Peak2Cloud: Scientific Computing on the Cloud

JOSEPH ANTHONY C. HERMOCILLA, University of the Philippines Los Banos

Peak2Cloud (P2C) is an Openstack-based private cloud for scientific and high performance computing. First, we present how P2C was configured and tested. Then we describe vcluster, a tool for rapidly deploying message-passing clusters on P2C. Lastly, we analyze some benchmark results on the performance of P2C deployed virtual clusters.

Categories and Subject Descriptors: C.2.4 [**Computer-Communication Networks**]: Distributed Systems

General Terms: Network operating systems

Additional Key Words and Phrases: cloud computing, high-performance computing

ACM Reference Format:
Joseph Anthony C. Hermocilla, 2010.Peak2Cloud: Scientific Computing on the Cloud. *ACM Trans. Comput. Syst.* 2014, 1, Article 2 (June 2014), 5 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Cloud computing has become a buzzword in today's modern computing, though there is no agreed upon meaning of the term. In 2011, [Mell and Grance 2011] of NIST published a definition that is widely quoted and used. The popularity of cloud computing mainly comes from its ability to provision additional resources on demand with minimum intervention from the provider. It leverages advances in virtualization and web services technologies. For example, a website with a sudden increase in workload can start another server machine (possibly virtual) almost instantaneously to accommodate the additional load.

Cloud computing offers service models which include *Software-as-a-Service(SaaS), Platform-as-a-Service(PaaS), and Infrastructure-as-a-Service(IaaS)*. IaaS allows the consumer to provision computing resources(hardware, network, storage) to run arbitrary software including operating systems [Mell and Grance 2011].

A cloud can be deployed in several ways. *Private* clouds are operated for an organization. *Community* clouds are shared by several organizations to support a community with shared concerns. *Public* clouds are available to the public. Lastly, hybrid clouds are composition of two or more clouds [Mell and Grance 2011].

1.1. Openstack

Openstack an open source software framework for deploying clouds. [Sefraoui 2012] It is based on Nebula used by NASA. It provides a public interface that is compatible

---

This work is supported by the Philippine Department of Science and Technology (DOST) Accelerated Science and Technology Human Resource Development Program (ASTHRDP).
Author's address: J. A. C. Hermocilla, Institute of Computer Science, University of the Philippines Los Banos

DOI:http://dx.doi.org/10.1145/0000000.0000000

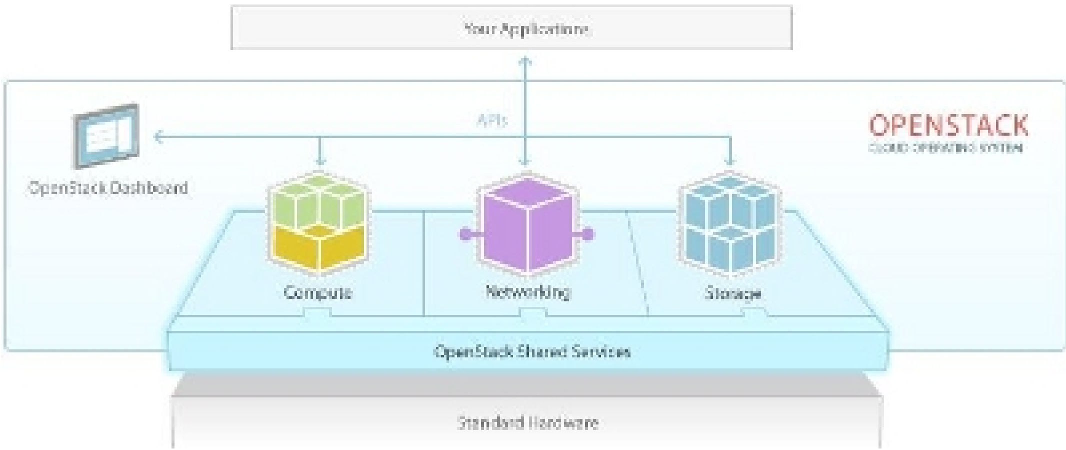

Fig. 1. Openstack at a glance.

with Amazon EC2. Several components of Openstack which are developed separately. The names in parenthesis represent the project name used by the developers.

— *Object Store("Swift")* - provides storage
— *Image("Glance"* - provides a catalog and repository for virtual disk images
— *Compute("Nova")* - provides virtual servers upon demand
— *Dashboard("Horizon")* - provides a modular web-based user interface for all the Openstack services
— *Identity"Keystone"* - provides authentication and authorization for all Openstack services
— *Networking("Quantum")* - provides "network connectivity as a service"
— *Block Storage("Cinder")* - provides persistent block storage for guest VMs

Figure 1 shows the interaction of the major Openstack components.

2. RELATED WORK

Studies have been published to evaluate the applicability of the cloud for scientific computing. The works described below focuses on performance.

[Walker 2008] showed that a performance gap between running HPC applications on a baremetal cluster and on an Amazon's EC2 provisioned cluster. They suggested that in order for cloud computing to be a viable alternative for HPC, providers must improve in the area of network interconnection.

[Evangelinos and Hill 2008] found that Amazon's EC2 may be a credible solution for on-demand and small-sized HPC applications. They supported this conclusion by running a low-order coupled atmosphere-ocean simulation on EC2.

[Ekanayake and Fox 2010] presented performance analysis of HPC applications on virtualized resources. They concluded that cloud techonologies work well for pleasingly- parallel problems. The main limitation of cloud technologies is the high overhead for applications with complex communication patterns, even with large data sets.

[Jackson et al. 2010] compared the performance of conventional HPC platforms to Amazon EC2. Their results showed that EC2 is six times slower than a typical mid-range Linux cluster, and twenty times slower than a modern HPC system. This is mainly because of the communication overhead. They also noted that variability in performance can be significant due to the shared nature of the cloud environment.

Fig. 2. Hardware used in P2C.

[Zhai et al. 2011] conducted a comprehensive comparison of the performance of a baremetal cluster(connected using Infiniband) and a cluster deployed using Amazon's Cluster Compute Instances (CCI). The study also revealed that running MPI applications in the cloud yielded more positive results compared to published results. They also highlight the flexibility and elasticity advantage of using cloud.

[Mauch et al. 2013] presented the High Performance Cloud Computing (HPC2) model. This model enables the provisioning of elastic virtual clusters which avoids the initial cost for physically owned hardware. They also presented a novel architecture for HPC IaaS clouds which support InfiniBand with QoS mechanisms since existing platforms still use Ethernet.

[Expsito et al. 2013] concluded that HPC application scalability depends mainly on the communication performance. Their study involved the use of Amazon's EC2 Cluster Compute Instances (CCI) platform targeted to HPC applications. This platform provides access to a high-speed network(10 Gigabit Ethernet).

[Ludescher et al. 2013] presented a novel code execution framework (CEF) to execute problem solving environment (PSE) source code in parallel on a cloud. The paper emphasized that the use of a public cloud can result to a magnitude of cost savings.

Most of these utilized the public cloud, specifically Amazon EC2 as their testbed.

3. METHODOLOGY

3.1. Hardware

P2C uses commercial-off-the-shelf(COTS) hardware. The cloud controller(1 unit) and compute nodes(2 units) is a four-core Intel(R) Core(TM) i3-2000 3.10GHz CPU with 4GB RAM and 100GB disk. A 1GBps, 16-port Dell PowerConnect 2716 switch connects the controller and the nodes.

Fig. 3. Switch used in P2C.

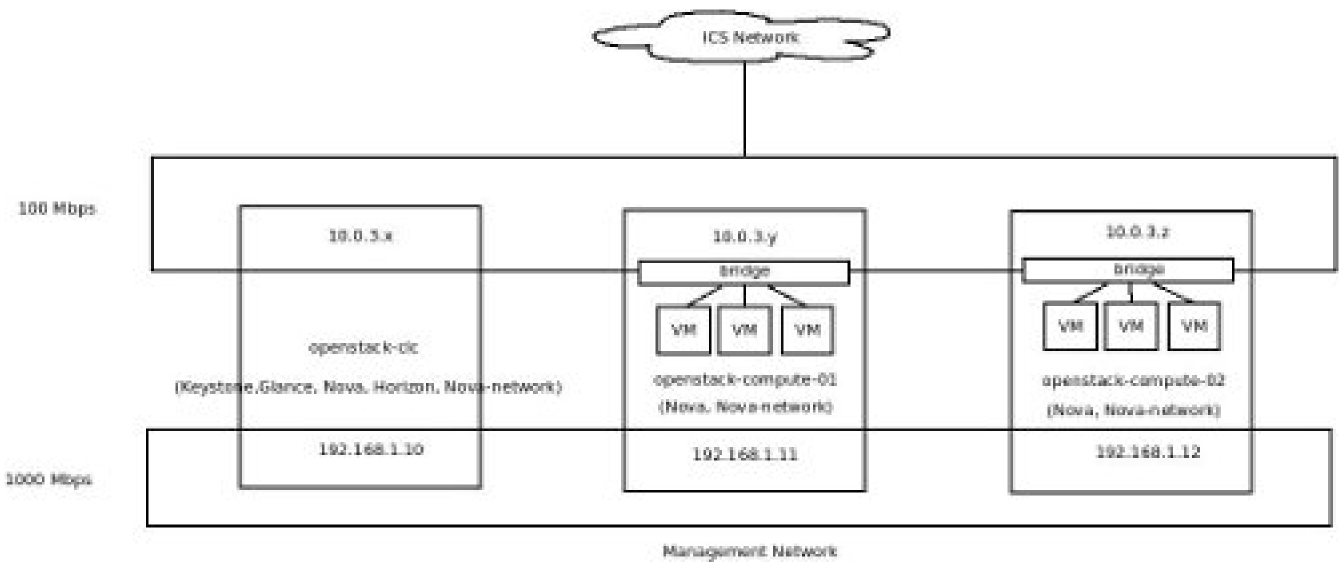

Fig. 4. Network topology

3.2. Network Topology

3.3. vcluster

3.4. Benchmarks

4. RESULTS AND DISCUSSION

5. CONCLUSIONS

ACKNOWLEDGMENTS

The author would like to thank the Lord.

REFERENCES

Jaliya Ekanayake and Geoffrey Fox. 2010. High performance parallel computing with clouds and cloud technologies. In *Cloud Computing*. Springer, 2038. http://link.springer.com/chapter/10.1007/978-3-642-12636-9_2

Constantinos Evangelinos and C. Hill. 2008. Cloud computing for parallel scientific hpc applications: Feasibility of running coupled atmosphere-ocean climate models on amazons ec2. *ratio* 2, 2.40 (2008), 234. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.296.3779&rep=rep1&type=pdf

Roberto R. Expsito, Guillermo L. Taboada, Sabela Ramos, Juan Tourio, and Ramn Doallo. 2013. Performance analysis of HPC applications in the cloud. *Future Generation Computer Systems* 29, 1 (Jan. 2013), 218–229. DOI:http://dx.doi.org/10.1016/j.future.2012.06.009

Keith R. Jackson, Lavanya Ramakrishnan, Krishna Muriki, Shane Canon, Shreyas Cholia, John Shalf, Harvey J. Wasserman, and Nicholas J. Wright. 2010. Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud. IEEE, 159–168. DOI:http://dx.doi.org/10.1109/CloudCom.2010.69

Thomas Ludescher, Thomas Feilhauer, and Peter Brezany. 2013. Cloud-Based Code Execution Framework for scientific problem solving environments. *Journal of Cloud Computing: Advances, Systems and Applications* 2, 1 (2013), 11. http://www.journalofcloudcomputing.com/content/2/1/11

Viktor Mauch, Marcel Kunze, and Marius Hillenbrand. 2013. High performance cloud computing. *Including Special sections: High Performance Computing in the Cloud & Resource Discovery Mechanisms for P2P Systems* 29, 6 (Aug. 2013), 1408–1416. DOI:http://dx.doi.org/10.1016/j.future.2012.03.011

Peter Mell and Timothy Grance. 2011. The NIST definition of cloud computing (draft). *NIST special publication* 800, 145 (2011), 7. http://pre-developer.att.com/home/learn/enablingtechnologies/The_NIST_Definition_of_Cloud_Computing.pdf

Omar Sefraoui. 2012. OpenStack: Toward an Open-source Solution for Cloud Computing. *International Journal of Computer Applications* 55, 3 (Oct. 2012), 38–42.

Edward Walker. 2008. Benchmarking Amazon EC2 for high-performance scientific computing. *;Login* 33, 5 (Oct. 2008). https://www.usenix.org/legacy/publications/login/2008-10/openpdfs/walker.pdf

Yan Zhai, Mingliang Liu, Jidong Zhai, Xiaosong Ma, and Wenguang Chen. 2011. Cloud versus in-house cluster: evaluating Amazon cluster compute instances for running MPI applications. In *State of the Practice Reports*. ACM, 11. http://dl.acm.org/citation.cfm?id=2063363

Received May 2014; revised June 2014; accepted June 2014