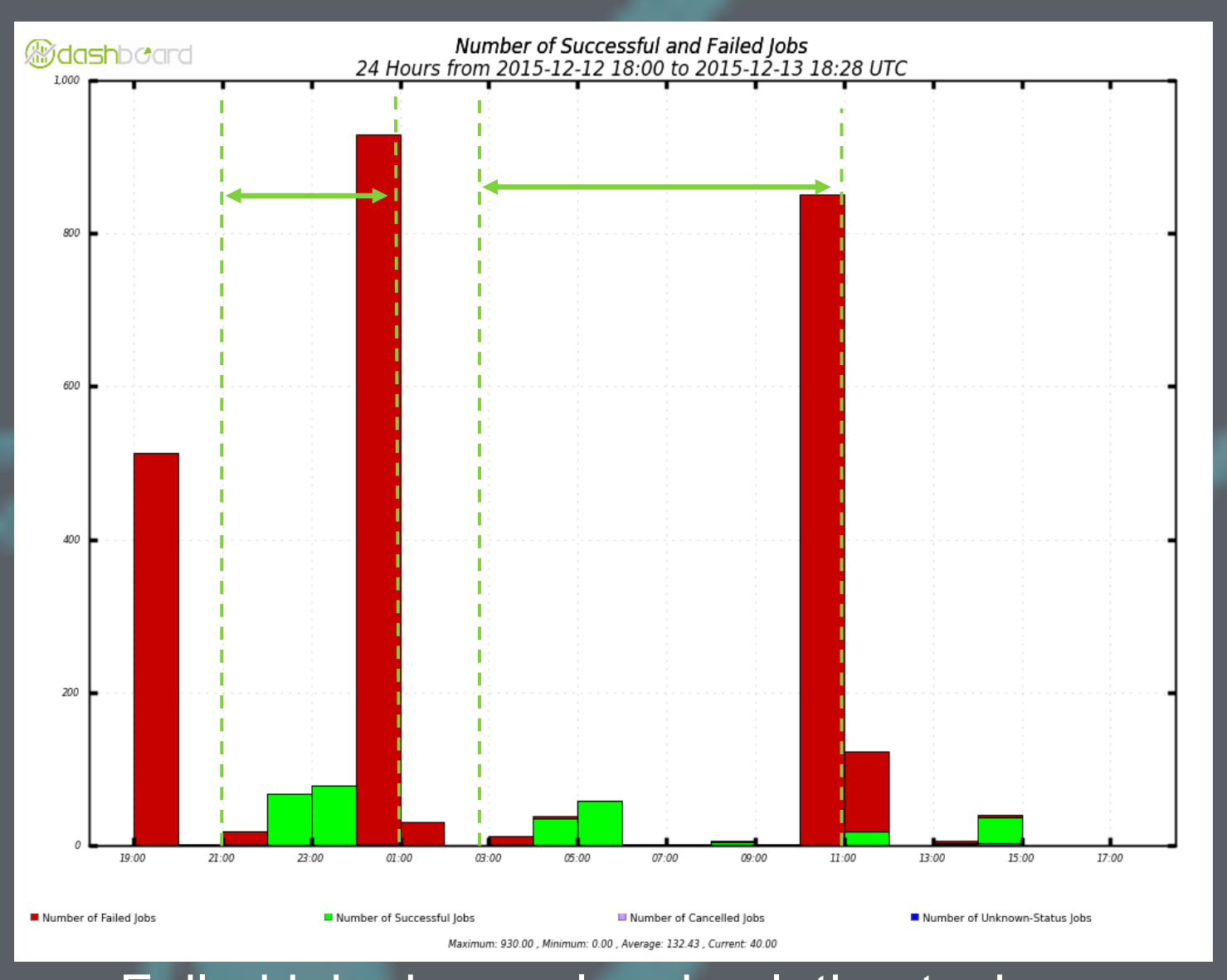
After the setup phase, during 2015 the Sim@P1 was reconfigured to accept multicore jobs and to help with MC reconstruction jobs in certain periods.
It delivered more than 46 million CPU-hours and it generated more than 0.5 billion Monte Carlo events since Mar 1, 2015



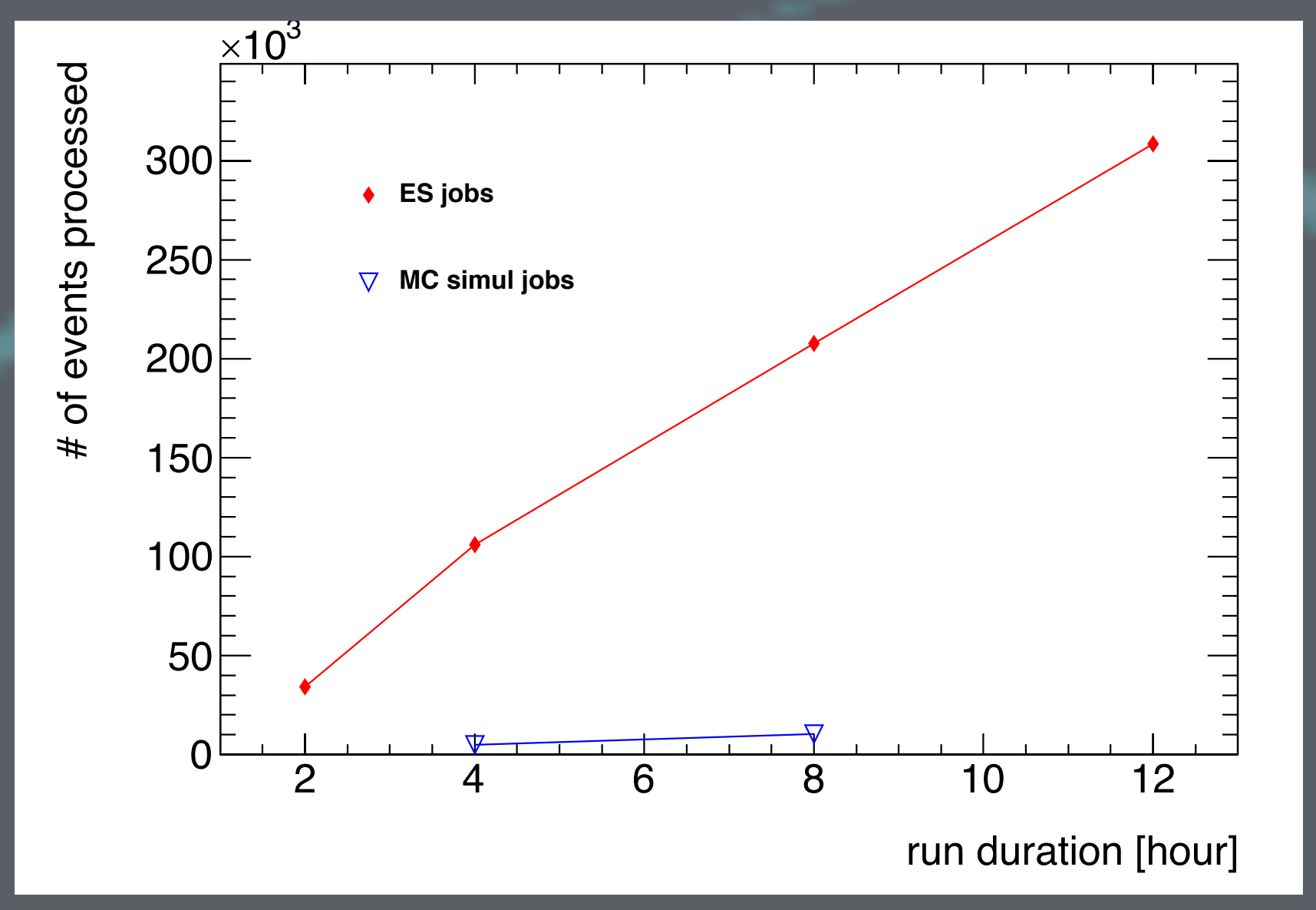
by design Sim@P1 runs only SIM+EVGEN jobs

Event Service pilot test

- The Sim@P1 project has piloted in Event Service (ES) test as the first opportunistic site.
- The ES itself is described on CHEP2016 poster "Production Experience with the ATLAS Event Service"
- The regular simulation jobs will fail and lose all progress when the resources vanish, but the Event Service jobs with the event-by-event processing and staging-out can minimize the loss of the resources changes.
- This is very useful for the Sim@P1 project to achieve successful simulation jobs with short runs (less than 12 hours) during the LHC run intervals.
- Many feedbacks and suggestions have been provided to Event Service team during the test, which helps to improve its performance and fix bugs.



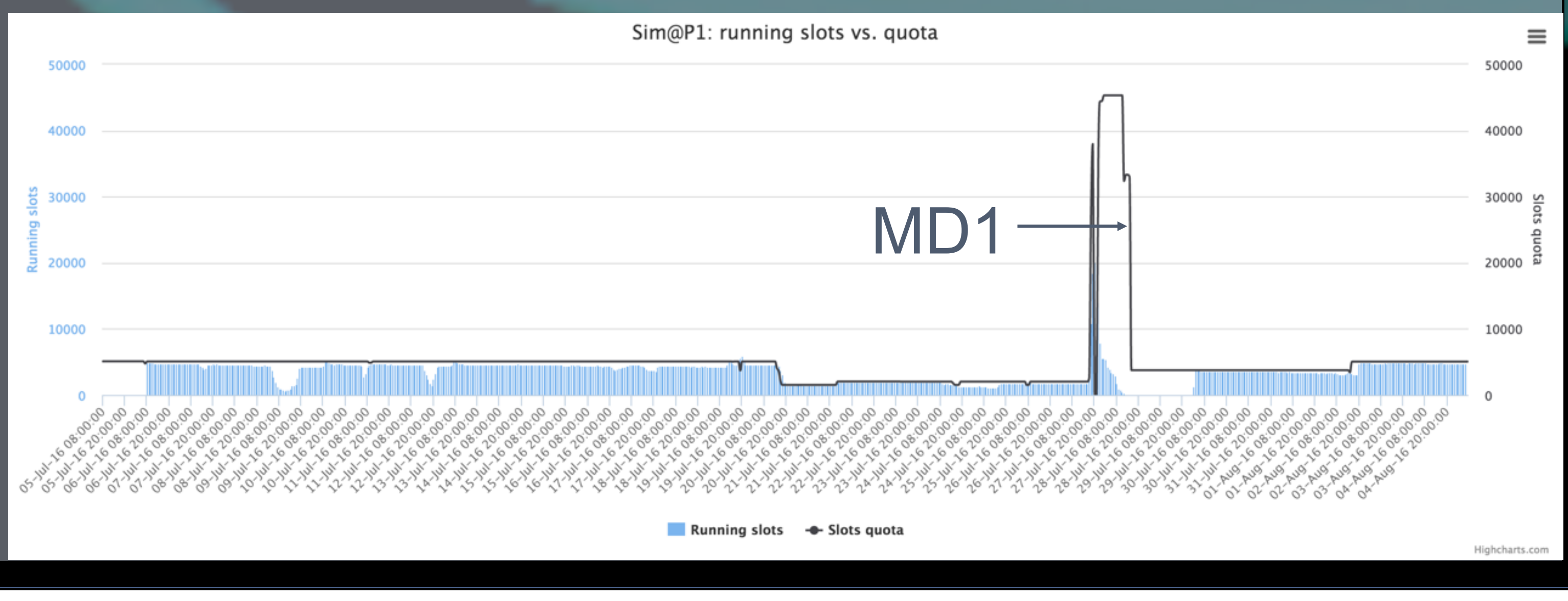
Failed jobs in regular simulation tasks for short periods



Efficiency comparison between regular and ES jobs

Utilization monitoring

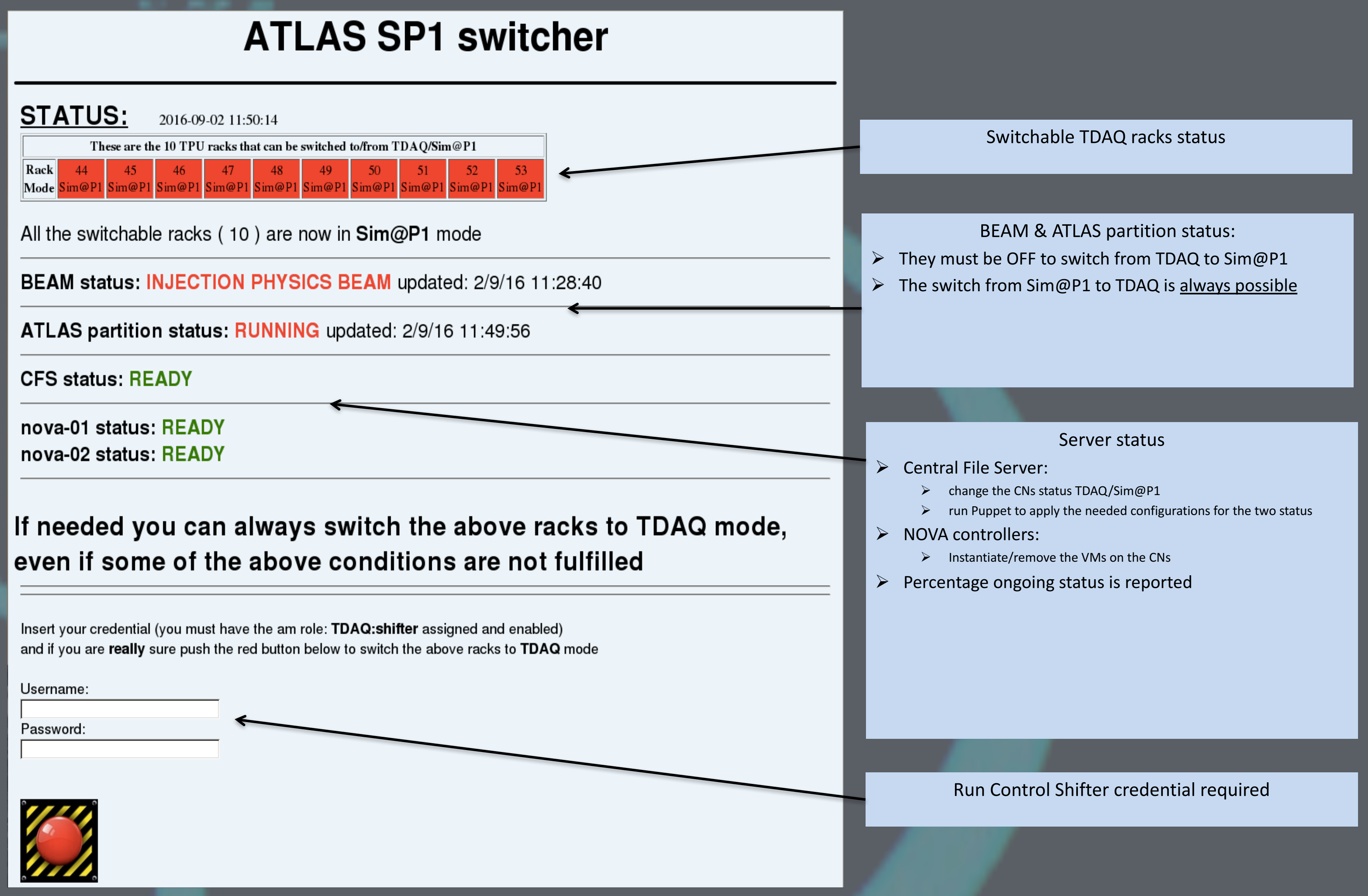
- We improved the monitoring to see utilization of the available resources by running grid jobs.
- If the utilization is below 80% the SP1 responsibles are notified.



Utilization plot of Jul-Aug 2016

Automated switcher: red button

- The red button is a web UI to permit the ATLAS Run Control shifter to automatically re-assign the HLT resource from TDAQ to Sim@P1 and back.
- After checking all the needed requirements, it submits a shell script to change the role of the TDAQ resources and to instantiate/remove the VMs on the CNs.



ATLAS SP1 switcher

STATUS: 2016-09-02 11:50:14

These are the 10 TPU racks that can be switched to/from TDAQ/Sim@P1										
Rack	44	45	46	47	48	49	50	51	52	53
Mode	Sim@P1	Sim@P1	Sim@P1	Sim@P1	Sim@P1	Sim@P1	Sim@P1	Sim@P1	Sim@P1	Sim@P1

All the switchable racks (10) are now in **Sim@P1** mode

BEAM status: INJECTION PHYSICS BEAM updated: 2/9/16 11:28:40

ATLAS partition status: RUNNING updated: 2/9/16 11:49:56

CFS status: READY

nova-01 status: READY
nova-02 status: READY

If needed you can always switch the above racks to TDAQ mode, even if some of the above conditions are not fulfilled

Insert your credential (you must have the am role: **TDAQ:shifter** assigned and enabled)
and if you are **really** sure push the red button below to switch the above racks to **TDAQ** mode

Username:

Password:

Switchable TDAQ racks status

BEAM & ATLAS partition status:
- They must be OFF to switch from TDAQ to Sim@P1
- The switch from Sim@P1 to TDAQ is always possible

Server status
- Central File Server:
  - change the CNs status TDAQ/Sim@P1
  - run Puppet to apply the needed configurations for the two status
- NOVA controllers:
  - Instantiate/remove the VMs on the CNs
- Percentage ongoing status is reported

Run Control Shifter credential required

Dynamic partitioning

- We switched from static HTCondor slots (8 CPU, no RAM limit) to dynamically allocated slots of two types (8 CPU, 8GB RAM + 8 CPU, 15GB RAM)
- The limits are enforced with Linux Cgroups
- The change helped to test the reconstruction jobs that often caused memory exhaustion
- We still see heavy I/O node from reconstruction jobs. This needs to be fully understood before we allow more reco jobs to be processed at Sim@P1.

Future updates

- CC7 based CNs: initial test in 2017
- Latest OpenStack version (we use last SLC6 supported version: Icehouse)
- Further test other type of jobs (reconstruction, reprocessing)
- Get more experience with very short (order of hours) availability of high number of resources
- Evaluate NUMA options for virtual machines

[1] University of Johannesburg, South Africa [2] CERN, Switzerland [3] Istituto Nazionale di Fisica Nucleare Sezione di Bologna, Italy
[4] University of California, Irvine, USA [5] Institute of Physics, ASCR, Czech Republic [6] University of Washington Department of Physics, USA
[7] University of Wisconsin-Madison, USA [8] Brookhaven National Laboratory (BNL), USA [9] University of Cape Town, South Africa