iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** II **Month of publication:** February 2016

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

High Performance Cloud Computing

Ms. Sajjan R.S. [1], Ms. Kalburgi Tayyaba Md.Ibrahim[2]
Department of Computer Science and Engineering, V.V.P. Institute of Engineering and technology, Solapur, Maharashtra, India

Abstract— Abstract- HPC applications nowadays are gaining huge attention in cloud computing arena. Majority of the HPC applications are used for scientific applications or purposes. They require huge CPU usage and tremendous data storage capabilities. Previously HPC applications required large set of interconnected computers in a network like clusters or supercomputers. These clusters are infeasible both financially and technically. But nowadays due to the advent of cloud computing scientific fraternity is able to reap the benefits of cloud computing due to the low cost and maintenance associated with HPC applications. Cloud computing also provides the benefits of Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) both beneficial to the scientific community. Also Quality of Service (QoS) is an added advantage. This paper focuses on the present state of HPC applications as well as previous research done in this field.

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

I. INTRODUCTION

THE idea of using clouds for scientific applications has been around for several years, but it has not gained traction primarily due to many issues such as lower net-work bandwidth or poor and unstable performance. Sci-entific applications often rely on access to large legacy data sets and pre-tuned application software libraries. These applications today run in HPC environments with low latency interconnect and rely on parallel file systems. They often require high performance systems that have high I/O and network bandwidth. Using commercial clouds gives scientists opportunity to use the larger re-sources on-demand. However, there is an uncertainty about the capability and performance of clouds to run scientific applications because of their different nature. Clouds have a heterogeneous infrastructure compared with homogenous high-end computing systems (e.g. su-percomputers). The design goal of the clouds was to pro-vide shared resources to multi-tenants and optimize the cost and efficiency. On the other hand, supercomputers are designed to optimize the performance and minimize latency[1].

However, clouds have some benefits over supercom-puters. They offer more flexibility in their environment. Scientific applications often have dependencies on unique libraries and platforms. It is difficult to run these applica-tions on supercomputers that have shared resources with pre-determined software stack and platform, while cloud environments also have the ability to set up a customized virtual machine image with specific platform and user libraries. This makes it very easy for legacy applications that require certain specifications to be able to run. Setting up cloud environments is significantly easier compared to supercomputers, as users often only need to set up a vir-tual machine once and deploy it on multiple instances. Furthermore, with virtual machines, users have no issues with custom kernels and root permissions (within the virtual machine), both significant issues in non-virtualized high-end computing systems.

Scientific computing involves the construction of mathematical models and numerical solution techniques to solve scientific, social scientific and engineering problems. These models often require a huge number of computing resources to perform large scale experiments or to cut down the computational complexity into a reasonable time frame[2].These needs have been initially addressed with dedicated high-performance computing (HPC) infrastructures such as clusters or with a pool of networked machines in the same department, managed by some "CPU cycle scavenger" software such as Condor. With the advent of Grid computing new opportunities became available to scientists: in a complete analogy with the power Grid the computing Grid could offer on demand the horse power required to perform large experiments, by relying on a network of machines, potentially extended all over the world.

II. LITERATURE REVIEW

[1]Jackson has deployed a full application that performs massive file operations and data transfer on Amazon EC2 . The research mostly focuses on different storage options on Amazon. Walkerevaluates the performance of EC2 on NPB benchmarks and compares their performance on EC2 ver-sus NCSAABE supercomputer on limited scale of 1 and 4 instances. The paper suffices to bring the results without detailed analysis and does not identify what this gap contributes to[1]. Other papers have run the same benchmark on different infrastructures and provided bet-ter analysis of the results.

[2] Only a few of the researches that measure the applica-bility of clouds for scientific applications have used the new Amazon EC2 cluster instances that we have test-ed . Mehrotra compares the performances of Amazon EC2 HPC instances to that of NASA's Pleiades supercomputer . However the performance metrics in that paper is very limited. They have not evaluated dif-ferent performance metrics of the HPC instances. Rama-krishnan have measured the performance of the HPCC benchmark They have also applied two real appli-cations of PARATEC and MILC[1]. David Pellerin Stated That Cloud computing helps research and academic organizations, government agencies, and commercial HPC users gain fast access to grid and cluster computing resources, to achieve results faster and with higher quality, at a reduced cost relative to traditional HPC infrastructure.The AWS Cloud transforms previously complex and static HPC infrastructures into highly flexible and adaptable resources for on-demand or long-term use.

[3] Filippo Gioachin [5] Stated That Through a performance analysis of HPC applications and a comparison on a range of platforms, we have shown that different applications exhibit different characteristics that make them more or less suitable to run in a cloud environment. Applications with non-intensive communication patterns are good candidates for cloud deployments. For communicationintensive applications, supercomputers remain the optimal platform, largely due to the overhead of network virtualization in the cloud.

[4] Thomas Sterling [4] Stated That Computer resourses,storage,as well as application ,can be dynamically provisioned on a pay per use basis.The research & organization are dealing with problem than can be tests & experiments performed on Amazon's EC2,S3 & their result .HPC application WCD is Overcome. Computer resourses,storage,as well as application ,can be integrated.the resourses & organization are dealing with several challenge with deploying & supporting HPC application in the cloud.Another experiment HPC application called WCD in three different cloud environment so it occur some issues on Amazon's EC2.

III. CLOUD COMPUTING

A. Defining Cloud

Cloud Computing is type of computing as a service over the Internet on a pay-as-you-use. Cloud computing offers a new way of services by re-arranging various resources and providing them to users based on their demands[5]. It allows for provision of a variety of business and customer services. As organizations experience increasing pressure to change or move organization to respond continuously in changing environment. organizations experience increasing pressure to change or move organized to respond continuously to a changing environment.

General characteristics of cloud computing which will help the development and adoption of this rapidly evolving technology.The serviceoriented conceptional characteristic abstracts the details of inner implementations. As a technical characteristic, the loose coupling is the key technical feature of all kinds of cloud computing systems. The strong fault tolerant makes cloud computing much more adapting the widely adopte network understructure. The economic pattern is the main reason why so many companies jump into the hot pool of cloud computing and distinguish it from other research area such as HPC and grid computing. The ease use user experience characteristic hides the complexity of cloud service providers and supply cloud users with very simple interfaces. There are also other kinds of characteristics such as TCP/IP especially Internet based, virtualization and high security.There will be full war on cloud computing among big giants including Google, Microsoft, Apple, Amazon, IBM, Oracle and HP etc. It is hard to predict who will win the war. It is also hard to predict the impact of cloud computing on both information technology and society [7].

B. Cloud Services

The services of cloud computing are broadly divided into three categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Softwareas-a-Service (SaaS) [7].

Infrastructure-as-a-Service is the delivery of huge computing resources such as the capacity of processing, storage and network. Taking storage as an example, when a user use the storage service of cloud computing, he just pay the consuming part without buying any disks or even knowing nothing about the location of the data he deals with. Sometimes the IaaS is also called Hardware-as-a-Service (HaaS).

Platform-as-a-Service generally abstracts the infrastructures and supports a set of application program interface to cloud applications. It is the middle bridge between hardware and application. Because of the importance of platform, many big companies want to grasp the chance of predominating the platform of cloud computing as Microsoft does in personal computer time. The well known examples are Google App Engine [7] and Microsoft's Azure Services Platform.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Software-as-a-Service aims at replacing the applications running on PC. There is no need to install and run the special software on your computer if you use the SaaS. Instead of buying the software at a relative higher price, you just follow the pay-per-use pattern which can reduce you total cost. The concept of SaaS is attractive and some software runs well as cloud computing, but the delay of network is fatal to real time or half real time applications such as 3D online game.

Saas is the cloud environment where the application developed by the developer is provided to customers. The customer then uses the software on demand through the internet (world wide web) . Just like a all world data is in small smart phone.[1]

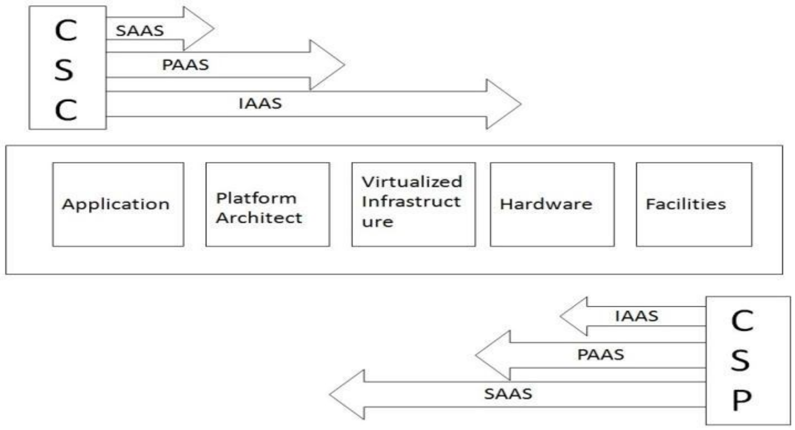


Figure 1. cloud services

IV. HIGH PERFORMANCE CLOUD COMPUTING

Cloud computing presents a unique opportunity for batch processing and analytics jobs that analyze terabytes of data and can take hours to finish. If there is enough data parallelism in the application, users can take advantage of the cloud's new "cost associativity": using hundreds of computers for a short time costs the same as using a few computer for a long time Programming abstractions such as Google's MapReduce and its open-source counterpart Hadoop allow programmers to express such tasks while hiding the operational complexity of choreographing parallel execution across hundreds of cloud computing servers.

Some works with MapReduce has already been done and tested over the clouds. Again, the cost/benefit analysis must weigh the cost of moving large datasets into the cloud against the benefit of potential speedup in the data analysis[1]. When we return to economic models later, we speculate that part of Amazon's motivation to host large public datasets for free may be to mitigate the cost side of this analysis and thereby attract users to purchase cloud computing cycles near this data.

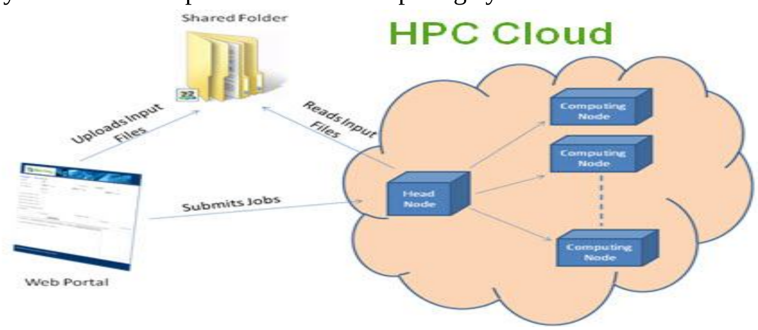


Fig 2.HPC In Cloud

Some commercial HPC applications that have been deployed with clouds have been described by focusing the nature of the application and the commercial benefits of the deployment with the clouds. For example, the Server Labs, Pathwork Diagnostics, Cycle computing and Atbrox and Lingit.Nonetheless, the cloud computing model, in spite of its promise, either imposes constraints in conflict with some HPC requirements or simply fails to adequately support them[2].Among these constraints is the underlying hardware architecture virtualization, which is valuable for generic usage of diverse cloud resources. Such resources generally provide portability but obstruct targeting algorithm optimizations to specific hardware structures, as is typical of HPC applications. The time-critical overhead that virtualization layers add further degrades the performance efficiency and scalability of some HPC workloads. Another performance issue related to clouds is that users share resources among multiple tasks for both computational and networking functionality. The resulting resource contention inserts sporadic andunpredictable delays, further degrading performance and making optimizations more difficult.

A. High Performance Computing As A Service

Because of current commodity clouds' limitations in serving general HPC applications, cloud providers realized the need to provide cloud solutions that are built specifically for HPC applications (for example, hardware with faster processors and interconnects). Some providers have even provided nonvirtualized hardware to deliver the bareto-the-bone performance that these applications require. This is commonly referred to as HPC as a Service—that is, running HPC applications on HPC resources that are exposed as on-demand services using cloud computing abstractions—to take advantage of the cloud model without sacrificing the HPC performance that scientific applications require.The two main approaches for providing HPC as a Service. The first approach develops abstractions enabling large HPC . HPC as a Service classification. Large HPC systems Smaller HPC clusters connected together Physical solutions Physical solutions Virtualized solutions Example: Blue Gene/P as a Service (CometCloud) Examples: Penguin on Demand, Silicon Graphics, and HPCynergy Examples: EC2 Cluster, Azure, Adaptive Systems, and CloudCycle Computing in Science & Engineering systems to be provisioned as clouds. We illustrated this approach in previous work using the IBM Blue Gene/P supercomputer[2].

The second approach uses HPC clusters that can be connected to form a large cloud. These HPC clusters can be virtualized or nonvirtualized to provide better performance. The Amazon EC2 Cluster, Azure, Adaptive Systems, and Platform HPC Computing are commercial providers that use the virtualized approach; Silicon Graphics and Penguin Computing on Demand are commercial providers that provide access to nonvirtualized resources[2]. Andrzej Goscinski and his colleagues12 proposed a similar approach in an academic environment, developing the HPCynergy system to provide HPC as a Service using small clusters. In an early experiment using CometCloud, we explored how the Cloud abstraction can be effectively used to provide a simple interface for current HPC resources and support real-world applications. In particular, we experimentally validated the Cloud paradigm's benefits, such as ease of use and dynamic allocation, and its application to supercomputers— specifically, an IBM Blue Gene/P system.

B. Benchmarking Hpc Performance

It is important for us to use wide-spread benchmarking tools that are used by the scientific community. Specifical-ly in Cloud Computing area, the benchmarks should have the ability to run over multiple machines and provide accurate aggregate results. For memory we use CacheBench. We perform read and write benchmarks on single instances. For network bandwidth, we use Iperf. For network latency and hop distance between the instances, we use ping and trac-eroute. For CPU benchmarking we have chosen HPL benchmark [1]. It provides the results in floating-point operations per second (FLOPS). In order to benchmark S3, we had to develop our own benchmark suite, since none of the widespread bench-marking tools can be used to test storage like this.

To gain thorough insights into the performance of selected platform, we chose benchmarks and applications from different scientific domains and those which differ in the nature, amount, and pattern of inter-processor communication. Similarly to previous work [5], we used NAS Parallel Benchmarks (NPB) class B (the MPI version, NPB3.3-MPI), which exhibit a good variety of computation and communication requirements. Moreover, we chose additional benchmarks and real world applications, written in two different parallel programming environments – MPI [18] and CHARM++ :

The scaling behavior of our testbeds for the selected applications. These results are averaged across multiple runs (5 executions) performed at different times. We show strong scaling results for all applications except Sweep3D, where we chose to perform weak scaling runs. For NPB, we present results for only embarrassingly parallel (EP), LU solver (LU), and Integer sort (IS) benchmarks due to space constraints. The first observation is the difference in sequential performance: Ranger takes almost twice as long as the other platforms, primarily because of the older and slower processors. The slope of the curve shows how the applications scale on different platforms. Despite the poor sequential speed, Ranger's performance crosses Open Cirrus, private cloud and public cloud for some applications at around 32 cores, yielding a much more linearly scalable parallel performance.

Iinvestigated the reasons for better scalability of these applications on Ranger using application profiling, performance tools, and microbenchmarking and found that network performance is a dominant factor (Section 4). We observed three different patterns for applications on these platforms. First, some applications such as EP, Jacobi2D, and NQueens scale well on all the platforms up to 128–256 cores. The second pattern is that some applications such as LU, NAMD, and ChaNGa scale on private cloud till 32 cores and stop scaling afterwards. These do well on other platforms including Open Cirrus. The likely reason for this trend is the impact of virtualization on network performance [1]-[5].

C. High Performance Computing 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 HPCvsystems are currently ranked on TOP500 list . This list ranks the most powerful supercomputers from around the world. High Performance Linpack of the LINKPACK 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[5]. 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 . 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[5] . These are considered to suit better to support multithreaded jobs.

D. Challenges For High Performance Computing Applications In The cloud

Moving HPC jobs to the cloud is a cultural, as well as technical, issue that affects users and IT infrastructure administrators. Users frequently act as if on- premise HPC resources are cost- free and thus don't always utilize them carefully. On the other hand, they appear more aware that the cloud is not free and must be accessed wisely. Users would benefit from a tool that helps them decide whether and how to use on-premise or cloud resources for various tasks.

Such a tool would be of value, because although the cloud offers many advantages, it also presents challenges. For example, cloud use entails latency. Tightly coupled parallel applications require processors to communicate among themselves via a high- speed network. Without such a network, many parallel applications don't scale well, causing users to choose to work with on- premise resources[3]. A significant bottleneck occurs between the user infrastructure (including systems ranging from laptops to clusters) and the cloud.

This can ruin the experience for users, who expect quick access to HPC cloud applications' output for purposes such as visualization and analysis. UberCloud—an online community and marketplace where engineers and scientists discover, try out, and buy computing as a service—reports challenges that companies face when moving HPC workloads to the cloud[3].The US Department of Energy's Magellan Report on Cloud Computing for Science contains analyses on running HPC and data- intensive applications in the cloud.

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 [5]. 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 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 [5].

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 . This isone 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 [2].

It is possible that some of the scientific applications may be supported by the government and the data may be considered

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

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.

E. Cloud Benefits For High Performance Computing

Evaluate the performance of High Performance Computing applications on a range of platforms varying from supercomputer to cloud. Also, we analyze bottlenecks and the correlation between application characteristics and observed performance, identifying what applications are suitable for cloud. Evaluate the performance when running the same benchmarks on exactly same hardware, withoutand with different virtualization technologies, thus providing a detailed analysis of the isolated impact of virtualization on High Performance Computing applications .

To bridge the divide between High Performance Computing and clouds, we present the complementary approach of (1) making High Performance Computing applications cloud-aware by optimizing an application's computational granularity and problem size for cloud and making clouds HPCaware using thin hypervisors, OS-level containers, and hypervisor- and application-level CPU affinity, addressing – how to use cloud for HPC[5]. We investigate the economic aspects of running in cloud and discuss why it is challenging or rewarding for cloud providers to operate business for HPC compared to traditional cloud applications. We also show that small/medium-scale users are the likely candidates who can benefit from an HPC-cloud.

Instead of considering cloud as a substitute of supercomputer, we investigate the co-existence of multiple platforms – supercomputer, cluster, and cloud. We research novel heuristics for application-aware scheduling of jobs in this multi-platform scenario significantly improving average job turnaround time and job throughput , compared to running all jobs on supercomputer. The insights from performance evaluation, characterization, and multi-platform scheduling are useful for both – HPC users and cloud providers. Users can better quantify the benefits of moving to a cloud. Cloud providers can optimize the allocation of applications to their infrastructure to maximize utilization and turnaround times.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 [5] 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 [5].

V. 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[4].

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[4]. 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[7].

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 [2] 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 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 [2].

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 [2] 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 [6]. WCD (pronounced wicked) 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 [2]. 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 [5] 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 [2] 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.

VI. BENEFITS

HPC as a Service provides users with a number of key benefits as follows.

A. HPC resources scale with demand and are available with no capital outlay-only the resources used are actually paid for.

B. Experts in high-performance computing help setup and optimize the software environment and can help trouble-shoot issues that might occur.

C. Faster time-to-results especially for computational requirements that greatly exceed the existing computing capacity.

D. Accounts are provided on an individual user basis, and users are billed for the time they use service.

E. A HPC platform for you and your applications: Support for ANSYS, OpenFOAM, LSTC, etc ... and third party support.

F. Access from anywhere in the worlds with high-speed data transfer in and out.

VII. FUTURE WORK

However, many anticipated properties of distributed cloud environments strongly suggest that clouds can only partly address HPC user needs and that some workload subdomains will remain beyond the capabilities of cloud services. Virtualization, uncertainty of hardware Structural details,lack of network control and memory access contention,repeatability,and protection and security all inhibit cloud paradigm adoption for certain critical uses.

Also, it's unlikely that a general business model, implicit with clouds, will provide the extreme computing and peak performance.

Finally, protected access to such facilities is a potential source of competitive edge for science, market, and national security, and the agencies that employ them will therefore limit or entirely preclude offering such systems to a cloud-covered processing world.

VIII. CONCLUSIONS

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 rganizations have infrastructures that can support smallscale 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] Iman Sadooghi, Jesús Hernández Martin, Tonglin Li, Kevin Brandstatter, Ketan Maheshwari, Tiago Pais Pitta de Lacerda Ruivo, Gabriele Garzoglio, Steven Timm,Yong Zhao, Ioan Raicu.” Understanding the Performance and Potentialof Cloud Computingfor Scientific Applications” IEEE TRANSACTIONS ON CLOUD COMPUTING, MANUSCRIPT ID.

[2] Manish Parashar, Moustafa AbdelBaky, Ivan Rodero, and Aditya Devarakonda.” Cloud Paradigms and Practices for Computational and Data-Enabled Science and Engineering”. IEEE Copublished by the IEEE CS and the AIP.

[3] Marco A.S. Netto, Renato L.F. Cunha, and Nicole Sultanum,” Deciding When and How to Move HPC Jobs to the Cloud**”** COMPUTER PUB L I SHE D B Y THE IEEE COMPUTE R SOC I E T Y.

[4] Thomas Sterling and Dylan Stark.” A High-Performance Computing Forecast: Partly Cloudy”, 1521-9615/09/$25.00 © 2009 IEEE Copublished by the IEEE CS and the AIP.

[5] Abhishek Gupta, Paolo Faraboschi, Fellow, IEEE, Filippo Gioachin, Laxmikant V. Kale, Fellow, IEEE,Richard Kaufmann, Bu-Sung Lee, Verdi March, Dejan Milojicic, Fellow, IEEE, and Chun Hui Suen.” Evaluating and Improving the Performance and Scheduling of HPC Applications in Cloud”, 2168-7161 (c) 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

[6] Robert C. Green, II, Lingfeng Wang, and Mansoor Alam.” Applications and Trends of High Performance Computing for Electric Power Systems: Focusing on Smart Grid”, IEEE TRANSACTIONS ON SMART GRID, VOL. 4, NO. 2, JUNE 2013.

[7] Chunye Gong, Jie Liu, Qiang Zhang, Haitao Chen and Zhenghu GongDepartment of Computer Sciences National University of Defense Technology.” The Characteristics of Cloud Computing”, 1530-2016/10 $26.00 © 2010 IEEE DOI 10.1109/ICPPW.2010.45.

[8] Craig A. Lee, Member, IEEE, Samuel D. Gasster, Senior Member, IEEE, Antonio Plaza, Senior Member, IEEE, Chein-I Chang, Fellow, IEEE, and Bormin Huang.” Recent Developments in High Performance Computing for Remote Sensing: A Review”, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 4, NO. 3, SEPTEMBER 2011.

[9] Ye Xiaotao1, Lv Aili1, Zhao Lin2, Research of High Performance Computing With Clouds Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT ’10) Jiaozuo, P. R. China, 14-15,August 2010, pp. 289-293.

[10] Christian Vecchiola1, Suraj Pandey1, and Rajkumar Buyya1,2”High-Performance Cloud Computing**: A** View of Scientific Applications’” Cloud computing and Distributed Systems (CLOUDS) Laboratory Department of Computer Science and Software Engineering.The University of Melbourne, Australia {csve, spandey, raj}@csse.unimelb.edu.au 2Manjrasoft Pty Ltd., Melbourne, Australia.

[11] Abhishek Gupta, Laxmikant V. Kale, Filippo Gioachin. **“**The Who, What, Why and How of High Performance Computing Applications in the Cloud”, HP Laboratories,HPL-2013-49.

[12] Benjamin P. Berman, Phil Maechling, Bruce Berriman, Gideon Juve, Ewa Deelman, Karan Vahi, Gaurang Mehta.” Scientific Workflow Applications on Amazon EC2” National Science Foundation under the SciFlow (CCF-0725332).

[13] Sanjay P. Ahuja, 2 Sindhu Mani.” The State of High Performance Computing in the Cloud” School of Computing, University of North Florida, Jacksonville, FL 32224.

[14] David Pellerin, Dougal Ballantyne, Adam Boeglin.A Document On “An Introduction to High Performance Computing on AWS” Amazon Web Services – August 2015.

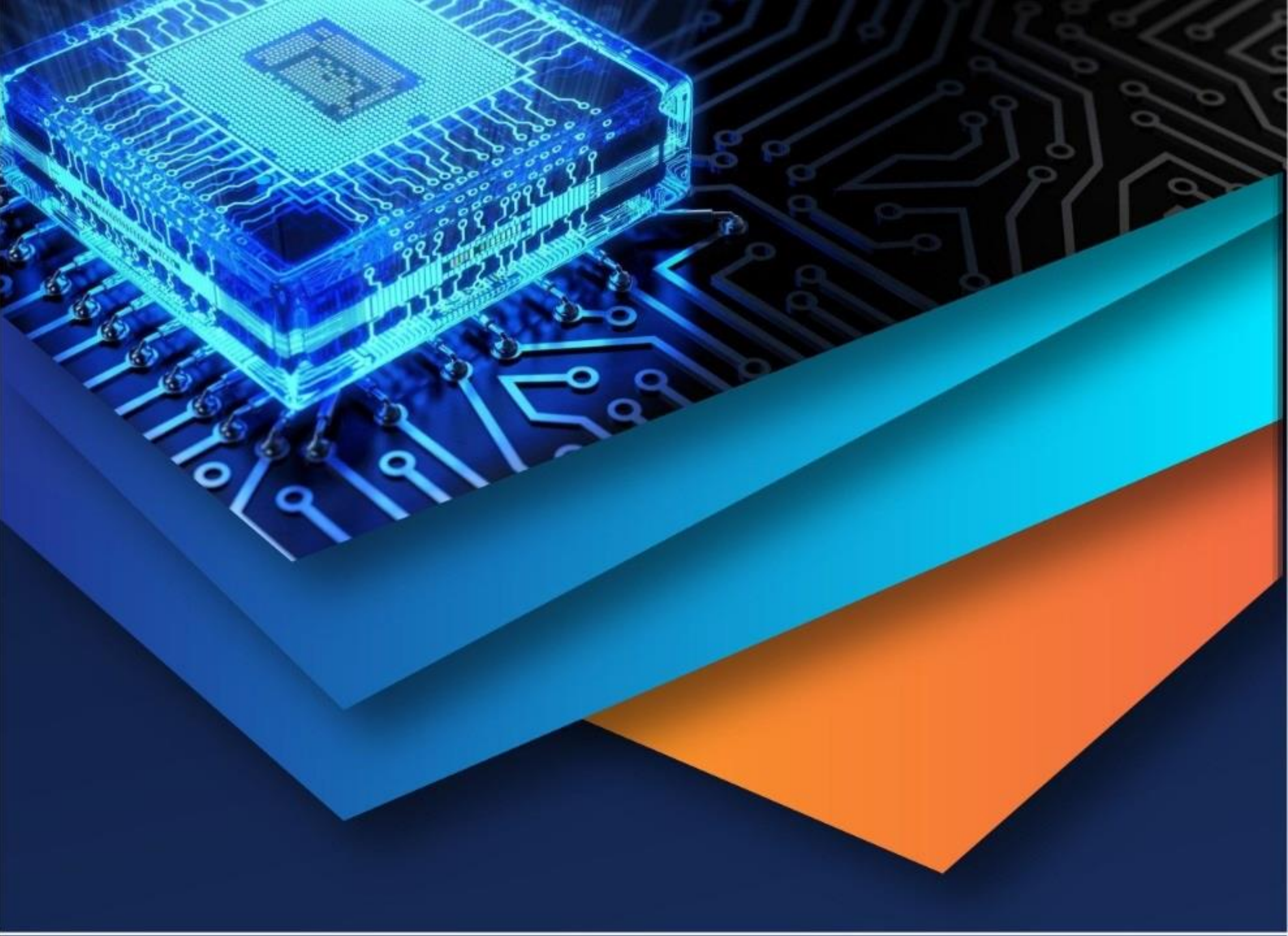
[15] Jaliya Ekanayake and Geoffrey Fox” High Performance Parallel Computing with Clouds and Cloud Technologies” School of Informatics and Computing, Indiana University, Bloomington, IN 47405, USA

[16] Raihan Masud” High Performance Computing with Clouds” raihan@cs.uoregon.edu.

[17] Shujia Zhou, , Ben Kobler, Dan Duffy,” Case Study for Running HPC Applications in Public Clouds” NASA Goddard Space Flight Center {shujia.zhou,benjamin.kobler,daniel.q.duffy,thomas.a.mcglynn}.

[18] Karthik Paladugu Sumanth Mukka,” Systematic Literature Review and Survey on High Performance Computing in Cloud” Master’s Thesis Electrical Engineering September 2012.

[19] S. Wagner • A. Bode • H. Satzger • M. Brehm.A Book On” High Performance Computing in Science and Engineering Garching/Munich 2012”.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY