Journal of Physics: Conference Series **513** (2014) 032067 doi:10.1088/1742-6596/513/3/032067

Helix Nebula and CERN: A Symbiotic approach to exploiting commercial clouds

Fernando H. Barreiro Megino[α], Robert Jones[α], Katarzyna Kucharczyk[β], Ramón Medrano Llamas[α], Daniel van der Ster[α]

α: CERN, Geneva, Switzerland
β: Warsaw University of Technology, Warsaw, Poland

E-mail: ramon.medrano@cern.ch

Abstract. The recent paradigm shift toward cloud computing in IT, and general interest in "Big Data" in particular, have demonstrated that the computing requirements of HEP are no longer globally unique. Indeed, the CERN IT department and LHC experiments have already made significant R&D investments in delivering and exploiting cloud computing resources. While a number of technical evaluations of interesting commercial offerings from global IT enterprises have been performed by various physics labs, further technical, security, sociological, and legal issues need to be address before their large-scale adoption by the research community can be envisaged.

Helix Nebula - the Science Cloud is an initiative that explores these questions by joining the forces of three European research institutes (CERN, ESA and EMBL) with leading European commercial IT enterprises. The goals of Helix Nebula are to establish a cloud platform federating multiple commercial cloud providers, along with new business models, which can sustain the cloud marketplace for years to come.

This contribution will summarize the participation of CERN in Helix Nebula. We will explain CERN's flagship use-case and the model used to integrate several cloud providers with an LHC experiment's workload management system. During the first proof of concept, this project contributed over 40.000 CPU-days of Monte Carlo production throughput to the ATLAS experiment with marginal manpower required. CERN's experience, together with that of ESA and EMBL, is providing a great insight into the cloud computing industry and highlighted several challenges that are being tackled in order to ease the export of the scientific workloads to the cloud environments.

1. Introduction

Helix Nebula is a consortium formed by public research institutions (CERN, EMBL and ESA) and several commercial providers of cloud services (T-Systems, Atos, CloudSigma, The Server Labs, Interoute, etc.) with the aim to make the private resources accessible by research institutions in Europe in a transparent way. Also, a marketplace is being designed, which should allow better competition across providers.

The role of each research institution is to showcase their flagship use case (e.g. LHC event processing in the case of CERN) and the role of each commercial company is to get prepared for a production environment, learn from the use cases and define a business model for the consortium which is sustainable and guarantees free competition.

2. The CERN Flagship

The CERN use case is based on ATLAS Monte Carlo simulation. In this way, clusters of virtual machines were deployed across providers since 2012 and simulation was sent to them, both on a benchmark fashion [1] and on real Higgs simulation of data.

The clusters were deployed using a plain HTCondor configuration, having the master sitting at CERN, and accessing data throufg CMVFS and EOS.

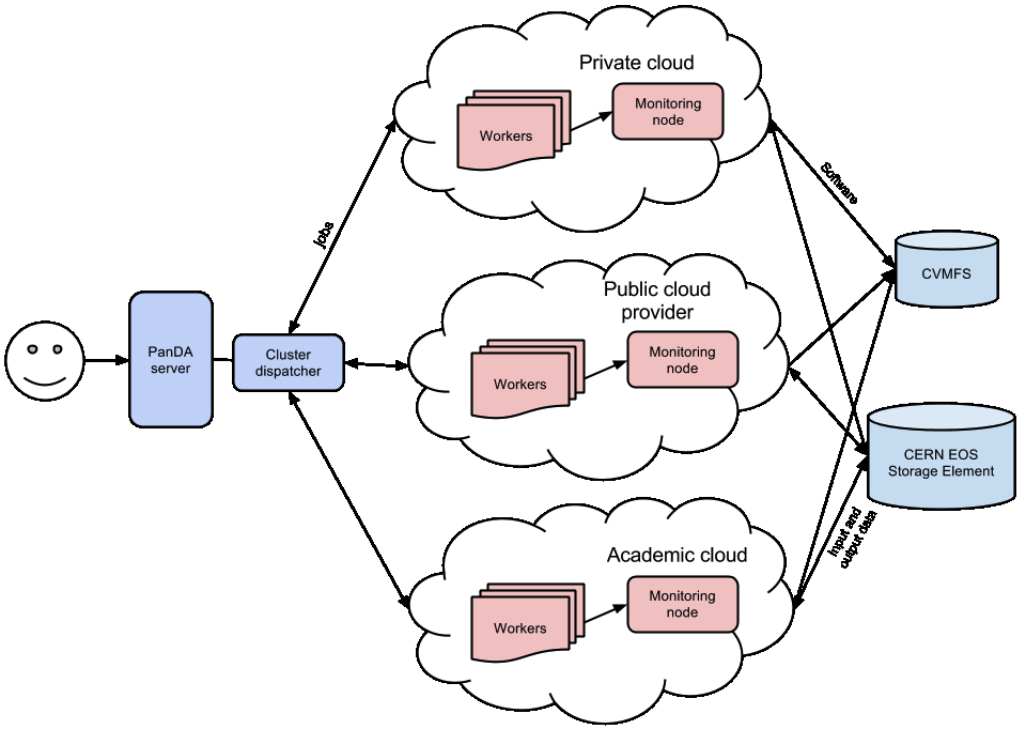

Figure 1: Schematic of the deployment on several clouds

The CERN flagship is one of the most simple use cases for a cloud: all of the worker nodes are independent and they should allow linear scalability in an easy way. On the other hand, deploying more services inside the cloud resources, such as CMVFS proxies or the head node of the HTCondor cluster would reduce the network resources, while introducing possible issues on the communication between different clouds.

3. The Blue Box

In order to achieve some level of federation between clouds and to have transparent, the Technical Architecture group of the consortium, which defines the architecture of the global services to be implemented, has come with the Blue Box proposal.

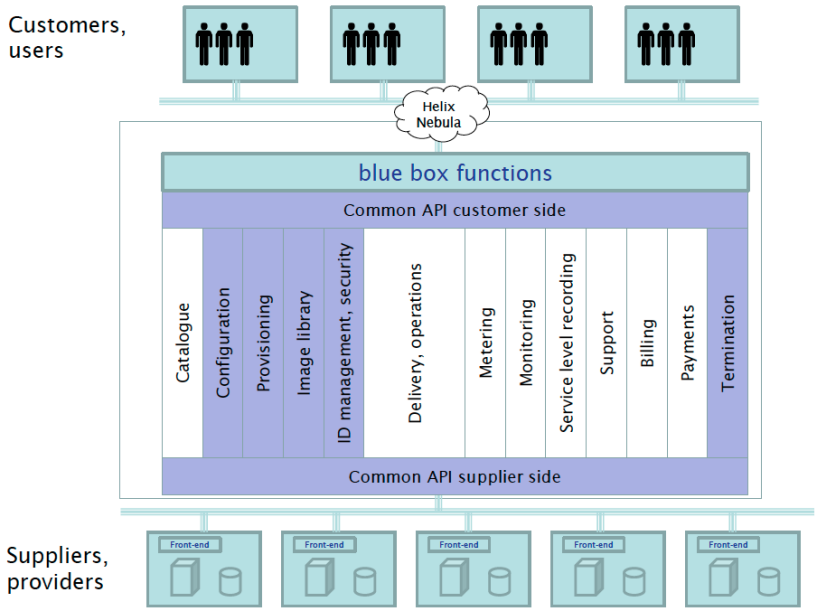

Figure 2: Schematic of the Blue Box proposal

The Blue Box is a service that will allow to transparently accessing the resources of each provider by means of a single interface and endpoint. It offers a northbound API for the clients, e.g. the research institutions and a southbound API where the providers can offer their resources.

Apart from the API translation and unification, there are several services to be offered by the Blue Box, namely an image catalog, billing tools and migration assistants.

The Blue Box was tested in two flavors: enStratius, now part of Dell Inc. and SlipStream from SixSq. A final decision has to be done to select one of them.

4. Testing of the providers

The tests performed came in two steps: the first iteration was centered on the feasibility of using commercial cloud resources for research purposes, which was carried out in early 2012.

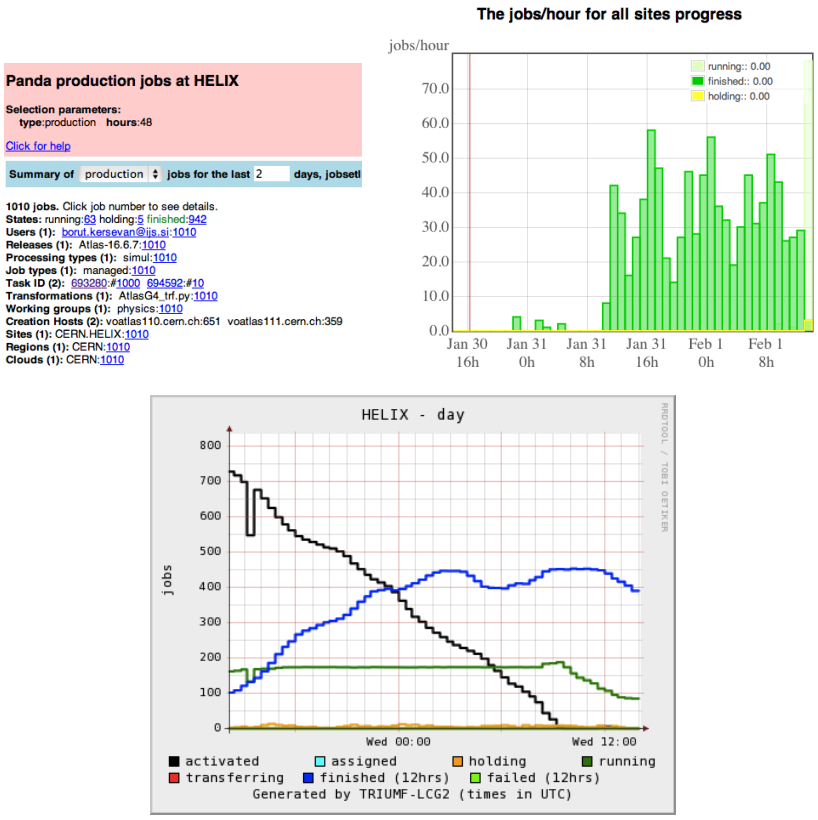

Figure 3: Results of the very first testing on Helix Nebula

This round of tests was not carried out using any Blue Box, but accessing the resources directly by means of the provider's API. Also, there was no multi cloud deployment and each provider was tested on by one. On the other hand, this test paved the way to access cloud resources to physics and the very worker node configuration and schema was then used for the deployment in-house.

The second round of tests was carried out on the summer of 2013 with the very same configuration, but accessing the resources via the Blue Boxes deployed.

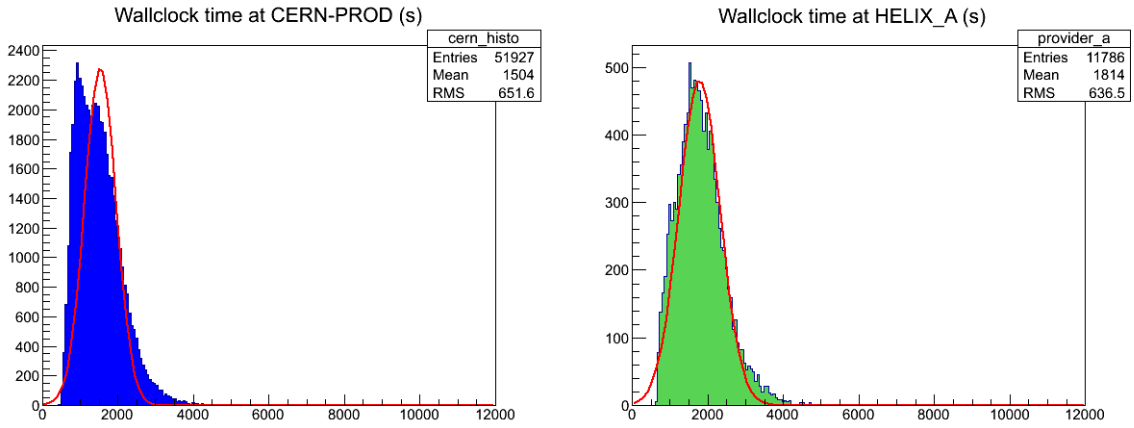

Figure 4: Comparison of performance of CERN's bare metal machines and one Helix Nebula provider

The results of this second round of tests is still not publishable at the moment of writing of this note.

5. Conclusions and future work

CERN's involvement in the Helix Nebula Science cloud has demonstrated that HEP computations can be offloaded to private cloud providers by reusing components of the WLCG in a non-Grid world.

The transparency of the Blue Box abstracts the various provider pecularities so that users may exploit different providers easily. Future features such as image interoperability, metering, and billing will further improve the offering.

Additional improvements to our client-side include better monitoring to enable us to countercheck the blue-box billing; and the architecture related to Condor may also be further simplified to enable scaling into may 10k's of processes.

The long term vision for future work in this area would see the development of technologies which enable further federation of private and public clouds. In the near future, we may see fluid movement of data and processing between the many cloud providers.

References

[1] Testing as a Service with HammerCloud. Medrano Llamas, Ramon and Elmsheuser, Johannes and Legger, Federica and van der Ster, Daniel and Sciacca, Giovanni and Barrand, Quentin. Computing in High Energy and Nuclear Physics (CHEP). 2013.