Journal of Physics: Conference Series

PAPER · **OPEN ACCESS**

Elastic extension of a local analysis facility on external clouds for the LHC experiments

To cite this article: V Ciaschini *et al* 2017 *J. Phys.: Conf. Ser.* **898** 052024

View the article online for updates and enhancements.

Related content

- Readiness of the ATLAS Spanish Federated Tier-2 for the Physics Analysis of the early collision events at the LHC
  E Oliver, J Nadal, J Pardo et al.
- High performance data transfer and monitoring for RHIC and USATLAS
  J Packard, D Katramatos, J Lauret et al.
- Commissioning of a CERN Production and Analysis Facility Based on xrootd
  Simone Campana, Daniel C van der Ster, Alessandro Di Girolamo et al.

This content was downloaded from IP address 128.141.59.222 on 23/02/2018 at 11:16

IOP Conf. Series: Journal of Physics: Conf. Series **898** (2017) 052024 doi:10.1088/1742-6596/898/5/052024

Elastic extension of a local analysis facility on external clouds for the LHC experiments

V Ciaschini[1], G Codispoti[2], L Rinaldi[3], D C Aiftimiei[1], D Bonacorsi[3], P Calligola[4], S Dal Pra[1], D De Girolamo[1], R Di Maria[5], C Grandi[4], D Michelotto[1], M Panella[1], S Taneja[1], F Semeria[4]

[1] INFN-CNAF, Bologna, IT
[2] CERN, Geneva, CH
[3] Bologna University and INFN, Bologna, IT
[4] INFN-BOLOGNA, Bologna, IT
[5] Imperial College, London, UK

E-mail: lorenzo.rinaldi@bo.infn.it

Abstract.
The computing infrastructures serving the LHC experiments have been designed to cope at most with the average amount of data recorded. The usage peaks, as already observed in Run-I, may however originate large backlogs, thus delaying the completion of the data reconstruction and ultimately the data availability for physics analysis. In order to cope with the production peaks, the LHC experiments are exploring the opportunity to access Cloud resources provided by external partners or commercial providers. In this work we present the proof of concept of the elastic extension of a local analysis facility, specifically the Bologna Tier-3 Grid site, for the LHC experiments hosted at the site, on an external OpenStack infrastructure. We focus on the Cloud Bursting of the Grid site using DynFarm, a newly designed tool that allows the dynamic registration of new worker nodes to LSF. In this approach, the dynamically added worker nodes instantiated on an OpenStack infrastructure are transparently accessed by the LHC Grid tools and at the same time they serve as an extension of the farm for the local usage.

1. Introduction

The physics experiments (such ATLAS [1] and CMS [2]) operating at the Large Hadron Collider at CERN are collecting large amount of collision data and in the future the overall data volume is foreseen to steadily increase. Data processing requires large computing resources, nevertheless the data processing campaigns are often not constant over the time, thus having periods of peak usage when resources usage greatly increases with respect to periods of standard usage. During those peak periods, the world-wide distributed WLCG [3] computing centers may run over the pledged resources and generate long backlog queues to completely absorb the computing workload.

Instead of buying new resources to be fully used only for short time periods, a possible solution could be to access external Cloud resources provided both by external partners and commercial providers, operating in the so called Cloud Bursting mode.

In this contribution we describe a mechanism based on the DynFarm [4] software to dynamically extend the Bologna local farm of the INFN-BOLOGNA-T3 data center to the external cloud facility of the INFN-CNAF both for the ATLAS and CMS experiments.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

2. Dynamic extension of the Bologna local farm

The INFN-BOLOGNA-T3 site is hosted in the INFN-CNAF Tier-1 data center, although the two sites are topologically independent. In order to extend the Bologna farm on an external cloud, initially we created light-weight images of the User Interfaces (UI) and of the Worker Nodes (WN) based on the existing configuration of the Bologna Tier-3 farm. The images are based on CentOS6 and include the software components used by the experiments. The experiments software is accessed through CVMFS [5] and the users mapping is managed by LDAP [6] and Grid Pool Account services. The EMI [7] ARGUS authentication service and the GLEXEC [8] package were installed for the grid site configuration. The access to the distributed experimental data has been set up with the SRM and XrootD [9] protocols, and the local GPFS [10] file system has been exported via NFS [11]. The LSF [12] batch system has been configured with a dedicated setup for remote access (see Section 3 for more details). A schematic view of the farm configuration is reported in Fig. 1.

After that, we loaded the WNs images into the CNAF Cloud infrastructure, an Infrastructure as a Service (IaaS) based on OpenStack [13] Juno. We configured the virtual farm instantiated in OpenStack in order to have the virtual WNs receiving jobs directly from the grid interfaces of the experiments, defining specific experiment testing queues both in the LSF instance and in the experiment Grid information systems.

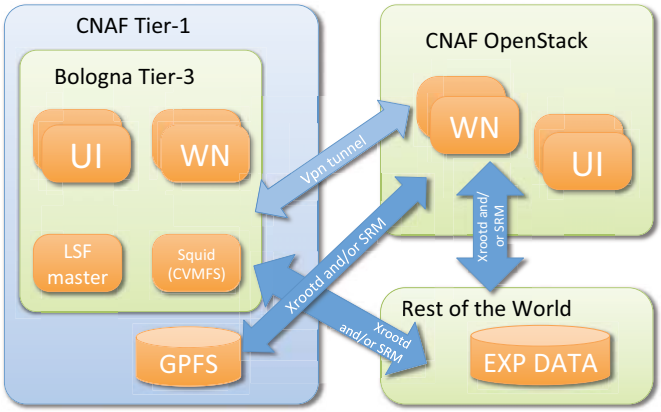

Figure 1. Schematic view of the Bologna Tier-3 farm extension.

3. The DynFarm framework architecture

The main piece of the dynamic extension mechanism is the DynFarm framework, that is based on a custom LSF configuration which allows to extend a standard site on external resources, as shown in Figure 2. This is achieved by instantiating virtual machines (VM) on a remote site and making them part of the local farm. Each VM at startup contacts a Configuration Service in the bursting site: after the authorization phase, it receives a set of configuration files and commands allowing the VM to establish a connection with a VPN server inside the site. The VPN server establishes a GRE (generic routing encapsulation) tunnel with the machines that must be visible to the VM to allow it to work and adjust the proper routes. The VPN connection will make the VM and the OpenVPN server visible to each other, but it will have no further effect on the network connectivity of the VM. This will ensure that all and only the necessary traffic will reach the site, therefore limiting the problems that may be caused by an increased network latency due to the geographical distance from the site. This mechanism

allows a computing center to accept workloads greatly superior to those it has been created to accept. The basic functionalities have been tested by executing the aforementioned analysis workflow using one VM as UI and one as WN, and a dedicated LSF master and queue. The WN was initially instantiated in the local Tier-3 cluster. Then the test was repeated with the VMs instantiated in the OpenStack instance of the CNAF Cloud infrastructure.

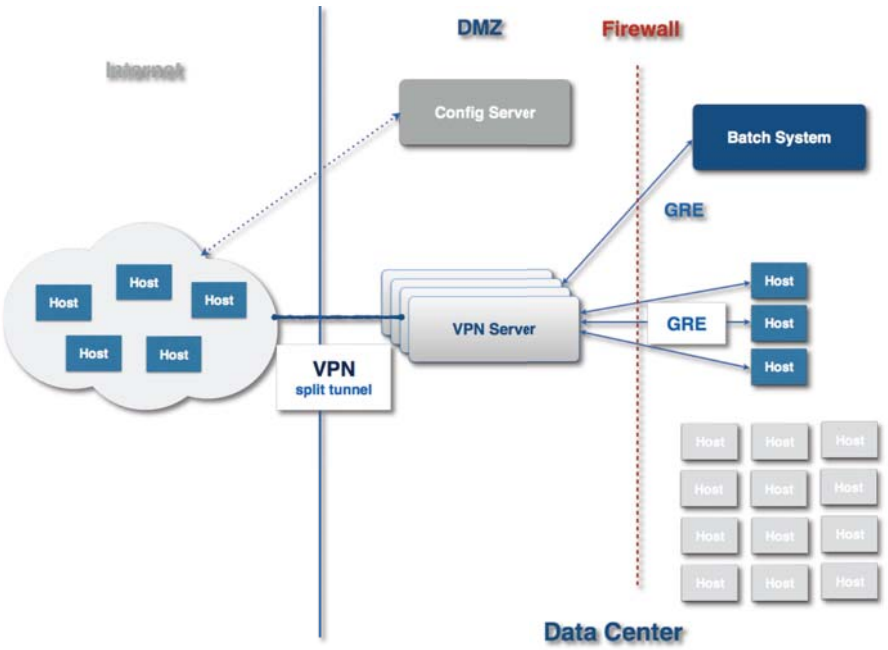

Figure 2. Schematic view of the DynFarm framework

4. Experiments configuration and results

4.1. ATLAS

The ATLAS configuration has been carried out using the ATLAS Grid Information System (AGIS): within the INFN-BOLOGNA-T3 site, we have defined two new PANDA resources [15], one for *Analysis* and one for *Production*, and we tested the queues using the HammerCloud (HC) stressing test system [14]. Bunches of thousands of jobs were sent to the queues and for each test we measured the Wall Clock time, the CPU time and the Event rate. In Figure 3 the CPU-time/Wall-time efficiency distributions are shown for *Analysis* and *Production* jobs. The *Analysis* test showed a 25% average efficiency, while for *production* test we obtained a 80% average efficiency: these value are compatible with typical user analysis and central production jobs. We also obtained similar results when sending the same HC test jobs to the physical nodes of the Tier-3 center.

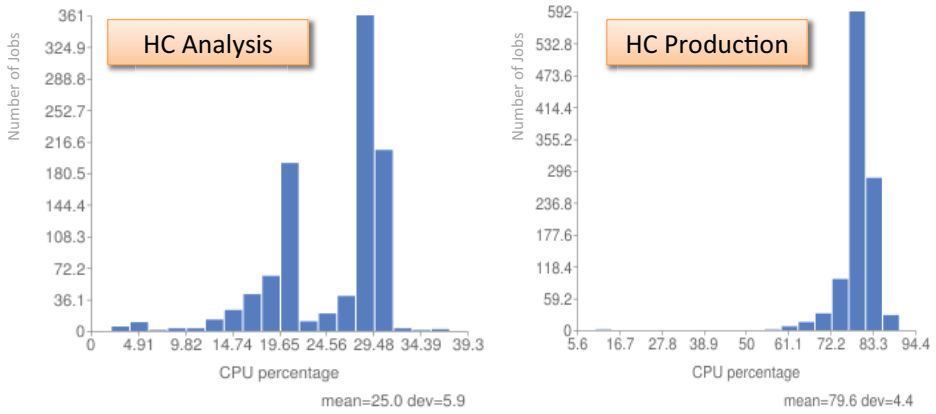

Figure 3. Efficiency of the ATLAS jobs run over the virtual nodes instantiated as extension of the Tier-3 in OpenStack.

4.2. CMS

The CMS setup has already been presented in several occasions [16]. We submitted a Top Quark skimming workflow, accessing remote data through XrootD and copying the results with a standard Grid command to the destination storage. The test setup included five virtual nodes statically allocated with QuadCore CPU with 4 cores and 8 GB RAM. Since the virtual nodes were inserted in a production queue, the jobs were distributed between standard WN and virtual nodes, with only about 5% to the latter. Over more than 3000 jobs submitted, 172 jobs ran in the virtual nodes. Since the jobs were submitted in bunches, they hit all the variability of a production systems, encountering the concurrency of other users jobs, variation of the load. We measured the job efficiency as a ratio between the CPU time over the Wall Clock time for both the jobs running on the OpenStack nodes and the jobs running on the normal WNs. The distribution of the efficiencies does not show a peak, as it can be observed in Figure 4, given all the fluctuation affecting a productions system. Nevertheless, a relevant part of the jobs reached more than 95% efficiency while almost their totality has an efficiency greater than 80%. No job failures on the virtual nodes were observed.

5. Conclusions

The Bologna Tier-3 has been a realistic use case for the CNAF OpenStack infrastructure. The LSF DynFarm extension served as development environment for the tools and setup. We implemented a mechanism to extend the Bologna Tier-3 local farm to an external Cloud, for

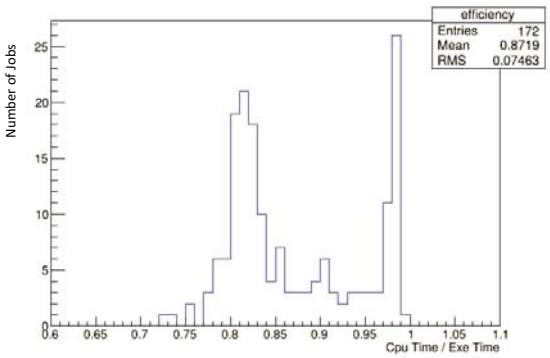

Figure 4. Efficiency of the CMS jobs run over the virtual nodes instantiated as extension of the Tier-3 in OpenStack.

the computing activities of the ATLAS and CMS experiments. Furthermore the mechanism has been used in production to extend the CNAF Tier-1 to external resources, such as the Aruba commercial cloud.

After this successful experience, the Bologna Tier-3 is evaluating to become a pure Cloud site in order to reduce maintenance costs and profit from the CNAF Tier-1 infrastructure.

References

[1] ATLAS Collaboration 2008 *JINST* **3** S08003 (`http://cern.ch/atlas/`)
[2] CMS Collaboration 2008 *JINST* **3** S08004 (`http://cern.ch/cms/`)
[3] Shiers J D 2007 The worldwide LHC computing grid *Comp. Phys. Comm.* **177** 219-223
[4] Ciaschini V et al CHEP2016 proceedings, poster 364
[5] Buncic P et al 2010 CernVM - a virtual software appliance for LHC applications *J. Phys.: Conf. Ser.* **219** 042003
[6] Donley C 2002 LDAP programming, management and Integration *Manning Pubs* ISBN 1-930110-40-5
[7] Aiftimiei C et al 2012 Towards next generations of software for distributed infrastructures: The european middleware initiative *IEEE 8th International Conference* 1-10
[8] Groep D, Koeroo O and Venekamp G 2008 gLExec: gluing grid computing to the unix world *J. Phys.: Conf. Ser.* **119** 062032
[9] Andreeva J et al 2014 Monitoring of large-scale federated data storage: XRootD and beyond *J. Phys.: Conf. Ser.***513** 032004
[10] GPFS http://www-03.ibm.com/systems/software/gpfs/
[11] Sandberg R 1986 The Sun network file system: design, implementation and experience *Proceedings of the Summer 1986 USENIX Technical Conference and Exhibition*
[12] LSF www.ibm.com/spectrum-lsf
[13] OpenStack, http://www.openstack.org/
[14] Van der Ster D Cet al. 2011 *J. Phys.: Conf. Series* **331** 072036
[15] Maeno T et al 2011 *J. Phys.: Conf. Series* **331** 072036
[16] Codispoti G et al 2016 *J. Phys.: Conf. Series* **762** 012013
G. Codispoti et al 2016 *ISGC PoS* 023