The DII-HEP OpenStack based CMS Data Analysis for Secure Cloud Resources

L. Osmani[1],S. Toor[2],M. Komu[3],M. Kortelainen[2],T. Lindén[2],J. White[2],R. Khan[4], P. Eerola[2],S. Tarkoma[1]

[1]University of Helsinki, [2]Helsinki Institute of Physics, [3]Ericsson Research, [4]University of Alabama

Goals

Harness Grid and Cloud technologies to ensure a steady and seamless transition towards new ways of operating.

Motivation

- The High Energy Physics community is interested in performing simulations and data analysis on public or private cloud facilities.
- Currently the simulations and analysis are being performed mostly on computing and data Grids.
- The software and experience of operating on a Grid needs to be adapted for running on cloud facilities.

Contribution: Deploy a Cloud-based and Grid-enabled cluster

- We combine the elements of Cloud and Grid software components.
- We manage the VMs dynamically in an elastic fashion.
- We use the EMI authorization service (Argus) and the Execution environment Service (Argus-EES).
- Plugin developed for Argus-EES that can communicate with multiple OpenStack deployments to expand and shrink resources on-demand.
- Leverage HIP protocol for traffic management and security.

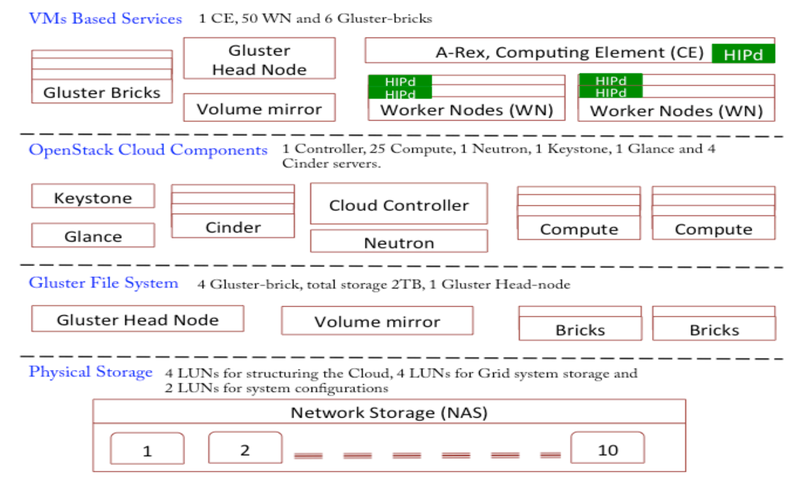


Figure 1. Architecture of DII-HEP Cloud

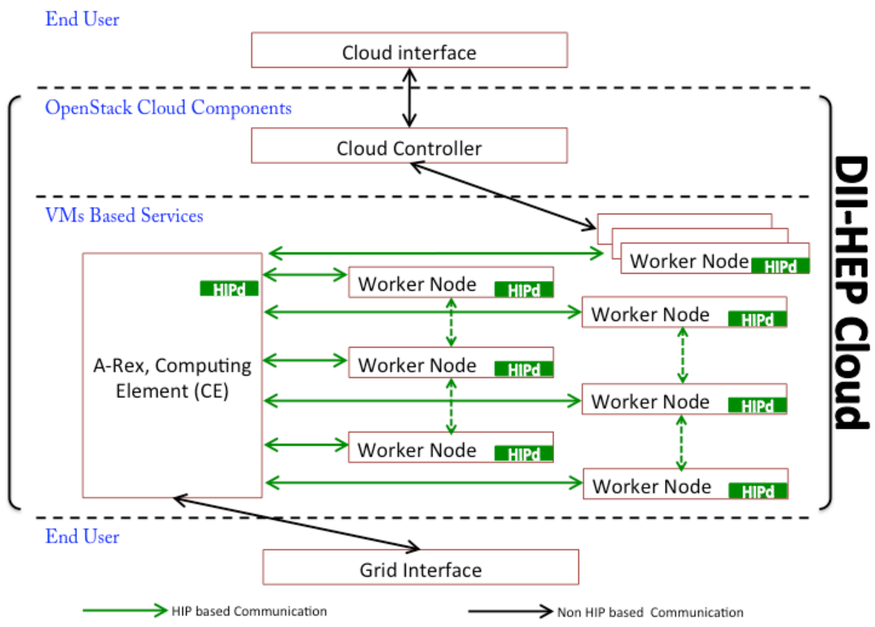


Figure 2. DII-HEP Cloud

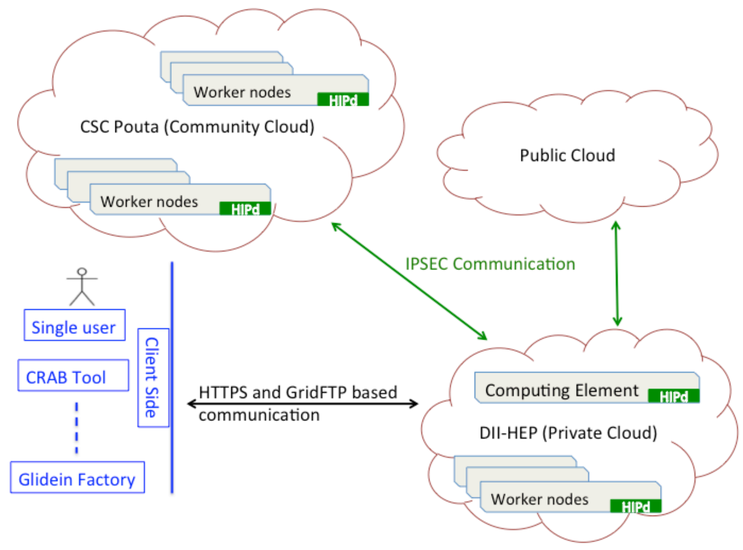
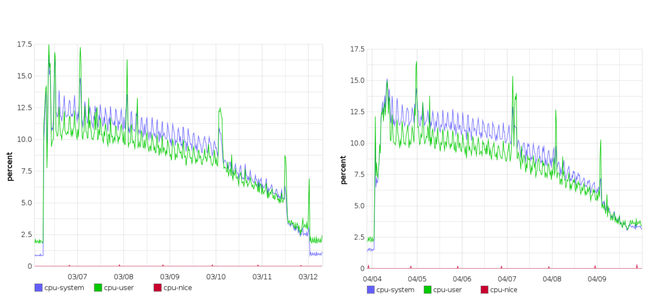


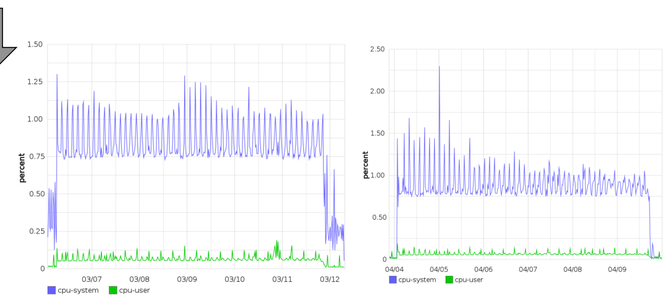
Figure 3. Cloudbursting

Evaluating our implementation

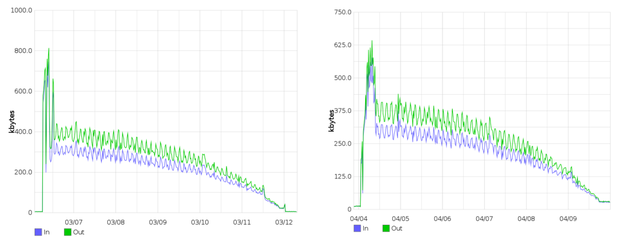
- The constructed virtual cluster is Cloud-based and Grid-enabled.
- OpenStack used for the Cloud and the Advance Resource Connector (ARC) for the Grid.
- Analysis software and libraries provided through CERNVMFS.
- Cloudbursting towards other community Clouds.
- CPU intensive CMS jobs CRAB jobs were run with and without the HIP protocol to study the difference in CE and workernode CPU and network usage. The jobs were run for about 170 min and each test was running for about a week on 200 concurrent jobs.



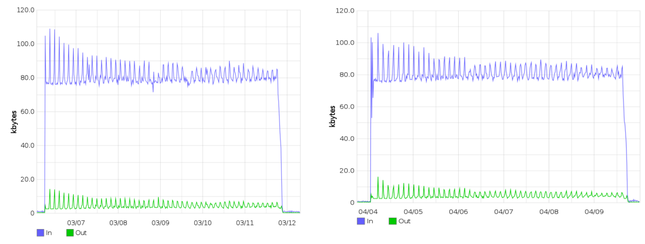
CPU CE vs CPU CE HIP



CPU workernodes vs CPU workernodes HIP



CE network vs CE network HIP



Network workernodes vs network workernodes HIP

The Host Identity Protocol (HIP) has been designed for mobile networks and it provides a secure method for IP mobility and multi-homing. HIP separates the end-point identifier and locator role for IP address which improves network agility of applications and the underlying virtual machines.

Future work

- The virtual cluster has been migrated from the test setup at the University of Helsinki to the cPouta IaaS at the CSC Kajaani Datacenter.
- The new setup is being taken into test usage.
- The largest Finnish CMS/ALICE physical cluster is planned to be replaced by this cloud setup.

REFERENCES:

S.Toor,L.Osmani,P.Eerola,O.Kraemer,T.Lindén,S.Tarkoma,J.White, *A scalable infratructure for CMS analysis based on OpenStack Cloud and Gluster file system*, J. Phys.: Conf . Ser. 513 062047

J.White, S.Toor, P.Eerola,T.Lindén,O.Kraemer,L.Osmani,S.Tarkoma, *Dynamic Provisioning of Resources in a Hybrid Infrastructure*, PoS(ISGC2014)019, International Symposium on Grids and Cloud (ISGC) 2014,Taiwan

L.Osmani,S.Tarkoma,P.Eerola,M.Komu,M.J.Kortelainen,O.Kraemer,T.Lindén, S.Toor,J.White, *An overview of the DII-HEP OpenStack based CMS data analysis*, Submitted to J.Phys.:Conf.Ser.ACAT 2014.