A Performance Analysis of EC2 Analysis of Cloud Computing Applications on the Amazon Web Services

K.Siva sankara rao[#1], T.Rajendra Prasad[#2]

[1,2#] **(Kakinada Institute of Engineering and Technology, JNTU Kakinada, A.P, India)**

Abstract:

Cloud Computing is emerging today as a commercial infrastructure that eliminates the need for maintaining expensive computing hardware. Through the use of virtualization, clouds promise to address with the same shared set of physical resources a large user base with different needs. Thus, clouds promise to be for scientists an alternative to clusters, grids and supercomputers. However, virtualization may induce significant performance penalties for the demanding scientific computing workloads. Cloud computing is an emerging commercial infrastructure paradigm that promises to eliminate the need for maintaining expensive computing facilities by companies and institutes alike. Through the use of virtualization and resource time-sharing, clouds serve with a single set of physical resources a large user base with different needs. Thus, clouds have the potential to provide to their owners the benefits of an economy of scale and, at the same time, become an alternative for scientists to clusters, grids, and parallel production environments. Cloud computing has seen tremendous growth, particularly for commercial web applications. The on-demand, pay as you go model creates a flexible and cost effective means to access compute resources. For these reasons, the scientific computing community has shown increasing interest in exploring cloud computing. However, the underlying implementation and performance of clouds are very different from those at traditional supercomputing centers

Key Words: cloud computing, bottleneck, Amazon

Corresponding Author: T. Rajendraprasad (Associate Professor, M.tech CSE Department)

1.Introduction

Scientific computing requires an ever-increase number of resources to deliver results for growing problem sizes in a reasonable time frame. In the last decade, while the largest research project s were able to afford expensive super computers, her project s were forced to opt for cheaper resources such as commodity clusters and grids .Cloud computing pro poses an alternative in which resources are no longer hosted by the researcher 's computation a facilities, but leased from big data centers only when needed. Despite the existence of several cloud computing vendors, such as Amazon an d Go Grid, the potential of clouds remains largely unexplored. To address this issue, in this paper we present a performance analysis of cloud computing services for scientific computing. The cloud computing paradigm holds good promise for the performance hungry scientific community. Clouds promise to be a cheap alternative to supercomputer sand specialized clusters, a much more reliable platform than grids, and a much more scalable platform than the largest of commodity clusters or resource pools. Clouds also promise to "scale by credit card, " that is, scale up immediately and temporarily with the only limits imposed by financial reasons, scientific computing requires an ever-increasing number of resources to deliver results for ever growing problem sizes in a reasonable time frame. In the last decade, while the largest research projects were able to afford (access to) expensive super computers, many projects were forced to opt for cheaper resources such as commodity clusters and grids. Cloud computing proposes an alternative in which resources are no longer hosted by the researchers' computational facilities, but are leased from big data centers only when needed. Despite the existence of several cloud computing offerings by vendors such as Amazon [1] and Go Grid [2], the potential of clouds for scientific computing remains largely unexplored. To address this issue, in this paper we present a performance analysis of cloud computing services for many-task scientific computing. Cloud computing has emerged as an important paradigm for accessing distributed computing resources. Commercial providers such as Amazon, Rack space, and Microsoft, all offer environments for developing and deploying applications in the cloud. There are many definitions of cloud computing, but some characteristics exist in most definitions, e.g., virtualized environments and on-demand provisioning of compute resources. The goal of the recently funded DOE Magellan project is to evaluate the ability of cloud computing to meet DOE's computing needs. The project is evaluating existing commercial cloud offerings and technologies. The purpose of this paper is to examine the performance of existing cloud computing infrastructures and create a mechanism for their quantitative evaluation. Our initial work focuses on Amazon EC2 and its performance, which we believe is representative of current mainstream commercial cloud computing services.

2. Amazon EC2

We identify three categories of cloud computing services: Infrastructure-as-a Service (IaaS) , that is, raw infrastructure and associated middleware, Platform as-a-Service (PaaS) , that is, APls for developing applications on an abstract platform, and Software-as-a-Service (SaaS), that is, support for running software services remotely. The scientific community has not yet started to adopt PaaS or SaaS solutions , mainly to avoid porting legacy applications and for lack of the needed scientific computing services, respectively. Thus, in this study we are focusing on IaaS providers. Unlike traditional data centers , which lease physical resources, most clouds lease virtualized resources which are mapped and run transparently to the user by the cloud's virtualization middleware on the cloud's physical resources. For example, Amazon EC2 runs instances on its physical infrastructure using the open-source virtualization middleware Xen [7] . By using virtualized resources a cloud can serve with the same set of physical resources a much broader user base; configuration reuse is another reason for the use of virtualization. Scientific software, compared to commercial mainstream products, is often hard to install and use . Pre- and incrementally-built virtual machine (VM) images can be run on physical machines to greatly reduce deployment time for software .A Performance Analysis of EC2 Cloud Computing . The Amazon EC2 instance

3. RELATED WORK

A number of different groups have conducted feasibility studies of running their scientific applications in the Amazon cloud. In addition, previous work has examined the performance of individual Amazon AWS components, e.g., the simple storage service Hazel hurst examines the performance of the bioinformatics application WCD [1]. The performance and storage costs of running the Montage workflow on EC2 are detailed by Deelman et. al. [2]. The High-Energy and Nuclear Physics(HENP) STAR experiment has examined the costs and challenges associated with running their analysis application on the EC2 cloud [4]. In previous work we examined the usefulness of cloud computing for e-Science applications Standard benchmarks have also been evaluated on AmazonEC2. Napper et. al. examine the performance of the Linpack benchmarks on different EC2 instance types [5]. The NAS benchmarks have been run by Evangelinos et. al. [6] and Masud [7]. Osterman et al. ran a variety of micro benchmarks and kernels. Rehr. et al. show that Amazon EC2 is a feasible platform for applications that don't need advanced network performance. Wang et al study the impact of virtualization on network performance. This work is unique in examining the performance of a set of applications that represent the typical workload run at a major supercomputing center. The applications chosen represent both the range of science done and the algorithms typical of supercomputing codes. More importantly, by analyzing the running code using IPM we are able to profile the underlying characteristics of the application, and an quantitatively identify the major performance bottlenecks and resource constraints with respect to the EC2 cloud.

3.1 Scientific Computing

Job structure and source:

PPI workloads are dominated by parallel jobs [6], while grid workloads are dominated by small bags-of-tasks (BoTs) [3] and sometimes by small workflows comprising mostly sequential tasks. Source-wise, it is common for PPI grid workloads to be dominated by a small number of users. We consider users that submit many tasks, often grouped into the same submission as BoTs, as proto-MTC users, in that they will be most likely to migrate to systems that provide good performance for MTC workload execution. We focus in Section 3 on a more rigorous definition of MTC workloads, and on demonstrating their presence in recent scientific workloads.

Bottleneck resources:

Overall, scientific computing workloads are highly heterogeneous, and can have either one of CPU, I/O, memory, and network as the bottleneck resource. Thus, in Section 4 we investigate the performance of these individual resources.

Job parallelism:

A large majority of the parallel jobs found in published PPI and grid traces have up to 128 processors [5], [6]. Moreover, the average scientific cluster size was found to be around 32 nodes and to be stable over the past five years. we look at the the performance of executing parallel applications of up to 128 processors.

4. CLOUD Performance Evaluation

In this section we present an empirical performance evaluation of cloud computing services. Toward this end, we run micro-benchmarks and application kernels typical for scientific computing on cloud computing resources, and compare whenever possible the obtained results to the theoretical peak performance and/or the performance of other scientific computing systems.

4.1 Method

Our method stems from the traditional system benchmarking. Saavedra and Smith have shown that benchmarking the performance of various system components with a wide variety of micro-benchmarks and application kernels can provide a first order estimate of that system's performance. Similarly, in this section we evaluate various components of the four clouds introduced in Section 2.2. However, our method is not a straightforward application of Saavedra and Smith's method. Instead, we add a cloud-specific component, select several benchmarks for a comprehensive platform independent evaluation, and focus on metrics specific to large-scale systems (such as efficiency and variability).

Cloud-specific evaluation

An attractive promise of clouds is that they can always provide resources on demand, without additional waiting time . However, since the load of other large-scale systems varies over time due to submission patterns [6] we want to investigate if large clouds can indeed bypass this problem. To this end, one or more instances of the same instance type are repeatedly acquired and released during a few minutes; the resource acquisition requests
.

Infrastructure-agnostic evaluation

There currently is no single accepted benchmark for scientific computing at large-scale. To address this issue, we use several traditional benchmark suites comprising micro-benchmarks and (scientific) application kernels. We further design two types of test workloads: SI–run one or more single process jobs on a single instance (possibly with multiple cores), and MI–run a single multi-process job on multiple instances.

5. Conclusion:

The emergence of cloud computing as the paradigm in which scientific computing is done exclusively on resources leased only when needed from big data centers, e-scientists are faced with a new platform option . However, the initial target of the cloud computing paradigm does not match the characteristics of the scientific computing workloads. Thus, in this paper we seek to answer an important research question Is the performance of clouds sufficient for scientific c*omputing'?* To this end, we perform a comprehensive performance evaluation of a large computing cloud that is already in production. Our main finding is that the performance and the reliability of the tested cloud are low. Thus, this cloud is insufficient for scientific computing at large, though it still appeals to the scientists that need resources immediately and temporarily. With the emergence of cloud computing as a paradigm in which scientific computing can done exclusively on resources leased only when needed from big data centers, e-scientists are faced with a new platform option. However, the initial target workloads of clouds does not match the characteristics of MTC-based scientific computing workloads.

6. REFERENCES

[1] Amazon Inc., “Amazon Elastic Compute Cloud (Amazon EC2),” Dec 2008, [Online] Available: http://aws.amazon.com/ec2/.

[2] GoGrid, “GoGrid cloud-server hosting,” Dec 2008, [Online] Available: http://www.gogrid.com.

[3] A. Iosup, O. O. Sonmez, S. Anoep, and D. H. J. Epema, “The performance of bags-of-tasks in large-scale distributed systems,” in *HPDC*. ACM, 2008, pp. 97–108.

[4] I. Raicu, Z. Zhang, M. Wilde, I. T. Foster, P. H. Beckman, K. Iskra, and B. Clifford, “Toward loosely coupled programming on peta scale systems,” in *SC*. ACM, 2008, p. 22.

[5] A. Iosup, C. Dumitrescu, D. H. J. Epema, H. Li, and L. Wolters, “How are real grids used? The analysis of four grid traces and its implications,” in *GRID*. IEEE, 2006, pp. 262–269.

[6] U. Lublin and D. G. Feitelson, “Workload on parallel supercomputers: modeling characteristics of rigid jobs,” *J.Par.&Distr.Comp.*, vol. 63, no. 11, pp. 1105–1122, 2003.

[7] D. G. Feitelson, L. Rudolph, U. Schwiegelshohn, K. C. Sevcik, and P. Wong, “Theory and practice in parallel job scheduling,” in *JSSPP*, ser. LNCS, vol. 1291. Springer-Verlag, 1997, pp. 1–34.