



Helix Nebula and CERN: A Symbiotic Approach to Exploiting Commercial Clouds

Fernando H. Barreiro Megino, Robert Jones, Katarzyna Kucharczyk
Ramon Medrano Llamas, Daniel van der Ster
CERN, IT Department, CH-1211, Geneva 23, Switzerland



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.

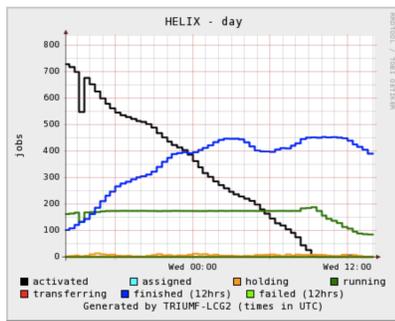
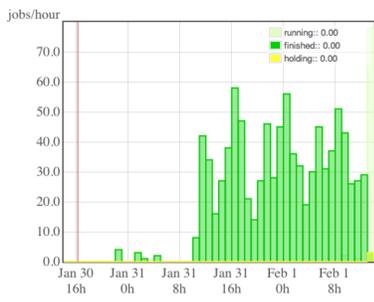
This contribution will summarize the participation of CERN in Helix Nebula. We will explain CERN's flagship use-case and the model used to integrate several cloud providers with an LHC experiment's workload management system. During the first proof of concept, this project contributed over 40.000 CPU-days 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.

Early Success in January 2012

100 nodes, 200 CPUs used at commercial provider for G4 production tasks

Panda production jobs at HELIX
Selection parameters:
type:production hours:48
Click for help
Summary of production jobs for the last 2 days, jobset

The jobs/hour for all sites progress



Infrastructure

We have built a cloud execution framework using Condor to schedule the job slots, CVMFS as the machine images, EOS for data storage, and Ganglia for monitoring:

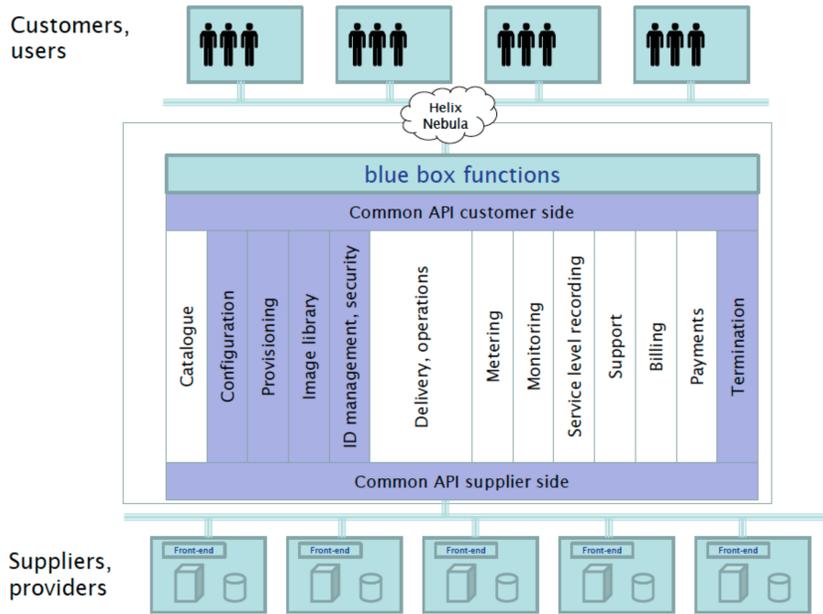
Condor: orchestrates the jobs; a condor_schedd boots with the cloud VM and calls home to a condor_master at CERN to get its workload.

CVMFS: holds all the software and configuration.

EOS: input and output of physics data. Allows remote access from cloud sites across the WAN.

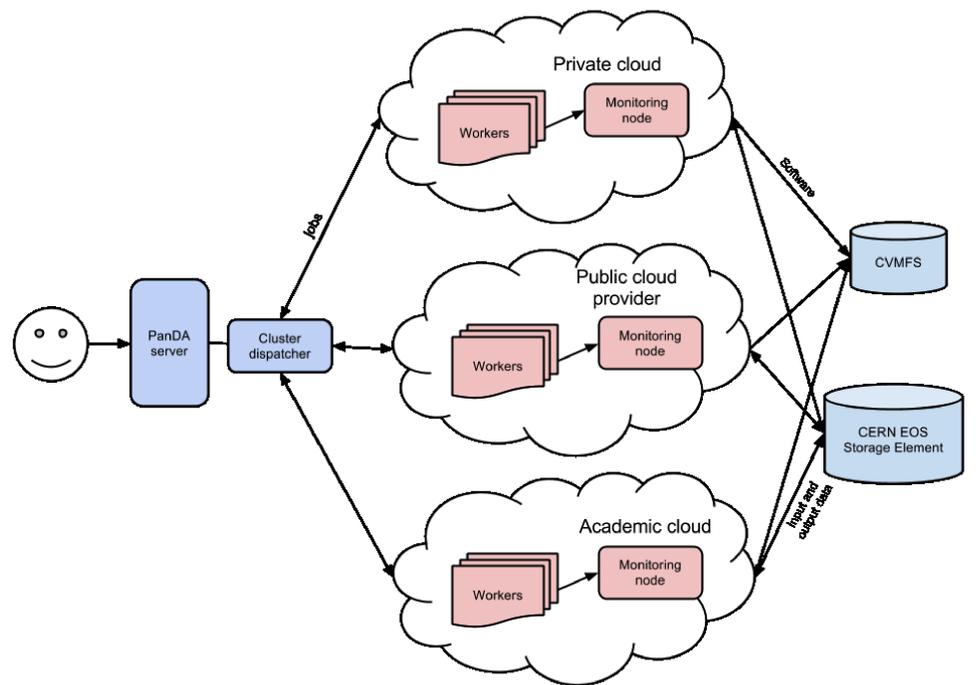
Ganglia: basic monitoring of machines.

The Blue Box



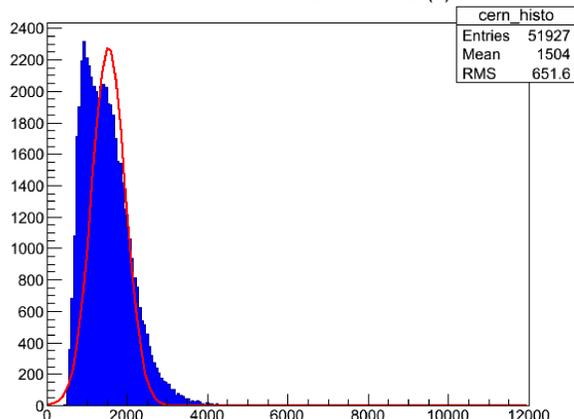
Source: Helix Nebula Deliverable D4.2

Running HEP Jobs on the Cloud with CVMFS and EOS

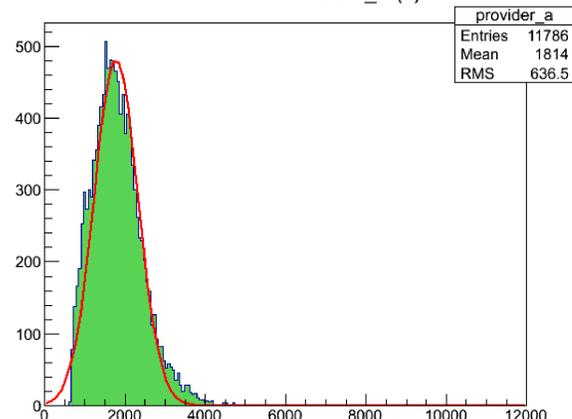


Results

Wallclock time at CERN-PROD (s)



Wallclock time at HELIX_A (s)



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 peculiarities 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.