International Journal of Scientific Research and Engineering Development-– Volume X Issue X, Year

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

High Performance Computing As A Service

Chetan B*, Dr.Krishnappa H.K**

*(Computer Science and Engineering, R V College of Engineering, Bengaluru, India
Email: chetanb.cs17@rvce.edu.in)

** (Computer Science and Engineering, R V College of Engineering, Bengaluru, India
Email: krishnappahk@rvce.edu.in)

----------**********----------

Abstract:

This paper highlights the advantage of using a high performance computing as a service in the modern fast growing and changing world. This paper also highlights cost benefits of high performance computing as a service.

Index Terms—High performance as a service, Cloud

----------**********----------

I. INTRODUCTION

The Supercomputer once were reserved for government, medical researchers, academia and forward-thinking movie makers. Today due to huge serge in data intensive technologies like artificial intelligence and machine learning, which demanded what IDC describes as massive parallel compute capabilities to function properly, has broader variety of enterprises are looking into high-performance computing (HPC) solutions.

A High Performance Computing

High performance computing is a method of processing large amounts of data at rapid speeds by combining several computers and storage devices into a cohesive fabric. HPC makes it easy to find solution for the problem which are data intensive and which need huge amount of processing. HPC has a greater application in artificial intelligence and machine learning industries as they tend to analyze huge size of data-set to provide more and more accurate results. To build HPC system we need 3 key components: 1)compute 2)network 3)storage. Compute servers are clustered together to provide a high-performance computing architecture. On the cluster's servers, software programs and algorithms operate in parallel.To capture the output, the cluster is connected to the data storage.Compute servers are clustered together to provide a high-performance computing architecture. On the cluster's servers, software programs and algorithms operate in parallel. To capture the output, the cluster is connected to the data storage. These components work together to execute a wide range of activities in a seamless manner.

B As A Service

As a service model falls under the wide category of cloud computing and remote access. It highlights the large number of products, tools, and technology that are now available as a service over the internet to users. Essentially, any IT function can be remodeled into a service that can be consumed by businesses. The payment model is also changed to pay-for-use rather than as an upfront purchase or license. Cloud computing created basis for as a service model and also technologies such as IOT (Internet of things) and edge computing have helped it to grow rapidly and forge strong foundation for its future.

II. ARCHITECTURE

In this section we will be focusing on a high-level view of a HPC architecture in a cloud environment. The components are arranged in figure 1 and they are explained in the following subsections.

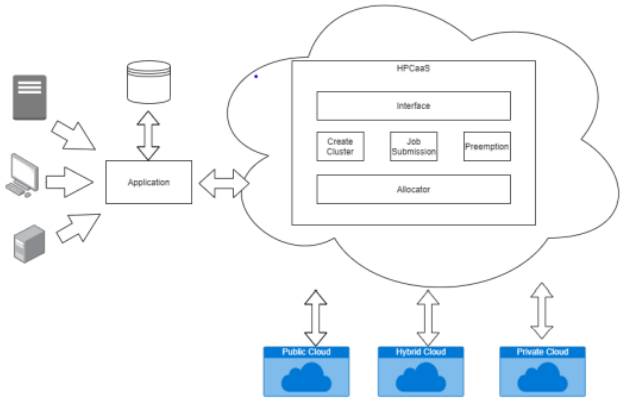

Fig. 1 Architecture of HPC in cloud environment

C HPCaaS

This is the key component of the architecture. It provides a graphical user interface for developing and deploying applications or services that need the setup of a cluster architecture for HPC, job submission, preemption services, and optimal allocation on computing clouds. Figure 1 projects the idea of enabling platform as a service for HPC services. This architecture supports various types of clouds such as private, public and hybrid for resource allocation.

D Create Cluster

This component is in charge of creating a cluster on the cloud. This component takes data computational clouds as input, for which it must create virtual machines, configure them, and enable HPC programs to run. It takes advantage of allocator component to offer communication and cluster deployment in the cloud

E Preemption

This module is responsible for scheduling policies that allow the HPC platform to interleave between processes (cluster in the cloud). Some scheduling policy techniques that the user can specify are permitted to be used here. This component is used to facilitate the the various jobs to access the shared resource fairly.

F Job Submission

This component is in charge of binding the HPC application and kick start the services and application which are running on the cloud.

G Allocator

This component plays a vital role in platform's cloud communication. It is in charge of credentials, user creation and use, virtual machine creation (typically in response to request of Create Cluster component), and cloud component and service deployment. It's also responsible for allocating resources in the configured clouds.

H Personal Database

The data from the user's applications is stored in this database.Personal data settings are frequently saved by application for later modification or adaptation to user needs. Furthermore, using these data, we may track resource utilisation, conduct audits, and make suggestions based on user behaviour, such as scheduling restrictions.This database can be accessed locally or over a computing cloud.

I Application

Any platform can be used to construct the app (e.g.,mobile, desktop or web).The developer/user can take benefit of HPC applications running cloud services after the application implements the accessible functionalities from HPCaaS. Furthermore, the service provider can make this application available to users. It is feasible to create SaaS (Software as a Service) services for access to HPC cloud services using this component.

III. WORKING OF HPCAAS

The batch-processing concept, which dates back to the early mainframe period, is similar to HPC as a service. At their most basic level, HPC services are a cluster of servers connected by an interconnect. You have the capacity to use this virtual computing infrastructure to bring together a large number of different servers in a parallel construct to solve the problem when you submit it. Multiple computer servers are networked together to form HPC clusters in the HPC architecture. Each cluster is made up of

hundreds or thousands of computer servers known as nodes, each of which is equipped with an HPC processor. These nodes work in tandem to increase processing rates and provide great performance. On the cluster's servers, algorithms and software packages are also run. To obtain the results, the cluster is then networked to data storage. Each of these components works together to complete a variety of activities.

IV. ADVANTAGE OF HPC IN CLOUD ENVIRONMENT

There are several advantage of using hpc as a cloud service such as speed, cost, fault tolerance etc. although benefits may vary from system to system.

Characteristic	On-premises HPC resource	HPCaaS
Flexibility	Low	High
Reliability	Low	High
Setup Time	High	Low
Maintenance	expensive and complex	N/A
Resource Utilization	Low	High
Scalability	Low	High
Flexibility	Low	High

Table 1

J Maintenance and Administration

The majority of HPC users in STEM fields are not computer scientists, but engineers, biologists, physicists, and other academic scholars with little or no computing experience. They would rather avoid the headaches of managing and maintaining HPC clusters or supercomputers. Nonetheless, most on-campus HPC users are currently required to hire a server administrator to keep the "on-premises" hardware and software resources up to date. In contrast, a cloud service provider keeps the infrastructure up to date for running applications, freeing users from the hassles of resource management. Galaxy is a good example of a cloud-based tool that adds a layer of transparency to life science computational research. For their data-intensive biomedical research, biologists can use the interactive web-based tool to use predefined software libraries and tools.

K Resource Utilization

Cloud computing, in general, provides business and online applications with a dynamic and scalable foundation. In other words, cloud resources can be dynamically scaled up or down depending on user and application demand. When it comes to HPC users in academic settings, the number of such demand surges will grow. HPC customers may not need the resources for a long time or may have a large unexpected demand for the resources from time to time, depending on the scheduling of scientific research. As a result, leveraging cloud-based resources and, in particular, taking advantage of the pay-per-use model, capital expenditure will be converted to operational cost.

L Performance to Cost Ratio

When compared to a physical cluster with the same setup, a cloud-based virtual cluster often has a performance disadvantage. The pay-as-you-go concept and the scalability of cloud resources can lower the cost of running HPC programmes with little communication and I/O overhead.

V CHALLANGES

Several studies and analyses compared the performance of HPC applications on the cloud to on-premises infrastructure, concluding that today's clouds cannot compete with supercomputers

M Virtualization Overhead

Virtualization is important in the cloud because it allows for more elasticity, resource pooling, and flexibility. Virtualization, and particularly the hypervisor, however, adds unnecessary cost by introducing a software layer that prevents applications from having direct access to hardware resources. The virtualization overhead

varies depending on the type of hardware. Virtualization overhead for CPUs is substantially lower than network virtualization overhead due to hardware support. Passing through GPUs rather than virtual GPUs is often more efficient for particular hardware types, such as GPUs.

N Network Bandwidth and Latency

Several tenants share network interconnects and I/O resources in the cloud. As a result, network capacity and latency may be unpredictable. In most circumstances, the available bandwidth is far less than anticipated. Furthermore, the network latency on the cloud is not consistent. As a result, HPC applications, particularly data-intensive ones, would suffer performance degradation.

O Multi-tenancy

One of the properties of the cloud is multi-tenancy. It is also one of the cloud's revenue-generating aspects for cloud providers. It allows cloud service providers to share resources across multiple tenants. The number of tenants sharing a single cloud resource is referred to as the degree of multitenancy. Cloud providers can overprovision resources to users by increasing the degree of multi-tenancy. Multi-tenancy is the exact opposite of what HPC requires. HPC applications require direct access to dedicated hardware and batch scheduling, but cloud-based shared resources hamper HPC application performance expect HPC to continue to have an impact on the industry as a whole, and to have an even greater influence in regard to 5G and the prospects it presents.

VI CONCLUSION

This paper explores potential benefits of High Performance Computing as a Service. The use of High Performance Computing is predicted to skyrocket in the next years, as data and analytics become such an important corporate tool.This is primarily due to the increased accessibility of HPC solutions, made possible by HPCaaS.This technology gives telcos the information they need to improve their product offerings, find new revenue streams, and improve customer service. We expect HPC to continue to have an impact on the industry as a whole, and to have an even greater influence in regard to 5G and the prospects it presents

ACKNOWLEDGMENT

We would like to thank all the faculty of RV College of Engineering and others for their knowledge transfer and the support provided for this work on HPCaaS.

REFERENCES

[1] Dr. Hassan Rajaei, Saba Jamalian"HPC as a Service in Education," ASEE's 123rd Annual., no. 16060, June 2016.

[2] Emanuel Coutinho,Gabriel Paillard,Ernesto Trajano de Lima,Leonardo Moreira, "An Architecture Proposal for High Performance Computing in Cloud Computing Environments", 4th International Workshop on ADVANCEs in ICT INfrastructures and Services (ADVANCE 2015)At: Recife., December 2015.

[3] NetApp. "What Is High-Performance Computing (HPC)? How It Works." NetApp, NetApp, 1 Jan. 1970, www.netapp.com/data-storage/high performance-computing/what-is-hpc/.

[4] "What Is High-Performance Computing?" What Is High-Performance Computing? — Oracle India, www.oracle.com/in/cloud/hpc/what-is-high performance-computing/.

[5] 15, June, et al. "SaaS vs PaaS vs IaaS: What's The Difference amp; How To Choose." BMC Blogs, 15 June 2019, www.bmc.com/blogs/saas-vs paas-vs-iaas-whats-the-difference-and-how-to-choose/ref1.

[6] L. O. Moreira,V.A.E.deFarias, F.R.C.Sousa,G.A. C.Santos, J. G. R. Maia, and J. C. Machado, "Towards improvementson the quality of service for multi-tenant RDBMS in the cloud,"inWorkshops Proceedings of the 30th International Conference onData Engineering Workshops, ICDE 2014, Chicago, IL, USA, March31 - April 4, 2014.IEEE, 2014, pp. 162–169. [Online]. Available:http://dx.doi.org/10.1109/ICDEW.2014.6818322

[7] Y.Chi, H. J. Moon, H. Hacig um us , and J. Tatemura, "Sla tree:A framework for efficiently supporting sla-based decisions in cloudcomputing," inProceedings of the 14th International Conferenceon Extending Database Technology, ser. EDBT/ICDT '11.NewYork, NY, USA: ACM, 2011, pp. 129–140. [Online]. Available:http://doi.acm.org/10.1145/1951365.1951383.

