PARALLEL CODE ON CLOUD PLATFORMS AND HPC EQUIPMENTS

Author: ILKER OZTURK

Date: 18 November 2021

Authorisation: Module Leader

Distribution: All students

School of Computing and Mathematical Sciences



Table of Contents

[1. Definition of cloud computing and how it can be useful 4](#_Toc89290782)

[1.1 Cloud Computing definition 4](#_Toc89290783)

[1.2 Deployment Models 4](#_Toc89290784)

[1.3 Service Models 4](#_Toc89290785)

[1.4 How cloud computing can be beneficial? 5](#_Toc89290786)

[2. Comparing HPC Cloud and traditional HPC 6](#_Toc89290787)

[2.1 Definition of HPC 6](#_Toc89290788)

[2.2 Definition of OPENMP and MPI 6](#_Toc89290789)

[2.3 HPC on Cloud and comparing to on premise HPC 7](#_Toc89290790)

[3. Cost Analysis 8](#_Toc89290791)

[3.1 My Recommendations to our consultant company 9](#_Toc89290792)

[3.2 References 11](#_Toc89290793)

[3.3 Bibliography 12](#_Toc89290794)

Table of figures

[Figure 1:Standart\_HB60RS&Standart\_HC444RS performance comparison table (Burness, 2018) 8](#_Toc89287182)

[Figure 2: Purdue university on sites HPC costs for 6 different campuses (Smith et al., 2019) 9](#_Toc89287183)

[Figure 3: Cloud costs vs on premise costs & ROI values (Smith et al., 2019) 9](#_Toc89287184)

# Definition of cloud computing and how it can be useful

In this section, a definition will be made about clod computing and an analysis will be made about how it can be useful.

## Cloud Computing definition

Cloud Computing is one of the most used IT terms today. Cloud computing is the term for internet-based computing services for computers and other devices that may be accessed at any time and share computing resources. As Mell, P. and Grance (2011, p.2) defines it:

“Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” (Mell, p. and Grance, 2011, p.2)

Melp, P. and Grace's definition of cloud computing is very comprehensive and simple as a very explanatory definition in the field of cloud computing. It is claimed by computer theorists that the future of the internet will be cloud computing. Accordingly, it is predicted that in the future, online clouds will be used instead of computer hard disks.

## 1.2 Deployment Models

In general, Private, public, hybrid, and community clouds are the four types of cloud. The private cloud deployment model is a sort of cloud technology that is favoured by larger businesses and businesses with more sensitive data. The private cloud is typically hosted in the customer's data centre and on-premises. Also, the private cloud can also be hosted on an independent external data centre. Service providers in Public Cloud applications give users with storage and other resource access. These services are either free or have a per-use fee. The public cloud is a multi-tenant environment, and all public cloud clients share the cloud provider's data centre infrastructure. Hybrid Cloud is a mix of public and private cloud services. Where security and privacy are more important and precautions must be taken, private cloud technology is utilised, but it is more appropriate to use public cloud in areas mentioned where the stated security measures can be reduced. The combination of these also varies according to the size of the companies. If any service we receive on the cloud is shared with several companies, it is called Community Cloud. The last cloud deployment model will be mentioned is the community cloud, which is not a widely used cloud type in the IT market.

## 1.3 Service Models

Another important point to be mentioned about cloud computing is service models. There are three main types of cloud computing. Each model represents a distinct component of the cloud computing infrastructure. The type of service will be mentioned about first is Infrastructure as a Service (IaaS), In this most basic cloud service model, cloud providers deliver servers as physical or virtual machines. Infrastructure as a Service (IaaS) is a type of cloud provides the highest level of flexibility and administrative control over your IT resources, as well as resources similar that are used by many IT departments and software developers today. It is considerably easier to construct testing and development environments. New applications can be brought to market considerably more quickly this way. IaaS storage, network and operating systems are leased, and you only pay for the resources you use. Second, the service model I want to mention about is Platform as a Service (PaaS), Customers can use Platform Services (PaaS) to get the developer tools they need to build and manage mobile and web apps without having to invest in or maintain current infrastructure. It eliminates the need for infrastructure (usually hardware and operating systems) management, allowing customers to focus on application deployment and management. This helps users work more efficiently so users don't have to worry about sourcing, capacity planning, software maintenance, patches, or other similar demanding tasks to run their application. Institutions that create, test, and implement cloud solutions for specific systems before making them available to third parties prefer it. As a conclusion, in addition to hosting, the PaaS service provider provides an isolated installation, test environment, operating system, maintenance, and update supports to the application developer. Another service model I want to mention about is SaaS. Gibson, J. et al. (2012, p. 202) defines the SaaS service model as:

“Software-as-a-Service gives subscribed or pay-per-use users access to software or services that reside in the cloud and not on the user's device. The consumer of a SaaS application only requires thin client software such as a web browser to access the cloud-hosted application. This reduces the hardware requirements for end-users and allows for centralized control, deployment, and maintenance of the software. Some examples of popular SaaS applications are Hotmail, Gmail, and Google Apps.” (Gibson, J. et al., 2012, p.202)

In addition to this definition, the pay-as-you-use method is valid in the SaaS service. For this reason, it helps to customers to save their money. Furthermore, users do not need to install, manage, or upgrade software. All these processes are handled by SaaS providers. In addition, as mentioned in the definition, users do not need to install the applications on their local devices, as the applications are located on a cloud network that can be accessed via the web or an API.

## 1.4 How cloud computing can be beneficial?

Its benefits were also mentioned in the definition section, but I would like to summarize it in general in this section. Cloud computing eliminates investment costs such as software costs, hardware costs, and building in-house data centre costs. The cloud technology provides opportunity to access the information technology resource when needed and in the desired geography. When cloud technology is used, infrastructure problems such as tracking, storing and user authorization are eliminated. Instead of spending a high-budget security and storage software, users pay as much as they use cloud computing. To summarize in general terms, flexibility, security, and efficiency stand out as the most important features offered by cloud technologies to users.

# 2. Comparing HPC Cloud and traditional HPC

In this section, an analysis will be made by comparing the advantages and disadvantages of Cloud HPC and traditional HPC.

## 2.1 Definition of HPC

HPC (High Performance Computing) is a technology that allows us to solve major problems in science, engineering, and business by combining the computing power of multiple servers to provide very high performance in general. It is a method of processing large volumes of data at very high speeds using multiple computers and storage devices as a unified structure. Nielsen, F. (2016, p. 3) defines the HPC as follows:

“High Performance Computing, or HPC for short, is an area encompassing among others the various paradigms of parallel programming, their related programming languages, and application programming interfaces (APIs), the dedicated software tools, the international specialized conferences (ACM/IEEE Super-Computing, or SC for short), etc. Loosely speaking, HPC is both the scientific and technical fields of study of “Super-Computers” (SCs).” (Nielsen, F., 2016, p.3)

A process that takes months in a normal system can be reduced to days, hours, or even minutes through HPC. The main uses of HPC can be enumerated as analysis, data mining, simulation, modelling, software development, visualisation of complex data, and complex mathematical analysis.

## 2.2 Definition of OPENMP and MPI

OpenMP is an application development interface developed for shared memory, multiprocessor architectures, and parallel programming with the help of compiler directives. The parts of the code to be run in parallel are distributed to the thread structures. With the ompgetthreadnum() function, each thread does its own work. After each thread finishes its assigned task, it enters the queue again to receive a new thread. This cycle continues until all the work is done. OpenMP is important for both task and data parallelism. For the definition of MPI, Nielsen, F. (2016, p. 22) gave a very detailed and simple definition as follows:

“The Message Passing Interface (or MPI for short) standard is a programming interface: namely, an Application Programming Interface (API) that defines properly the syntax and full semantic of a software library that provides standardized basic routines to build complex programs thereof. Thus the MPI interface allows one to code parallel programs exchanging data by sending and receiving messages encapsulating those data.” (Nielsen, F., 2016, p.22)

Its basic operation is in the form of sending and receiving messages between processors with discrete address spaces. An environment is created with the MPI\_COMM\_WORLD command to enable communication between processors. MPI\_Init,MPI\_Finalize, MPI\_Comm\_Size, MPI\_Comm\_Rank, MPI\_Send and MPI\_Receive are the most commonly used MPI commands.

## 2.3 HPC on Cloud and comparing to on premise HPC

High Performance Computing (HPC) uses a large number of CPU or GPU to solve complex mathematical tasks. Today, these features are also available in the cloud environment. Although HPC is new to the cloud environment, it is rapidly evolving. Amazon Web Services, Google Cloud Platform, Microsoft Azure, IBM Spectrum Computing, Penguin Computing on Demand (POD) appear as companies that provide HPC support on "cloud". The main difference from on premises systems is that the resources can be decreased or increased according to the needs. Thus, users get rid of extra costs as they will pay for the resource they use. One of other benefits of HPC on Cloud is not having to pay extra for ISV software. Another important benefit is the absence of installation, maintenance, and management costs. Also, since the information is stored in the cloud environment, it can be easily shared with the authenticated users wherever they are in the world. In addition to all these benefits, HPC on cloud is so beneficial for extremely parallel computations. Some studies show that the HPC on Cloud structure works better than the on premise one in applications with a lot of parallelism. Research about the evaluation of HPC on cloud shows that “Clouds are more cost-effective for low communication-intensive applications such as embarrassingly parallel and tree-structured computations and HPC-optimized clusters are better for the rest.” (Gupta & Milojicic, 2011, p. 25) and another research about scientific and business applications on HPC Cloud showed similar results “a strong support seemed to be present when talking about embarrassingly parallel applications, which showed good performance with current cloud resources.” (Netto et al., 2018). Furthermore, on premise HPC systems are generally too expensive for small and medium scale companies. Therefore, cloud may be the ideal solution to meet their HPC requirements. According to new research (Lynn et al., 2020, p. 135) concludes as follows:

“HPC in the cloud allows organizations, particularly small-to medium enterprises, to access HPC infrastructure on a more cost-effective basis and as a result better compete by overcoming internal constraints to generate new capabilities, enter new markets, adopt new resources, and develop new products. This may enable an acceleration of innovation resulting in a positive impact for industry and society as a whole” (Lynn et al., 2020, p.135)

But on the other hand, the negative result for cloud from (Netto et al., 2018) research is as follows:

“The attempts of moving applications with heavy CPU and memory requirements to the cloud started by verifying the cost benefit of running those applications in the cloud against running them on already owned on-premise clusters. Various researchers used well-known HPC benchmarks and a few applications also common in the area. The goal was to understand not only performance but also monetary costs and how sustainable it would be to decommission their own clusters and move everything to the cloud. The main conclusion was that applications that were compute intensive and with high interprocessor communication could not scale well in the cloud, especially due to the lack of low-latency networks such as InfiniBand.” (Netto et al., 2018)

But some cloud providers like Azure even solved that problem. According to research from Microsoft (Burness, 2018) explains the HB and HC performance:

“Both new H-series VM sizes feature 100 Gb/sec EDR InfiniBand from our technology partners at Mellanox. This is the same HPC interconnect technology featured in the world’s most powerful supercomputer, and now Azure is bringing it to the public cloud for the first time.” (Burness, 2018)

Azure almost completely solves the performance problems of the users with its new services.

The most important disadvantage for some research centres or companies might be cost analysis depends on the application that run on system and payment plan from cloud provider which will be detailed in cost analysis part.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Size** | **CPU cores** | **Memory: GB** | **Memory per CPU Core: GB** | **Local SSD: GiB** | **RDMA Network:** | **Azure Network** |
| Standard\_HB60rs | 60 | 240 | 4 | 700 | 100 Gbps | 40 Gbps |
| Standard\_HC44rs | 44 | 352 | 8 | 700 | 100 Gbps | 40 Gbps |

Figure :Standart\_HB60RS&Standart\_HC444RS performance comparison table (Burness, 2018)

# 3. Cost Analysis

Cloud's benefits and potential and possible disadvantages were mentioned. In this section, a cost analysis will be made by comparing the on-premise structures with other cloud providers. Scientific research for small application with low utilization showed that “Related to financially related decisions. When compared to low-utilization clusters and when running small applications, cloud environments can be preferred over on-premise clusters.” (Netto et al., 2018). In other words, for large-scale applications with high utilization, on-premise HPC is more cost-effective. According to research result from (Smith et al., 2019) in Purdue University. Researchers compared the cost of 6 super computers in 6 different campuses with cloud costs.

Table

Description automatically generated

Figure : Purdue university on sites HPC costs for 6 different campuses (Smith et al., 2019)

Note. Adapted from “Community Clusters or the Cloud. Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning),” Smith, P., Harrell, S. L., Younts, A., & Zhu, X. (2019), p. 2 Published. <https://doi.org/10.1145/3332186.3333155>

The researchers made an annual cost analysis of the supercomputers on campuses and the 3 most important cloud providers. According to same research from (Smith et al., 2019), Annual costs are more expensive in cloud, but it also changes which provider is used. The next chart shows the prices and ROI values of these different cloud providers in detail.

Table

Description automatically generated with low confidence

Figure : Cloud costs vs on premise costs & ROI values (Smith et al., 2019)

Note. Adapted from “Community Clusters or the Cloud. Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning),” Smith, P., Harrell, S. L., Younts, A., & Zhu, X. (2019), p. 4 Published. <https://doi.org/10.1145/3332186.3333155>

These results show that running high performance computing on the cloud seems expensive in terms of cost. In this experiment, it is observed that cloud services are not yet close to premise HPC in terms of cost, especially in long-term use. This result can be easily deduced from the ROI values. Since the cost of the 3 most common cloud providers in the market was analysed in this experiment, these values were used to reach a conclusion.

## 3.1 My Recommendations to our consultant company

Personally, my advice is that although cloud providers provide a lot of benefits to their users and constantly improve their services, it will be quite ineffective in terms of cost to use the services of completely cloud providers for the needs of our company, which has an average of 40,000 CPU hours per month. Since 40,000 CPU hours per month is a very serious amount of usage and our company is also a consulting firm, the option to use the cloud entirely will put our company in trouble financially. Since we are a consultancy firm and we know that there is a high variability in monthly usage amounts, it is a fact that only on-premise HPC will limit us in high usage periods. I deduce from the figures that the most effective way for our company's needs in terms of cost and performance is a hybrid model. In the hybrid model, our company will use the on-premise structure, as it is cost effective in periods of normal use and less than normal use. In periods when usage is at the highest level, excess load will be transferred over the cloud, thus providing a flexible service to customers as well as providing the most affordable service against the variable demands of customers in terms of cost. In addition, when these cloud and on-premise systems are owned, the company will have knowledge about these two technologies. Furthermore, when these cloud and on-premise systems are owned, the company will have knowledge about these two technologies. On a cost basis, I strongly recommend using a solid hybrid model for the reasons clearly stated above.

## 3.2 References

Burness, E. (2018). Introducing the new HB and HC Azure VM sizes for HPC. Microsoft Azure. <https://azure.microsoft.com/en-us/blog/introducing-the-new-hb-and-hc-azure-vm-sizes-for-hpc/>

Gibson, J., Rondeau, R., Eveleigh, D. and Tan, Q., 2012, November. Benefits and challenges of three cloud computing service models. In 2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN) (pp. 198-205). IEEE.

Gupta, A. and Milojicic, D., 2011, October. Evaluation of hpc applications on cloud. In 2011 Sixth Open Cirrus Summit (pp. 22-26). IEEE.

Lynn, T., Fox, G., Gourinovitch, A., & Rosati, P. (2020). Understanding the Determinants and Future Challenges of Cloud Computing Adoption for High Performance Computing. Future Internet, 12(8), 135. <https://doi.org/10.3390/fi12080135>

Mell, P. and Grance, T. (2011), The NIST Definition of Cloud Computing, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.SP.800-145 (Accessed November 29, 2021)

Netto, M. A. S., Calheiros, R. N., Rodrigues, E. R., Cunha, R. L. F., & Buyya, R. (2018). HPC Cloud for Scientific and Business Applications. ACM Computing Surveys, 51(1), 22–23. <https://doi.org/10.1145/3150224>

Nielsen, F. (2016). In Introduction to HPC with MPI for data science (pp. 3–3). essay, Springer International Publishing.

Smith, P., Harrell, S. L., Younts, A., & Zhu, X. (2019). Community Clusters or the Cloud. Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning). Published. <https://doi.org/10.1145/3332186.3333155>

## 3.3 Bibliography

Aljamal, R., El-Mousa, A. and Jubair, F., 2020, April. Benchmarking Microsoft Azure Virtual Machines for the use of HPC applications. In 2020 11th International Conference on Information and Communication Systems (ICICS) (pp. 382-387). IEEE.

Ferretti, M. and Santangelo, L., 2020. Cloud vs On-Premise HPC: a model for comprehensive cost assessment. Parallel Computing: Technology Trends, 36, p.69.

Hassan, H.A., Mohamed, S.A. and Sheta, W.M., 2016. Scalability and communication performance of HPC on Azure Cloud. Egyptian Informatics Journal, 17(2), pp.175-182.

Luis M. Vaquero, Luis Rodero-Merino, Juan Caceres, and Maik Lindner. 2009. A break in the clouds: towards a cloud definition. SIGCOMM Comput. Commun. Rev. 39, 1 (January 2009), 50–55. DOI:https://doi.org/10.1145/1496091.1496100

Posey, B., 2016. Dynamic HPC clusters within amazon web services (aws) (Doctoral dissertation, Clemson University).