Cloud Computing Based HPC: A Framework for Ethiopian Universities

Samuel Fentahuen, Sreenivas Velagapudi

MSc Student, Dept of Computing, Adama Science and Technology University, Adama, Ethiopia.

Assistant Professor, Dept of Computing, Adama Science and Technology University, Adama, Ethiopia.

ABSTRACT: Cloud computing is becoming an adoptable technology for many of the organizations with its dynamic scalability and usage of virtualized resources as a service through the Internet. It is growing rapidly with applications in almost any area, including academia. [1]Cloud computing is model for enabling a convenient, on- demand and network access to a shared pool of configurable computing resources, (networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimum management effort or service provider interaction. High Performance Computing (HPC) enables scientists and researchers to solve complex problems that require many computing capabilities. Cluster computing has become popular in academia and industry. Clusters of servers are used for a variety of distributed applications like simulation, data analysis, web services, and so on.

Hence, one of the areas that need to be researched more is handling of large data sets in the cloud. La rge data set is one of important characteristics of HPC applications. In this paper, we will discuss private cloud infrastructure implementation for each university using OpenStack by considering HPC. Then, create a collaboration mechanism for each University private clouds throughout the country using a federated cloud architecture.

KEYWORDS: Cloud Computing, HPC, Education Cloud Federated Cloud Computing

I. INTRODUCTION

With the rapid development of processing and storage technologies and the success of the Internet, computing resources have become cheaper, more powerful and more ubiquitously available than ever before. This technological trend has enabled the realization of a new computing model called cloud computing, in which resources (e.g., CPU and storage) are provided as general utilities that can be leased and released by users through the Internet in an on-demand fashion. Now a days governments, Academic institutions, research centers, different governmental and non-governmental institution are now adopting cloud computing as a solution for ever increasing IT related problems and needs, for example plenty of academic institutions are tend to use Google's email app as a solution for their enterprise email system, another example now a day's every individual is storing files on a cloud storage like Google Drive, Dropbox, SurDoc etc. So, in one ways or another we are now using cloud offerings. At present, cloud is giving services beyond common cloud service like SaaS, PaaS, and IaaS it is used for High Performance Computing Infrastructure. Cloud computing presents a unique opportunity for batch processing and analytics jobs that analyze terabytes of data and can take hours to finish.

Cloud technologies such as Google MapReduce, Google File System (GFS), Hadoop and Hadoop Distributed File System (HDFS), Microsoft Dryad, and CGL-MapReduce adopt a more data-centered approach to parallel runtimes [2] [3]. In these frameworks, the data is staged in data/compute nodes of clusters or large-scale data centers. The main goal of this paper is to create a framework for Cloud base HPC infrastructure for Ethiopian Universities. The motivation of this paper work laid on EthERNet [4] project, the project is aimed to build and deliver highly interconnected and high performance networks for Universities and other Educational and Research Institutions in Ethiopia. More specifically, EthERNet was aimed to build and deliver high performance networking that connected these institutions with each other and similar institutions in the world, and by doing this to enable them to share educational resources and collaborate both within Ethiopia and globally

The paper is organized as follows, the first section will discuss on cloud computing, its service and deployment types, second section will discuss on HPC and related technologies, the third section will discuss on Cloud federation, the fourth section presents related works and finally we will present our proposed cloud enabled HPC infrastructure framework for Ethiopian Universities and finally conclusion and future work.

II.   RELATED WORK

The Computational Intelligence Research Group (CIRG) at University of Pretoria, South Africa [5], they were doing research on CI algorithms, but students face challenge that the problems they are trying to solve are not trivial. This means that the search space for CI algorithms can become extremely big, resulting in very computationally expensive workloads. To achieve statistical significance results, each students workload needs to include in order of thousands experiment using different parameters, inputs, problem types, etc. each student's work load could potentially take day or weeks and for some extra cases it takes even months to compute on a single workstation running 24 hours 7 days. Students from CIRG attempt to solve their challenge by running their experiments on more than one workstation simultaneously. This provide some improvement on throughput, scalability and failover but it has many problems on scheduling and management.

At the end they come up to automate their problems with cloud computing, cloud computing describe both a platform and a type of application. A cloud computing platform dynamically provisions, configures, reconfi gures and de-provisions servers as needed. Servers in the cloud can be physical machines or virtual machines. A cloud is a pool of virtualizes computer resources that can:

- Host a variety of different workloads, including batch-style- back-end jobs and interactive, user-facing applications
- Allow workloads to be deployed and scaled-out quickly through the rapid provisioned of virtual or physical machines
- Support redundant, self-recovering, highly scalable programming models that allow workloads to recover from many unavoidable hardware/software failures
- Monitor resource use in real time to enable rebalancing of allocations when needed

So for these researches cloud computing simplifies management, scheduling and booking of computing resources easily. Another technology they use to automate their research is create Grid application with Apache Hadoop [6], it is an open-source framework for running parallel computing applications on large clusters of commodity hardware. Apache Hadoop is based on the MapReduce algorithms. MapReduce [7] is a programming model that allows a large task to be broken down (or mapped down) into multiple smaller tasks that can be processed as individual jobs. The reduce function combines the output of all the smaller jobs in a specified manner to provide the output of the original large task. The Apache Hadoop framework takes care of job management aspects such as keeping track of which jobs run on which nodes, which jobs complete successfully, which jobs need to be restarted due to failures, and other tasks.

As Hadoop jobs run on distributed cluster, data management is of key importance. Apache Hadoop uses the Hadoop distributed file system (HDFS) to create multiple replicas of data items across different nodes in the cluster. Using this application, data reliability is increased through redundancy. Data also kept close to computing resource that uses it, which increases performance.

The combination of MapReduce algorithms and HDFS enables parallel grid applications to be developed and deployed rapidly and easily, with minimal development time being spent on grid management aspects.

Using the IBM cloud and Hadoop, the CIRG students at the University of Pretoria realize a number of benefits:

- Reduce research computational time from weeks to days
- The solution provides a platform that allows students to rapidly deploy grid applications
- A new mechanism for the research team to use open ideation

- The ability to easily manage and rapidly deploy new capacity to their infrastructure.

[6] A more economical solution for acquiring the necessary computational resources is cloud computing. A common pattern is to have bulk data that needs to be transformed, where the processing of each data item is essentially independent of other data items; that is, using a single-instruction multiple-data (SIMD) algorithm. Hadoop Core provides an open source framework for cloud computing, as well as a distributed file system.

On [8] HPC Cloud is the continuation of clouds main philosophy with one key difference. As virtualization is not suitable for all workloads, HPC Cloud must support both virtualized and direct access computing resources. This allows workloads that can be virtualized to be scaled with demand without interfering with other physical hosts. Virtualization gives HPC a flexibility it has not had before. As the processing core density of computer nodes increases a single operating system starts to make less sense. With virtualization a single node can run multiple operating systems at the same time to allow multiple users to use the same resource. This allows the HPC infrastructure to have a higher utilization from users within its organization and just its continued investment. Virtualization has another key benefit in decoupling a user's job from the physical resource running it. There are mainly two implications for the High Performance Computing (HPC) level of machine utilization in universities [9] . First, the need for HPC machines in universities - it is mainly to solve/compute demanding problems. Second, it is very hard to accommodate the hardware resources for HPC. An example of universities using HPC is a VCL.
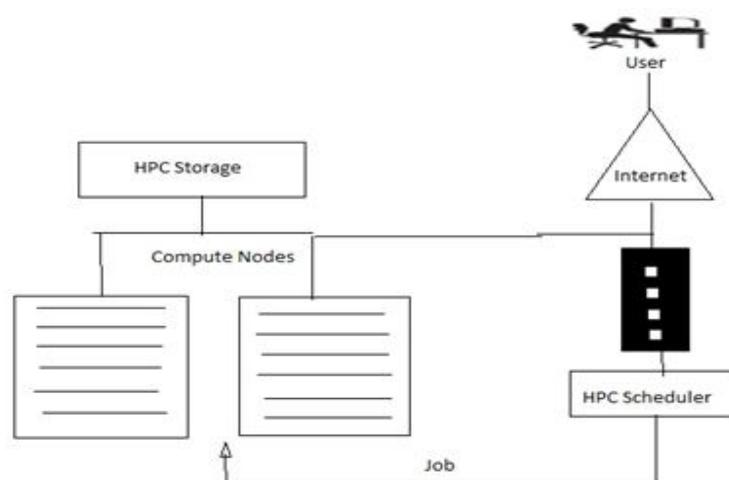

Figure 1: A Simple HPC Model for VCL

III. PROPOSED FRAMEWORK

The proposed framework consists of a private cloud infrastructure and a HPC infrastructure, we use OpenStack [10] open source cloud operating system for deploying a cloud infrastructure and we use Hadoop [6] as an HPC cluster. This technologies are in their own right; however, when they are joined together, the benefits a University experiences are substantial. Although the environment will be complex, an enterprise will see substantial synergies by joining an OpenStack private cloud with an Apache Hadoop environment.
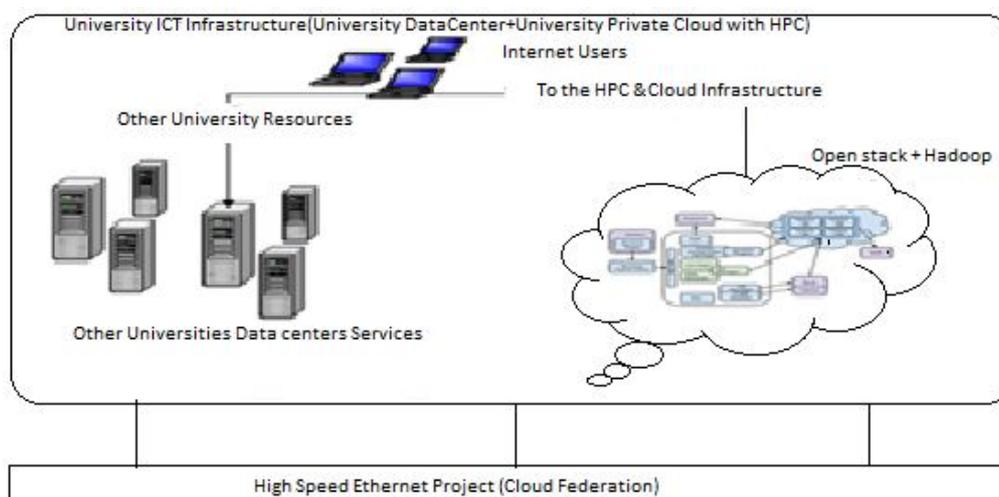

Figure 2: Proposed Framework for Cloud Enabled HPC for Ethiopian Universities

IV. SWIFT, NOVA, APACHE HADOOP, AND MAPREDUCE

The next sections walk through how an organization can integrate this proposed framework to private cloud and HPC technologies. A common deployment model for HPC in a private cloud environment is to deploy Open Stack's Swift storage technology joined to an Apache Hadoop MapReduce cluster for processing. The advantage of using this architecture is that each University will have a scalable storage/compute node.
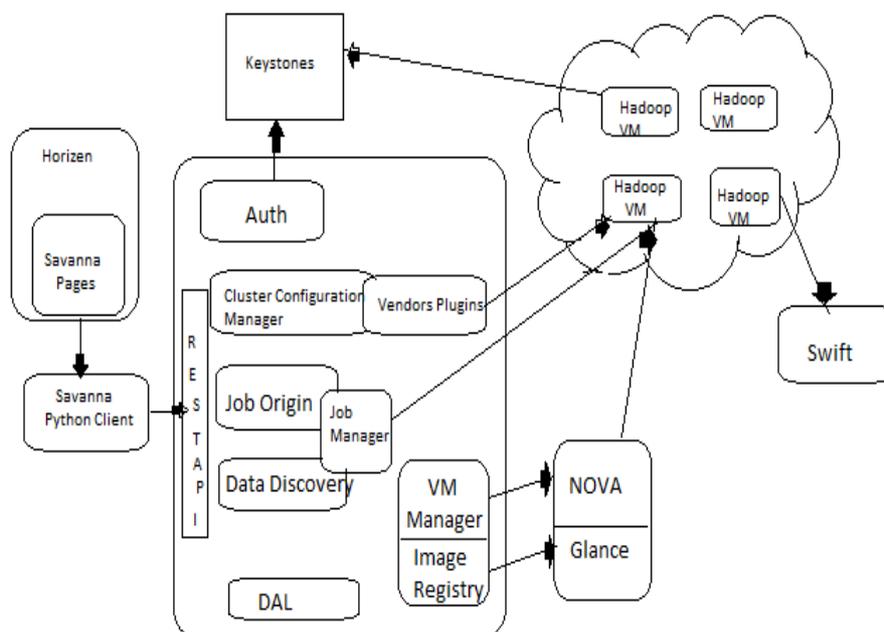
Figure 3: Detail architecture of Hadoop cluster and OpenStack

We can use the advanced level of flexibility, scalability, and autonomy within HPC environment, universities can leverage the native abilities of the open source offerings provided by Apache and OpenStack. To have a fully scalable and flexible HPC environment, it must run on a private cloud environment that provides both storage and compute nodes. To do that, the Universes must build the private cloud first, then add HPC. So, at this point, Swift, Nova and RabbitMQ are certainly needed as well as controller nodes for managing and maintaining the environment.

In order to integerate this technologies which are OpenStack based private cloud infrastructure and Hadoop based luster we use a special API called Savana controller, this controller allow as to give HPC as a service for cloud users. Figure 4, show the savanna API to integrate Hadoop infrastructure and OpenStack Infrastructure.
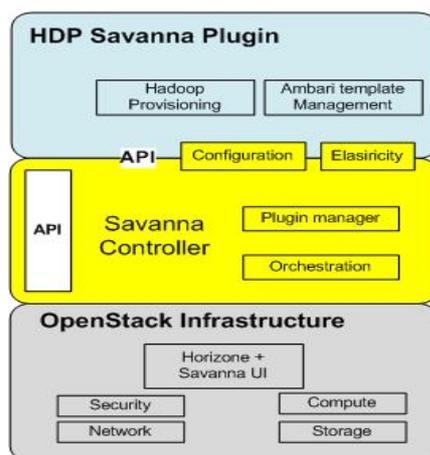
Figure 4: Hadoop on OpenStack architecture

V.The Savanna product communicates with the following OpenStack components:

- Horizon—Provides a GUI with the ability to use all of Savanna's features.
- Keystone—Authenticates users and provides a security token that is used to work with the rest of OpenStack, hence limiting a user's abilities in Savanna to his or her OpenStack privileges.
- Nova—used to provision VMs for Hadoop Cluster.
- Glance—Stores Hadoop VM images each of which contains an installed OS and Hadoop.
- Swift—Used as a storage for data that will be processed by Hadoop jobs.

VI. CONCLUSION AND FUTURE WORK

The future of HPC and cloud computing is bright, with the advent of different intercloud API's, and HPC technologies. This opportunities will answer many problems raised in academia and research centers in regard to on-demand resource provisioning access and high performance cluster computing infrastructure. In this paper we have discussed the potential opportunities and the current state-of-the-art high-performance computing on private clouds for academia. The adoption of Cloud computing as a technology and a paradigm for the new era of computing which occupy cluster of computing nodes in a local and/or remote clouds and has definitely become popular and appealing within academia and research centers. It has also widely spread among end users, students, researchers, which can help them to host their data to the cloud. For what concerns scientific computing, this trend is still at an early stage. We also discuss federated cloud infrastructure to enable universities private clouds to share and collaborate through high performance MoE fiber network backbone.

After this proposed framework which is federated private cloud deployment model that enables HPC, there are issues related to cost, performance and other parameters that we do not focus in this paper. Cloud computing holds a lot of promise with emphasis on HPC applications. The IaaS and PaaS most likely are better fits for hosting HP C applications in the cloud. The author suggest as a future work to measure the performance issues, cost, network latency and other parameters with different deployment models, and HPC. One more thing the author suggests to deploy this model for a universities to get the real advantages of cluster computing (HPC) and cloud computing.

REFERENCES

1. Sanjay P. Ahuja, Sindhu Mani "The State of High Performance Computing in the Cloud", Journal of Emerging Trends in Computing and Information Sciences, ISSN 2079-8407, PP:262-266, VOL. 3, NO. 2, February 2012
2. Qi Zhang , Lu Cheng, Raouf Boutaba "Cloud computing: state-of-the-art and research challenges" Springer, PP: 7-18, April 2010
3. Yuri Demchenko, Canh Ngo, Marc X. Makkes, Rudolf Strijkers, Cees de Laat " Defining Inter-Cloud Architecture for Interoperability and Integration" The Third International Conference on Cloud Computing, GRIDs, and Virtualization, cloud computing 2012
4. Seyyed Mohsen Hashemi, Amid Khatibi Bardsiri , Cloud Computing Vs. Grid Computing:, ARPN Journal of Systems and Software, ISSN 2222-9833, PP: 188-194, VOL. 2, NO.5, MAY 2012
5. Indu Gandotra, Pawanesh Abrol, Pooja Gupta, Rohit Uppal and Sandeep Singh, "Cloud Computing Over Cluster, Grid Computing: a Comparative Analysis", Journal of Grid and Distributed Computing, Volume 1, Issue 1, pp-01-04, 2011
6. Spodarets, Yegorov , Dragan. "Cloud and Grid Technology Based Educational and Research Computing System", Third International Conference "High Performance Computing", HPC-UA 2013 (Ukraine, Kyiv, October 7-11, 2013), PP: 370-373
7. Idafen Santana-P´erez , Mar´ıa S. P´erez-Hern´andez , "A semantic scheduler architecture for federated hybrid clouds", 2012 IEEE Fifth International Conference on Cloud Computing, 978-0-7695-4755-8/12, 2012 IEEE, PP: 384-391
8. Punit Gupta, Deepika Agrawal "Trusted Cloud Platform for Cloud Infrastructure", International Journal of Computers & Technology, ISSN 22773061, Vol. 10, No. 8, S e p t 0 5, 2 0 1 3, PP: 1884- 1891
9. Balu Srinivasulu, Kaliyaperumal Karthikeyan, Mungamuru Nirmala, Merugu Chandramouli "Comparative Analysis of Cloud Computing Utilities", Journal of Emerging Trends in Computing and Information Sciences, Vol. 3, No.10 Oct, 2012, ISSN 2079-8407, PP: 1352-1357
10. V.Sreenivas ,Dr.C.Narasimham" Performance Evaluation of Encryption Techniques and Uploading of Encrypted Data in Cloud",IEEE 2013,Computing, Communications and Networking Technologies (ICCCNT),2013, pp. 1-6, 2013.
11. V.Sreenivas,A.Tejaswi "Efficient Use of Cloud Computing in Medical Science", American Journal of Computational Mathematics¸ Vol:2, PP NO : 240-243 Sep 2012.
12. Wang, L.Z., G. von Laszewski, D. Chen, et al. Provide Virtual Machine Information for Grid Computing [J]. Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans, 277 2010.
13. M. Pokharel, Y.-H. Yoon and J. S. Park, "Cloud Compu-ting in System Architecture," International Symposium on Computer Network and Multimedia Technology, Wuhan, 18-20 January 2009, pp. 1-5.
14. V. Sreenivas and C. Narasimham "A Cloud Model to Implement SaaS," Advanced Materials Research, Vol. 341-342. 2011, pp. 499-503.
15. Khaled M. Khan and Qutaibah Malluhi, "Establishing Trust in Cloud Computing", 2010 IEEE.
16. Manish Pokharel, YoungHyun Yoon, Jong Sou Park, "Cloud Computing in System Architecture", 2009 IEEE.
17. K.A. Delic & J.A. Riley, "Enterprise Knowledge Clouds," Next Generation Km Syst. Int. Conf. Inform., Process, Knowledge Management, Cancun, Mexico, pp. 49–53, 2009.
18. Takada T., et al. Awase-E: Image-Based Authentication for Mobile Phones Using User's Favorite Images. Human-Computer Interaction with Mobile Devices and Services, Springer Berlin/ Heidelberg. 2003. 2795; 347-351.

BIOGRAPHY

Mr. Samuel Fentahuen is a graduate student in Software Engineering from Adam Science and Technology University, Ethiopia. And he has more than 6 years of system administration and system development experience and his area of interest is Cloud Computing, Datacenter, System Development and System Administration, Linux, VMWare and Hyper-V

Dr. Sreenvias Velagapudi: is working as a Assistant Professor in department of Computing at Adama Science and Technology University, Ethiopia. He published 17 papers in International and national journals he also published a text book in area of Software Engineering, he has nearly 8 years of teaching experience in India and abroad. His area of interest are Cloud Computing, software Engineering, Knowledge management, Artificial Intelligence.