ATLAS computing on Swiss Cloud SWITCHengines

S Haug and F G Sciacca on behalf of the ATLAS Collaboration

Albert Einstein Center for fundamental Physics, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

E-mail: `sigve.haug@lhep.unibe.ch, gianfranco.sciacca@lhep.unibe.ch`

Abstract. Consolidation towards more computing at flat budgets beyond what pure chip technology can offer, is a requirement for the full scientific exploitation of the future data from the Large Hadron Collider at CERN in Geneva. One consolidation measure is to exploit cloud infrastructures whenever they are financially competitive. We report on the technical solutions and the performances used and achieved running simulation tasks for the ATLAS experiment on SWITCHengines. SWITCHengines is a new infrastructure as a service offered to Swiss academia by the National Research and Education Network SWITCH. While solutions and performances are general, financial considerations and policies, on which we also report, are country specific.

ATL-SOFT-PROC-2017-045
09 February 2017

1. Introduction

The computing capacity increase due to technological evolution is not sufficient to fully exploit the scientific potential in the future data from the Large Hadron Collider (LHC) at CERN in Geneva [1]. As budgets for LHC computing are assumed to stay approximately constant, consolidation and innovation at all levels of the computing infrastructure and operation is an obligation. One measure is to exploit computing infrastructures as a service, so-called cloud solutions, whenever they are cost efficient.

In 2016 Switzerland outside CERN contributes with six production sites to the World wide LHC Computing Grid (WLCG). The computing resources are standard linux clusters at the size of 1000 cores. Depending on the supported experiments (ATLAS, CMS and LHCb) [2][3][4], the national contributions amount from a few to ten percent of the so-called tier two computing capacity of the experiments.

In addition to an innovation project investigating the running of LHC computing on the CRAY super computers at the Swiss National Supercomputing Center (CSCS), see *ATLAS and LHC on computing on CRAY* in these proceedings, we have enabled and studied simulation tasks for ATLAS on SWITCHengines [5]. SWITCHengines is an OpenStack infrastructure for academic usage on a pay per usage basis. It is offered by the Swiss National Research and Education Network (NREN) SWITCH [6]. Since 2015 this cloud infrastructure is in production and since April 2016 ATLAS runs LHC proton-proton collision and detector simulation calculations on the infrastructure in a 24/7 mode.

We report on SWITCHengines as an Infrastructure as a Service (IaaS), the computing cluster setup, the integration of the IaaS in WLCG, and some performance and cost considerations.

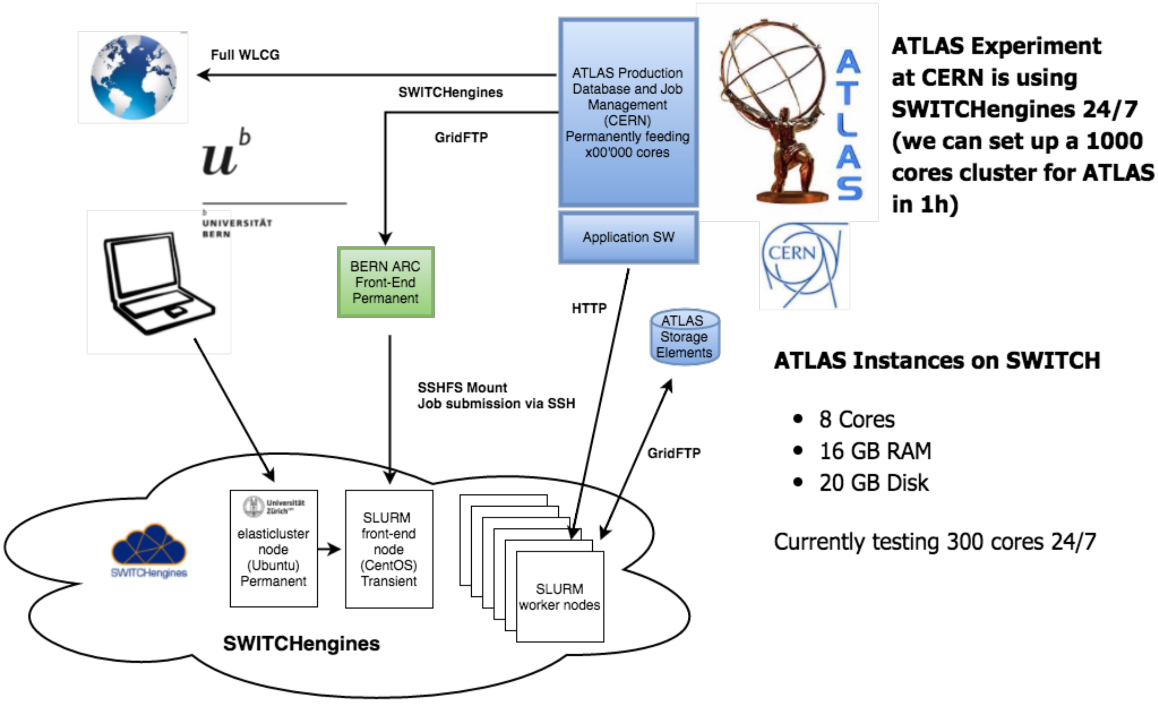

Figure 1. Integration of a Slurm cluster on IaaS SWITCHengines within the ATLAS production system. With the ElastiCluster software and the ARC computing element an elastic compute resource can be initiated within an hour.

2. The Infrastructure as a Service

SWITCHengines is an IaaS based on OpenStack [7]. The physical machines are located in two sites, one in Zurich and one in Lausanne. The IaaS provides compute and storage services to Swiss academic institutions. The users of these institutions have on-demand access to create and manage virtual machines and storage. Prices are available on the website. Commercial customers can get pricing details on request. The IaaS is in operation since 2015. As of 2016 SWITCHengines does not offer GPU machines and uses standard ethernet interconnect.

In 2016 we were given access to a testbed quota with 400 virtual CPU cores with 2 GB RAM and 2.5 GB disk storage per core. The instances used have 8 virtual CPU cores 16 GB RAM and 20 GB disk storage. One instance with a basic CentOS 6 image provided by SWITCHengines was launched and access enabled via an *ssh-key* pair. Some post-installation steps were performed, basically installing and mounting *cvmfs* in order to create an ATLAS worker node image. With the finished worker node a snapshot was created to provide the cluster image [8].

3. The Cluster

The ElastiCluster software developped and maintained by the University of Zurich is used to configure, setup, start and resize a private Slurm batch cluster, i.e. a cluster with only one user and one application [9]. For the ElastiCluster front-end an Ubuntu instance with the same capacities as the worker nodes was launched. ElastiCluster was installed from *github*. The configuration file contains a few values like batch system type, image identifier etc. A 304 virtual CPU core Slurm cluster was then started with one command on the command line. This process took about one hour. A few post-launch steps were needed before the cluster was production ready. However, a skilled system administrator can setup a 1000 core elastic Slurm cluster on the SWITCHengines within half a day. As a result the cluster becomes a transient or non-critical

2016-08-30 CEST 13:27:44

Processes: Grid Local

Country	Site	CPUs	Load (processes: Grid+local)	Queueing
Switzerland	ATLAS BOINC	98139	7894+6083	1571+4063
	ATLAS BOINC 3	98139	5815+8163	1253+4371
	ATLAS BOINC TEST	644	0+0	0+0
	Bern ce01 (UNIBE-LHEP)	1513	1048+0	156+0
	Bern ce02 (UNIBE-LHEP)	770	624+0	159+0
	Bern ce04 (UNIBE-LHEP>	304	304+0	192+0
	Bern UBELIX T3	4472	385+2822	208+2450
	CSCS BRISI Cray XC40	1500	576+0	154+0
	Geneva (UNIGE-DPNC)	720	168+349	169+0
	Lugano PHOENIX T2 arc>	1920	1526+4040	411+14
	Lugano PHOENIX T2 arc>	2240	2065+3504	391+4
	Lugano PHOENIX T2 arc>	2048	1864+3704	407+1
TOTAL	*12 sites*	*212409*	*22269 + 28665*	*5071 + 10903*

Figure 2. Swiss compute resources serving ATLAS. In addition to standard linux clusters there are volunteer based resources (BOINC), high-end HPC (Cray) and recently IaaS (ce04) represented.

component. In case of failure one can just start a new one, within the time it would take to get a hard disk exchanged.

4. The Grid Integration

The Slurm cluster was integrated into the ATLAS distributed computing system with a remote Advanced Resource Connector (ARC) Computing Element (CE) [10]. It is hosted on a virtual machine and receives the computing task instructions via *gridftp*. It generates the Slurm submission scripts which are submitted via an *ssh* connection and mount of the home directory on the Slurm front-end in SWITCHengines. The needed wrapper scripts for the ARC CE back-end were formerly written and used by the University of Bern to integrate Cray machines at CSCS into WLCG [11][12]. Figure 1 provides a sketch of the solution. Figure 2 shows how the IaaS shows up in the ATLAS Nordugrid monitor as of end August 2016.

5. The Performance

Within the ATLAS computing workflow simulations of the LHC collisions and the energy deposit in the dector from the particles created in the collisions are run on SWITCHengines. These simulations require almost no input data and produce only some MB of output per hour and core. Thus, no special bandwith considerations are needed for clusters up to the size of many thousand cores. Noting that simulations make up about 50% of the needed computing time and that the ATLAS distributed computing production system can direct spesific parts of the workflow to specific sites, having sites running only simulations is very reasonable .

In Figure 3 the delivered CPU time is shown from April to August relative to cost. In the end this is the interesting measure for the research. How fast the CPU time is delivered is often of less importance, as the production of the scientific results are not that time critical. The figure shows that the experiment did not manage to use the IaaS very efficiently in the beginning, i.e.

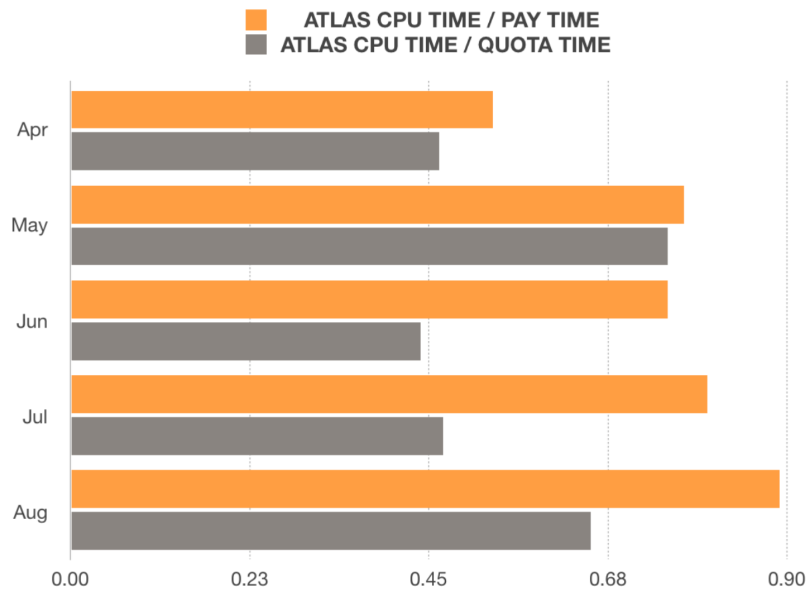

Figure 3. CPU time seen by ATLAS relative to PAY time and QUOTA time seen by the IaaS provider. The PAY time is the up-time of the instances in an elastic cluster. The QUOTA time is the time of all available instances running all the time.

in April. In August ATLAS could use almost 90% of the CPU time it would have had to pay for. This is a good fraction and a higher efficiency is hard to achieve due to the complexity of the full production system. Interesting is the fact that ATLAS in June, July and August did not manage to fill the quota. It is not completely clear if this was due to lack of tasks, instable heart beats from the cluster, e.g. poor network connections letting the site appearing as down, or a combination or both. However, the conclusion is then that costs can be significantly reduced with an elastic cluster, compared to a fixed quota model.

Additional and continuous performance tests, optimally together with every computing task, or job, should be performed to ensure that oversubscription or very slow hardware on the IaaS does not result in significant CPU time dilatation. Such tests are available for the ATLAS production system and will be implemented.

The stability of the IaaS, as observed from April to November, is remarkable. There has not been a single downtime, only a couple of network outages not longer than half an hour and with close to no impact on operation. OpenStack updates are transparent to the customer. Compared to traditional linux clusters with typically monthly downtimes, this is a real gain which may compensate inefficiencies at the percentage level.

6. Prices

SWITCHengines has been in 24/7 production for ATLAS simulations over several months. Maintainence effort has been close to zero, not comparable at all to the operation of own hardware. Performance indications are promising and comparable to what can be achieved on traditional linux clusters. The price is the remaining question.

The 2016 pricing for SWITCHengines are available on their webpage. A 1000 virtual CPU core cluster with 2 GB RAM per core would cost about 70 kCHF per year. For Swiss academic institutions there is no charge for network usage. Comparing with commercial cloud providers or other academic outsourcing options in Switzerland, this price is competitive.

7. Conclusions

We have addressed the question, if an IaaS based on OpenStack is a suitable and cost efficient computing resource for certain LHC computing tasks. SWITCHengines has been used for production simulations for the ATLAS experiment over several months. Such simulations amount for about 50% of the computing needs.

First conclusion is that medium sized clusters for such LHC computing, i.e. at the order of 1000 virtual CPU cores, can be setup and brought into production within a day. The cluster computing resource has become a transient part of the infrastructure. The second conclusion is that an IaaS as SWITCHengines has a close to 100% uptime and maintenance on the user side is close to zero. Third conclusion is that the price is competitive.

Longer term behaviour of the elasticity, i.e. the automatic resizing of the clusters according to demand, and the possible virtual CPU time dilation will be studied.

Acknowledgments

We acknowledge the support of *swissuniversities* and thank SWITCH for the provision of SWITCHengines.

References

[1] Evans L Bryant P (editors) 2008 *JINST* **3** S08001
[2] ATLAS Collaboration 2008 *JINST* **3** S08003
[3] CMS Collaboration 2008 *JINST* **3** S08004
[4] LHCb Collaboration 2008 *JINST* **3** S08005
[5] www.switch.ch/de/engines
[6] www.switch.ch
[7] www.openstack.org
[8] Buncic P et al. 2010 *J. Phys.: Conf. Ser.* **219** 042003
[9] elasticluster.readthedocs.io
[10] Ellert M et al. 2007 *Future Generation Computer Systems* **23** 219-240
[11] Hostettler M 2015 Enabling the ATLAS Experiment at the LHC for High Performance Computing, *Master Thesis from the Science Faculty, University of Bern, Switzerland*
[12] Sciacca F G et al. 2015 *J. Phys.: Conf. Ser.* **664** 062057