IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

IJARSCT
Impact Factor: 6.252

Volume 2, Issue 4, March 2022

The State of Cloud-Based High Performance Computing

Dr. M. Mohamed Ismail
Associate Professor, Department of Computer Science
Mazharul Uloom College, Ambur, Tamil Nadu, India

Abstract: *HPC applications have been gaining lot of attention in the cloud computing world. Most of these applications are scientific applications that require large CPU capabilities and are also data intensive requiring large data storage. Traditionally they have always required large number of computers inter-connected in a network such as clusters or supercomputers. These clusters are difficult to setup and maintain both technically and financially. With the advent of cloud computing and benefits of Infrastructure as a Service (IaaS) and Platform as a Service (PaaS), scientists and researchers are able to deploy their HPC applications in the cloud without worrying about the costs associated with the infrastructure and other costs involved. They also give guarantees on the quality of service (QoS). This paper focuses on documenting some of the research already done in the field of HPC applications and their current state in cloud computing.*

Keywords: High Performance Computing, Cloud Computing, EC2, S3

I. INTRODUCTION

The high level of research and some of the biggest IT companies successfully implementing and providing some cloud services for various purposes give us a hint that cloud computing is fast emerging as the next generation technology for computational needs. Cloud refers to both hardware and software applications made available in the network and provided as services to perform tasks over the internet. The Cloud also provides applications as services to store, retrieve and share data from systems connected to the internet. In other words the applications are not actually on the client machine. Large data centers are used to build these 'clouds' that are highly scalable. These data centers usually consist of several thousand interconnected computing devices capable of handling remote requests to run applications. The companies such as Google, Amazon, Sun Microsystems, and Microsoft housing these data centers actually maintain the costs associated with maintaining them and also the software-updates. IT companies are providing these services to the general public for a fee on-demand. This type of service is called Public Cloud [1]. On the other hand if the service is solely used within an organization and not shared with people outside of the organization it is called Private Cloud [2]. There is also a third kind, a combination of public and private cloud. It is referred to as Hybrid [2]. Choosing which one to deploy purely depends on the needs. Two important concerns in a cloud based environment are security and performance [13]. A lot of research is currently underway to analyze how clouds can provide High Performance Computing (HPC) capabilities. Since the data centers are highly data intensive it is imperative that high performing applications should be developed and deployed in the cloud to truly take advantage of them. The purpose of this paper is to explore the current state of high performance computing in cloud.

II. SERVICES IN THE CLOUD

Before getting into HPC it is important to understand how and what type of services are currently out there in cloud computing and how they fit into the above mentioned models.

These are categorized as Software as a service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) [3]. Organizations provide SaaS on demand. So a software application is offered as a service such as Google Apps for managing pictures, email service, calendar etc. Another example is salesforce.com that provides software solutions for sales and marketing on the cloud. The SaaS can be provided to individuals as well as organizations as needed. PaaS lot of times provides developers a platform to build and deploy software applications. The support is provided in the form of OS, development environment, middleware etc. APIs are provided by the providers so developers can interact with the environment to connect to deploy their applications. It addition it also provides tools to maintain these applications.

DOI: 10.48175/IJARSCT-3433

Google App Engines is an example of PaaS that provides an infrastructure and environment application developers. And finally there is IaaS. This layer provides the storage, data center spaces, servers and other networking devices such as routers and also the provisioning computer clusters as needed. The primary purpose of this layer is to handle the workload for computational needs.

As mentioned above the Cloud provides the architecture of hardware and software for computational needs for organizations and consumers. A lot of services provided to consumers focus on web services that rely on relatively less intensive tasks and hence performance is not necessarily an issue in these situations

III. HPC IN THE CLOUD

The real challenge for Cloud is when the computational needs are increased manifold such as for scientific applications that need supercomputer capabilities. Examples could be building 3D models from large amount of data for scientific research and development and Grid computing. Today HPC systems use supercomputers and computer-clusters to solve advanced problems. They involve network of systems with parallel programming capabilities in multiple disciplines such as system software, architecture, computational techniques etc. The HPC technologies provide the tools to build the high performance computing systems. Something similar needs to be done in the cloud world to handle its high performance needs as the applications running on these systems may require hundreds of thousands of CPU-hours[4].

IV. HPC REQUIREMENTS IN CLOUD

As previously stated the cloud must be able to provide services similar to current HPC systems for high performance or at least comparable to them. The HPC systems are currently ranked on TOP500 list [5]. This list ranks the most powerful supercomputers from around the world. High Performance Linpack of the LINKPACK [5] benchmark helps evaluate and gauge the workloads. LINPACK is a software that provides solutions to complex linear equations on computers, supercomputers being one of them. Another requirement for HPC in cloud is scalability. The software applications built on cloud should be highly scalable to adequately handle the workload. Some experiments using the Amazon's EC2 web services with clusters show LINPACK providing comparable results to HPC systems. The results indicated that the performance of the benchmark scaled linearly with size of the cluster [5]. But the memory and network performance were not sufficient to make it to TOP500 list.

Next requirement is network Latency. Latency is the lag in communication between the systems. In networked systems this latency can add up quickly and thus impacting the overall performance. It can be measured by a simple ping test to check how long time it takes for a message to send and receive. Hardware requirements include high performance CPUs with multi-threading capabilities, fast memory and memory bus architecture and I/O devices [6]. These are considered to suit better to support multithreaded jobs.

V. CHALLENGES FOR HPC

5.1 Applications in the Cloud

As more and more applications and content are being hosted and supported in the Cloud there is an ever growing need for supporting high-performance applications too. So researchers are moving their HPC applications to the cloud to understand the bottlenecks and challenges it presents. Models are designed and developed to predict the failures so they can be fixed before they can negatively impact the performance. One such experiment was performed using Microsoft's Azure cloud with Numerical Generation of Synthetic Seismograms [7]. According to the authors in [7] their HPC application creates "seismic waves in three dimensional complex geological media by solving some complex equations". While the results from the experiment were encouraging they also exposed some of the challenges that lie ahead of HPC in cloud. The authors in [7] state that, "Real-time processing is a critical feature for synthetic seismogram". Massive amounts of data are used in computing and hence need the necessary infrastructure to support it. In the current state this is one of the challenges.

As the number of user requests increase the application must be able to support the increasing load. At the same time as the number of requests decrease the application should be able to scale down. So achieving dynamic scalability is a challenge for HPC applications in the cloud. Windows Azure randomly de-allocates the compute nodes when scaling down and hence follows an asynchronous process. This negatively impacts the performance. The traditional HPC applications take longer to schedule and run the applications [7]. So it is important that an effective load balancing infrastructure is in place so scalability can be optimized for performance purposes. Both IaaS and PaaS provide services [8] may be used to build and deploy scalable applications that can be optimized for parallel computing.

In case of IaaS the infrastructure is already built and is readily available for providing services on-demand. Several bottlenecks such as delays, maintenance, operating costs etc. are removed here. Virtual nodes are created ondemand to handle the load and to perform the computing tasks. So providing Virtual resources is another challenge for HPC in the cloud. If a virtual node fails while performing a task it becomes imperative to identify where in the system it failed and why [7]. In order for the performance to be not impacted it is important that the load is transferred to another node while it is identified and fixed. It is also referred to as fault tolerance [7]. This is one of the challenges while designing load balancing systems for high-performance applications. It is possible that a compute node may share its resources in running more than one application. As the number of applications increase the load on the compute node it can decrease the performance and may sometime fail if reached over capacity.

Security of the data hosted in the data centers is also an important point to consider [9]. It is possible that some of the scientific applications may be supported by the government and the data may be considered sensitive. It is important to store the data securely.

The HPC applications are mostly the scientific application focusing on simulating models seismic, earth science, weather etc. These applications are extremely data and memory intensive. So storage, memory and processing bandwidth are also some of the other challenges.

VI. CLOUD BENEFITS FOR HPC

With so many challenges, HPC applications hosted in the cloud also derive several benefits from the cloud. The on-demand Infrastructure as a Service (IaaS) provides the infrastructure to host the HPC applications. Scalable HPC applications can be of immense value for scientific purposes. Researchers, consumers and organizations can use this service to their benefit on pay per use [9] basis to satisfy their needs and avoid setup, maintenance and administrative costs. Further, HPC calculations can be accelerated using parallel processing and greater compute capacity.

Another major benefit for scientists and researchers is that since the data is stored in the cloud they can collaborate with other scientists across the world and share the data working towards achieving common goals. Since there are Service Level Agreements (SLAs) with the service providers, it ensures the Quality of Service [9].

VII. CASE STUDIES AND OPEN PROBLEMS WITH HPC IN THE CLOUD

While the researchers and organizations are dealing with several challenges with deploying and supporting HPC applications in the cloud, they are also dealing with problems with them. The problematic areas include security, latency, storage architecture, scaling efficiency, performance, networking issues etc. In this section discusses some of the tests and experiments performed on Amazon's EC2, S3, and their results and conclusions.

On April 21, 2011 a networking event caused remirroring of large volumes of EBS that Amazon's EC2 instances use. EBS is Amazon's block storage volumes that can be mounted on the EC2 instances. Several volumes can be mounted the same instance. This remirroring created a shortage of capacity and also further prevented the creation of new EBS volume creation in the Availability Zone [11]. This resulted in connectivity issues with the EC2 instances, increased error rates in several portions of the instances and latency issues. These EC2 instances supported hosting of websites for several small and big businesses like foursquare.com, reddit.com, Quora etc. It took several hours for Amazon to restore the EC2 instances in the US-EAST-1 Availability zone.

This exposed some of the storage architecture weaknesses and problems associated with the cloud architectures. Hence, one of the areas that need to be researched more is handling of large data sets in the cloud. Large data set is one of important characteristics of HPC applications. Applications in most of today's clouds compute the data stored in the local disks. It is not clear however, how well this will work with virtual machine instances accessing the data [10].

A group of researchers (Christian Vecchiola*et al*) in the Department of Computer Science at the University of Melbourne performed experiments to analyze the performance and accuracy of classification of gene expression datasets [9] on Aneka cloud hosted on Amazon's EC2. Aneka provides software platform and framework for developing distributed applications on the cloud. It can also be built on other cloud infrastructures such as Amazon's EC2 so the applications

DOI: 10.48175/IJARSCT-3433

can scale on demand. Aneka's cloud consists of several containers as basic building blocks that collectively create a runtime environment for the applications. Profiling of gene expression datasets is important for researchers to understand the relationships between genes and disease [9]. The application used large amounts of data for profiling of the genes. These profiles are then classified. Several classification methods were used to check for their accuracy levels. Among these methods CoXCS classifier [9] scored the best and produced more accurate results than others. However, one of the negative points about this classifier is that it took lot of time for computation. The tests were performed with different sizes of datasets to get data on performance and accuracy. Based on the results they concluded that high-performance applications can do well with scalability with on-demand services. The cloud does provide the resources required to run these application if provisioned to do so. However they point out that applications will perform even better if fully customizable runtime environments can be provided by the clouds.

Another set of experiments was performed by Scott Hazelhurst of School of Electrical and Information Engineering at the University of Witwatersrand, Johannesburg. In one of the experiments he ran HPC application called WCD in three different cloud environments Amazon's EC2, CHCP's iQudu and Meraka C4 Xeon cluster [4]. WCD (pronounced *wicked*) [12] consumes large datasets and is CPU intensive to compute complex algorithms. The first experiment referred to as MPI Experiment 1 in [4] was to test and document the efficiency of the slave processes. This efficiency was measured by calculating the time taken to run with a number of slaves. Based on the output the author concluded that Amazon's EC2 faired midway between iQudu and C4 [4]. Some of the problems or the disadvantages documented were latency due to variable network speeds. The latencies resulted in communication delays in transferring data [4]. These delays can cause significant cost differences. Lack of GUI tools caused some usability issues on Amazon's EC2.

Other experiments from Jaliya Ekanayake and Geoffrey Fox at the School of Informatics and Computing, Indiana University included private cloud infrastructure Eucalyptus [10] running MPI applications. Several MPI applications with different computational characteristics were used for the experimentations on different VM configurations. The goal of working with MPI applications was to analyze the overhead of virtualized resources. Another goal was to understand the performance of VMs on multi-core nodes. The authors of [10] concluded that while clouds provide good quality of service their overhead is very high for application that require complex communication patterns and need large datasets. So, this is one of the areas that needs more research. These limit the working of such applications in the cloud. The clouds provide good fault tolerance and monitoring when compared to conventional HPC systems.

Physical memory to support data intensive applications is another problem in general. A lot of memory is required to run the virtual machines and to compute this data to build models for scientific purposes.

VIII. CONCLUSION

From all the studies and research that resulted in several experiments, cloud computing holds a lot of promise with emphasis on HPC applications. The IaaS and PaaS most likely are better fits for hosting HPC applications in the cloud since the hardware required for the infrastructure and the necessary software layers are already available for consumers, researchers and organizations. These organizations can save a lot of money and overhead cost by taking advantage of applications that can scale dynamically in these infrastructures on a pay per use basis. However, there are several challenges for HPC applications in the cloud too. So far only a handful of organizations provide cloud services. Only some of these organizations have infrastructures that can support small scale HPC applications. Running large scale HPC applications is still an issue that are highly data intensive and that consume high CPU capabilities. Inconsistent results while experimenting with HPC applications indicate a lot still needs to done. When compared to current traditional HPC system the clusters in the cloud fall short a little bit.

REFERENCES

[1]. Gillam, L. and Antonopoulus, N., Cloud Computing: Principles, Systems and Applications, Springer, 2010.

[2]. Velte, A., Velte, T. J. and Elsenpeter, R.C., Cloud Computing: A Practical Approach , McGraw Hill Professional, 2010

[3]. Introduction to Cloud Computing Architecture, White Paper, 1st Edition, June 2009

[4]. Hazelhurst, S., Scientific computing using virtual highperformance computing: a case study using the Amazon elastic computing cloud, SAICSIT Proceedings of the annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries: riding the wave of technology, 2008

[5]. Napper, J. and Bientinesi, P., Can Cloud Computing Reach the TOP500?, Proceedings of the Workshop on UnConventional High Performance Computing, in conjunction with The ACM International Conference on Computing Frontiers, 18-20 May 2009

[6]. Labate, B. and Korambath, P., Use of Cloud Computing Resources in an HPC Environment - - IDREHPC Research Projects, 2009

[7]. Subramanian, V., Ma, H., Wang, L., Lee, E. and Chen, P., Azure Use Case Highlights Challenges for HPC Applications in the Cloud, HPC in the Cloud, (Web: http://www.hpcinthecloud.com) February 21, 2011.

[8]. Masud, R and, Sottile, M. J., High Performance Computing with Clouds, High Performance Computing with Clouds, Technical Report, University of Oregon, CISTR- 2010-06, January, 2010

[9]. Vecchiola, C., Pandey, S. and Buyya, R., High- Performance Cloud Computing: A View of Scientific Applications, Journal reference: Proceedings of the 10th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN 2009, IEEE CS Press, USA), Kaohsiung, Taiwan, December 14-16, 2009

[10]. Ekanayake J., Qiu XH, Gunarathne T., Beason S, Fox G., High Performance Parallel Computing with Clouds and Cloud Technologies , CloudComp2009, 2009

[11]. Dignan, L., Amazon's N. Virginia EC2 cluster down, 'networking event' triggered problems, http://www.zdnet.com/blog/btl/ amazons-n-virginia-ec2- cluster-down- networking-event-triggeredproblems/47679, April 21, 2011.Hazelhurst, S., Algorith

DOI: 10.48175/IJARSCT-3433