COMP 1680
Clouds, Grids and Virtualisation.
Cloud Computing Coursework

Student ID - 001002629

27th November 2020

Contents

1 Introduction

This document looks at the definition of what the cloud is, the benefits and drawbacks of different platforms for High Performance Computing (HPC). By using the information that is provided in the analysis, a recommendation is provided for the business to consider on a solution on how to implement a HPC for parallel computation.

2 Cloud Computing

The Cloud can be described as a network of servers which are distributed across the globe, unified through the internet to behave as a singular ecosystem. These combined resources are delivered to customers through the medium of the internet by a Cloud Service Provider as a service, hardware, applications and tools are part of these resources, such as databases, file storage, servers and software. The highly modular nature of cloud computing is it greatest strength but can also make it confusing to navigate, for this reason describing the parts that make the cloud is the only way to define what it really is.

2.1 Advantages of the Cloud

Cloud computing brings many different advantages with it, by expanding and acknowledging these modules we can understand what cloud computing is to a greater degree. With this better understanding of what cloud computing is, we can make informed decisions based on facts that best suit the business needs.

2.1.1 Speed and Agility

The ascendancy of Cloud Computing flows from the speed and agility of allocating new resources, *"only a click away, which means that you reduce the time to make those resources available to your developers from weeks to just minutes."* (Mathew 2014) This gives a company that decides to utilize cloud computing an extreme uplift in its agility as monetary and time costs are considerably reduced while testing and developing, granting an edge over competitors that have not yet adopted cloud computing. Virtual deployment of servers, containers and virtual machines can take seconds to setup, run and scale these resource to fit the needs of the business and then shut a service down as rapidly as they are started.

2.1.2 Reliability and Resiliency

The capacity for a system to return to it initial state after a failure is seen as its resiliency, consistency of performance is the measure for reliability of a system. Without a resilient system, a failure could take days, weeks or even months to recover from. Data that is lost for a prolonged period of time will

cause catastrophic damage financially with lost revenues and to the perceived image of the company.

Cloud computing provides resiliency with the ability to backup and restore applications should there be a disruption to business functions. The distribution of the servers across continents provides a layer of resiliency so that one natural or physical disaster in a location will not disrupt entire systems. Fig 1 illustrates the global infrastructure that Amazon Web Service (AWS) currently operates and planned expansion adding more capacity to the service.

Figure 1: AWS Global Infrastructure (AWS 2018)

High availability provides access to services while *"maintaining acceptable continuous performance despite temporary failures in services, hardware, or data centers, or fluctuations in load."* (Shinder 2019) As the cloud has multiple locations, this shield them from fault tolerance, like hardware failure or fluctuations in traffic flow, the reliability aspect works in favour for cloud computing.

2.1.3 Scalability and Elasticity

Scalability has two flavours, vertical and horizontal scaling. Vertical is also known as scaling up, where the cloud server has resources added to it, for instance, Central Processing Unit's (CPU), more Random Access Memory (RAM) or Hard Disk Drive's (HDD). Scaling down is the opposite and removing these resources. This type of resource allocation is usually automated, removing the need to have dedicated staff and hardware during strenuous periods. Horizontal scaling is where providers facilities more resources by adding or removing servers to meet demands of the customer and prevents a server being over loaded, in the cloud environment this is usually achieved using virtualisation and automation.

Elasticity is having capacity to meet short term fluctuations of high or low demand on resources such as CPU and RAM usage which is automated to fit

the needs at that current time. Allowing services to operate in the manner in which it was designed, even with a high traffic flow time which would overload a standard server and prevent services from being accessed.

2.1.4 Security and Compliance

There are huge quantities of data being stored on cloud servers, providers build in security with a ground up approach. Fig 2 shows this layered approach to security that protects data and software that is stored. Google incorporates restricted physical access to their server rooms for personal who are cleared and authorised only, to Denial of Service attacks and accessing the platform by authenticated users only. All these layers of security are rolled out across the platform for all users, reducing the operating cost for users, who also do not have to invest themselves to get to this level of security incorporated into their systems.

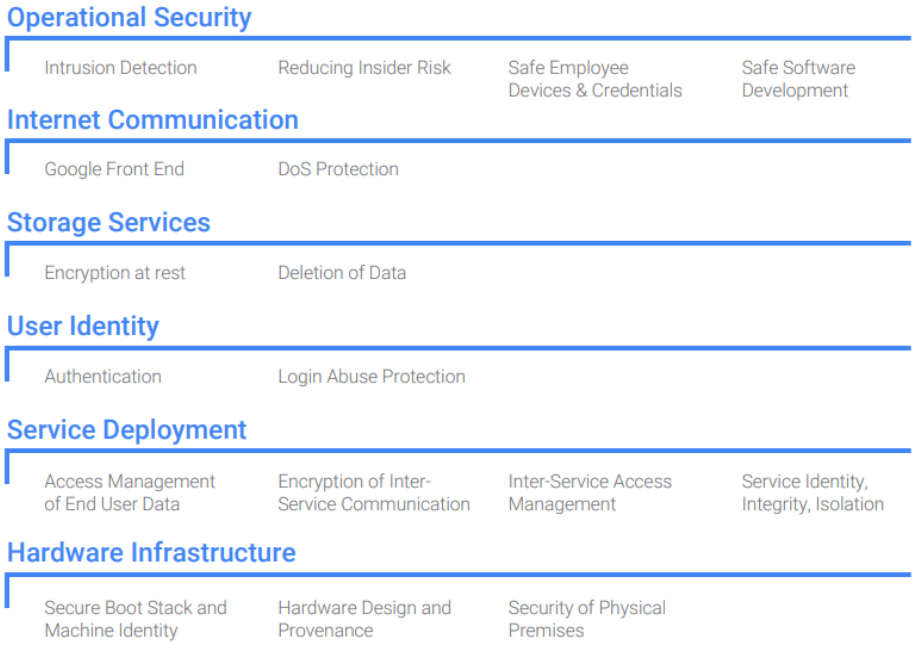

Figure 2: "Google Infrastructure Security Layers." (Google 2017)

Compliance for storing data differs by region, for instance, the protection of data and the processing of data is defined in chapter 2 of the (Data Protection Act 2018 2018) which incorporates (European Parliament and Council of European Union (2016) Regulation (EU) 2016/679 2016), more commonly known as General Data Protection Regulation (GDPR) that covers regulation on data processing and privacy in the European Union (EU) and European Economic Area (EEA). These regulations differ wildly across regions, therefore navigating

these complex protections can come at a high cost for businesses, Google however does offer *"a dedicated team where data protection related enquiries can be directed."* (Google 2018) Reducing the cost of operations.

2.1.5 Economies of Scale and Cost

The economies of scale *"refer to reductions in unit cost as the size of a facility increases"* (Wang et al. 2012) With this in mind, large cloud infrastructures should cost less to purchase for providers. This then benefits customers as savings that are made from purchasing large quantities of hardware can be passed on. With more users using cloud services, utilisation of hardware increases also reducing the running cost for the provider. For economies of scale, more is better as the costs are spread over a larger area. As cloud providers build more and larger facilities, this in turn reduces the cost of the hardware, especially when they are bought in large quantities for use in multiple data centers over a single company who are considering building a HPC.

2.2 Cloud Computing Models

There are three major types of service model to choose from, Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). These different models can directly translate to how much control users have and what the cloud service provider maintains. There are also four deployment models in Public Cloud, Private Cloud, Community Cloud and Hybrid Clouds, referring to the location and management of a cloud infrastructure.

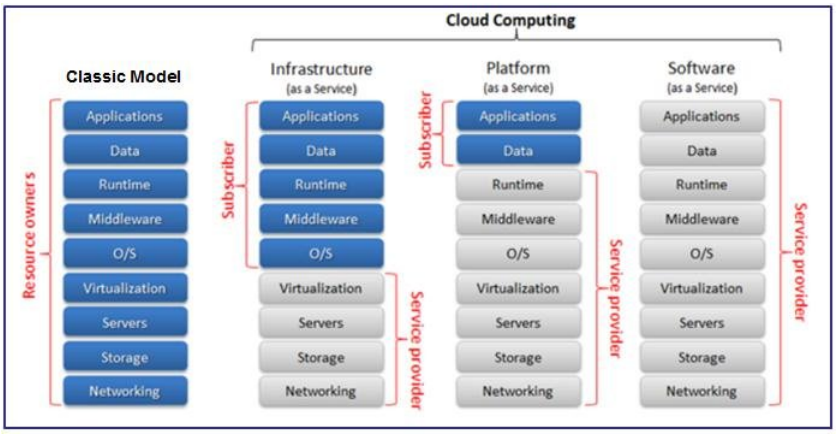

Figure 3: *"Separation of Responsibilities in Cloud Models"* (Kartit et al. 2014)

2.2.1 Infrastructure as a Service (IaaS)

IaaS provides users low level of control over resources such as network, compute and storage. *"IaaS enables end users to scale and shrink resources on*

an as-needed basis, reducing the need for high, up-front capital expenditures or unnecessary "owned" infrastructure" (IBM-Cloud-Education 2019) It forms the basis of the cloud combining physical hardware and virtualisation that users have access too. This has advantages over physically owning the hardware as there are no large investment costs and lengthy time constraints involved in acquiring hardware and configuring clusters before any computations are performed, it is also elastic therefore hitting resource limits is very unlikely in the cloud as more resources can be assigned for a period of time to cover the needed resources before being reducing again to levels that are required which saves on costs.

2.2.2 Platform as a Service (PaaS)

Platform as a Service (PaaS) includes all the infrastructure that IaaS includes, but also includes the Operating System (OS) that the servers are running and the development tools and environment needs to build applications. The user of this service only controls the data and the application that are run on this platform, which can be advantageous as it creates a single environment where the developer can build, debug, test and deploy, without the added costs of OS licences and the hardware that it runs on. All of this can be achieved using a compatible device that is connected to the internet through a web browser making it highly accessible and convenient.

2.2.3 Software as a Service (SaaS)

Software as a Service (SaaS) is a mechanism of software deployment from the operator of said software. The application is hosted in the cloud where the user does not need to worry about the hardware, OS or application version. The software is routinely sold as a subscription to the user, allowing flexibility for the end user to purchase a set period of access time to the applications. Some examples of SaaS would be Adobes Creative Cloud which offers a range of applications like Photoshop to storage and the ability to share the files, or Microsoft Office 365 which provides productivity applications, storage and communication.

2.2.4 Deployment Models

As stated, the deployment models are split into four main categories, Public Cloud, Private Cloud, Community Cloud and Hybrid Clouds and refer to the location of the server and who has access to the hardware.

- **Public Cloud** Public clouds are provided by third party companies like AWS, Google or Azure. They provide resources that are shared among the users publicly over the internet but keep each users data separate and confidential.
- **Private Cloud** These types of cloud are built with end users as the sole user of resources that are available. Usually, hardware is built, managed

and maintained by the user and not any third party. Although Red Hat does offer a different view stating, *"organizations are now building private clouds on rented, vendor-owned data centers located off-premise."*(RedHat 2019) These clouds can be highly specialised and focused for the business that decides to implement a private cloud.

- **Community Cloud** Community clouds are an amalgamation of business or organisations that have common or shared goals, the data on these clouds are available for everyone, therefore these types of clouds should not have any sensitive data stored on them. The National Institute of Standards and Technology (NIST) also states, *"It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises."*(Mell and Grance 2011) Therefore extra care must be taken when using these clouds

- **Hybrid Cloud** As the name suggests, hybrid clouds combine the use of one or more of the differing types of cloud. Implementing compute, storage and other infrastructure that is on both a private cloud and a public cloud, is the definition of hybrid. It gives the organisations that use this type of cloud great flexibility and agility that give it an advantage over their competitors.

3 Analysis of Platforms

There are many metrics that in-house HPC and Cloud Computing can be analysed on, understanding the benefits and drawbacks that each platform brings will lead to a better informed recommendation that can be used moving forward.

3.1 Traditional HPC Clusters

Traditional in-house implementations of HPC brings with it the full control over the clusters and how they are implemented. From operating systems to the type of CPU that is installed, every aspect of a given cluster is governed by the owner allowing the full optimisation of the hardware and software stack for the task that needs to be computed.

By utilising an on premises solution, expanding or upgrading cluster storage, networking capabilities and processing power requires large upfront cost from the business and time for the hardware to be installed into the cluster. This solution also requires maintenance which the business will have to finance to keep it operational and have specialist employees for this task.

Even with this increased cost of operation, there are advantages to keeping HPC on premises, with full control over the data that is fed into the HPC and the results that are generated there is little risk to data being while in transit through the network when compared to sending data over the internet and risking man in the middle attacks.

If there is highly sensitive or personal information that is being processed on the cluster, then an in house solution provides a layer of protection that no other solution can provide. An extreme version would be providing an air gap so that there are no connections to the outside world. Therefore, only way of moving data to or from the cluster would be a physical device. So although not the most convenient it would be a secure solution for highly sensitive data.

3.2 Cloud HPC

With the ability to instantly create a virtual cluster and use it almost instantaneously, no downtime for hardware upgrades, cloud computing has many advantages. Having no maintenance costs to the business or specialised network staff the cost overhead of using a cloud vendor can be lower upfront. With the increasing resource pool that these vendors have, the raw compute power is catching up to traditional HPC as seen in Fig 4 where Guidi et al. (2020) have performed benchmarks on the performance gap between two traditional HPC machines and two AWS Elastic Compute Clusters(EC2).

Platform	Age	Core/Node	Frequency (GHz)	Theoretical Peak (GFlops/Node)	Processor	Memory (GiB)	Network (Gbps)	L1	L2	L3
Cori Haswell	4	32	2.3	1,177	Xeon E5-2698V3	120	*82	64KB	256KB	40MB
Cori KNL	4	68	1.4	3,046	Xeon Phi 7250	90	*82	64KB	1MB	-
AWS r5dn.16xlarge	1	32	2.5	2,560	Xeon Platinum 8259CL	512	75	64KB	1MB	36MB
AWS c5.18xlarge	1	36	3.0	3,456	Xeon Platinum 8124M†	144	25	64KB	1MB	25MB

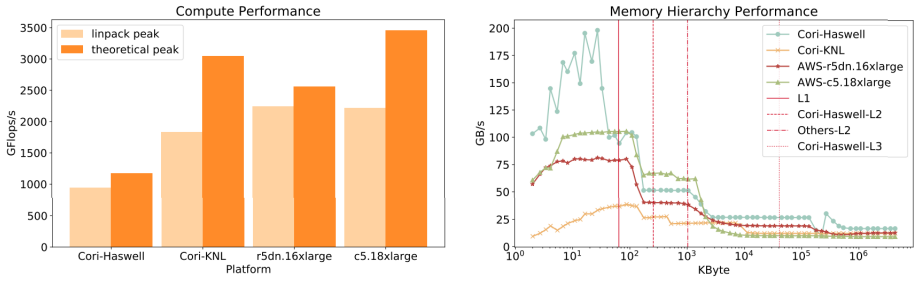

Figure 4: *"Details of the evaluated machines, Peak compute performance, CacheBench"*(Guidi et al. 2020)

The findings of this research from Guidi et al. (ibid.) state, *"Surprisingly, cloud systems offered higher bandwidth and lower latency than HPC systems in the point-to-point communication microbenchmarks."* Meaning that the performance gap that had existed between the two platforms has closed in recent years, making cloud computing a seriously viable option for pure performance.

However, there are vulnerabilities in the domain of security, *"For instance, a rapidly growing literature indicates that virtual machine vulnerabilities to side-channel attacks expose IaaS (Infrastructure as a Service), PaaS (Platform as a Service) and SaaS (Software as a Service) to breaches"* (Mthunzi et al. 2020). The petabytes of data that service providers contain, manage and process makes them a prime target for hackers. Evidence to this is a survey conducted by

Help Net Security of 300 Chief Information Security Officers (CISOs) found that *"Nearly 80% of companies had experienced at least one cloud data breach in the past 18 months"*(Help Net Security 2020) With a 80% of responding CISOs from the same survey stating configuration of access setting and the inability to to pinpoint superfluous access to sensitive data as a major issue.

3.3 Hybrid Cloud

By utilising the best of both traditional and cloud HPC, the cost to setup and maintain would be considerably lower then a full in-house cluster. Using a cluster in-house to process critical or sensitive data and sending less critical and sensitive to a provider that could be instanced for a very short time reducing the overall cost. The cost saving of the cloud could be used along with the data security of a physical solution, the flexibility to acquire more compute power instantly in the cloud while keeping the full control over the physical cluster is an attractive prospect.

4 Recommendation

Having a HPC be it traditional or in the cloud can be extremely beneficial to any business, allowing research and development that just is not possible on standard hardware, affording an edge over the competition in the same field. Innovation in research and development can be rapidly advanced over those who do not utilise this advantageous platform.

Although on the surface it would look like a cloud implementation would seem like a perfect solution with little upfront cost and no need for specialised employees, serious consideration needs to be spent on security and integrity aspects of data being transferred and processed on a cloud platform. With an uptick in data breaches involving users of cloud providers, and the nature of sending data over the internet being vulnerable while in transit to man-in-the-middle attacks. If the data that is being processed is highly sensitive then a cloud only solution is not viable.

As for a in-house solution, the initial cost and time to install can prove prohibitively expensive if not planned properly including long term maintenance which needs taken into consideration. This does offer the greatest level of control over hardware and software, especially if there is a custom OS that needs to be used to run the applications. Data can be more secure as it only transits between the clusters and not over the internet, so processing highly sensitive data should not be an issues with good access management.

In conclusion the best solution moving forward would be to use a hybrid HPC cloud solution. This give the business the highest control over the hardware that is owned and operated to process the critical data but also reduces the cost of ownership as the on-premises cluster can be made smaller. With less critical data being passed over to a cloud HPC cluster to be processed.

Reference

AWS (2018). *Global Infrastructure*. Accessed on 15-10-2020. URL: https://aws.amazon.com/about-aws/global-infrastructure/.

Azam, Md Gulnawaz (Jan. 2019). "Application of cloud computing in library management: innovation, opportunities and challenges". In: Accessed on 16-10-2020. DOI: 10.5281/zenodo.2536637.

Data Protection Act 2018 (2018). *Data Protection Act*. Accessed on 16-10-2020. URL: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted.

Emeras, J., S. Varrette, and P. Bouvry (2016). "Amazon Elastic Compute Cloud (EC2) vs. In-House HPC Platform: A Cost Analysis". In: *2016 IEEE 9th International Conference on Cloud Computing (CLOUD)*. Accessed on 27-10-2020, pp. 284–293. DOI: 10.1109/CLOUD.2016.0046.

European Parliament and Council of European Union (2016) Regulation (EU) 2016/679 (2016). *Council regulation (EU) no 2016/679*. Accessed on 16-10-2020. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/%20?uri=CELEX:32016R0679%5C&from=EN.

Google (2017). *Google Infrastructure Security Design Overview*. Tech. rep. Pages: 4 - 13 Accessed on 15-10-2020. Google Cloud. URL: https://cloud.google.com/security/infrastructure/design.

— (2018). *General Data Protection Regulation (GDPR)*. Tech. rep. Pages: 3 - 13 Accessed on 16-10-2020. Google Cloud. URL: https://cloud.google.com/security/gdpr/resource-center/pdf/googlecloud%5C_gdpr%5C_whitepaper%5C_618.pdf.

Guidi, Giulia et al. (2020). *10 Years Later: Cloud Computing is Closing the Performance Gap*. Accessed on 08-11-2020. arXiv: 2011.00656 [cs.DC].

Help Net Security (May 2020). *Most companies suffered a cloud data breach in the past 18 months*. Accessed on 08-11-2020. URL: https://www.helpnetsecurity.com/2020/06/03/cloud-data-breach/.

IBM-Cloud-Education (2019). *IaaS (Infrastructure-as-a-Service)*. Accessed on 17-10-2020. URL: https://www.ibm.com/cloud/learn/iaas.

Kartit, Zaid et al. (May 2014). "Network Issues in cloud computing and countermeasures". In: *JNS4 Tetouan*. Accessed on 17-10-2020.

Lawton, G (2008). "Developing Software Online With Platform-as-a-Service Technology". In: *Computer* 41.6. Accessed on 20-10-2020, pp. 13–15.

Mathew, Sajee (2014). *Overview of Amazon Web Services AWS Whitepaper*. Tech. rep. 7. Accessed on 14-10-2020. Amazon Web Services. URL: https://docs.aws.amazon.com/whitepapers/latest/aws-overview/aws-overview.pdf.

Mell, Peter M. and Timothy Grance (2011). *SP 800-145. The NIST Definition of Cloud Computing*. Tech. rep. Accessed on 21-10-2020. Gaithersburg, MD, USA.

Mthunzi, Siyakha N et al. (2020). "Cloud computing security taxonomy: From an atomistic to a holistic view". In: *Future Generation Computer Systems* 107. Accessed on 08-11-2020, pp. 620–644.

RedHat (Nov. 2019). *What is private cloud?* Accessed on 21-10-2020. URL: https://www.redhat.com/en/topics/cloud-computing/what-is-private-cloud.

Shinder, Debra (2019). *Trusted Cloud: Microsoft Azure Security, Privacy, Compliance, Reliability/Resiliency, And Intellectual Property*. Accessed on 14-10-2020. URL: https://go.microsoft.com/fwlink/?LinkId=392408%5C&clcid=0x809.

Wang, L. et al. (2012). "In Cloud, Can Scientific Communities Benefit from the Economies of Scale?" In: *IEEE Transactions on Parallel and Distributed Systems* 23.2. Accessed on 16-10-2020, pp. 296–303.

Bibliography

Bhardwaj, Sushil, Leena Jain, and Sandeep Jain (2010). "Cloud computing: A study of infrastructure as a service (IAAS)". In: *International Journal of engineering and information Technology* 2.1. Accessed on 20-10-2020, pp. 60–63. ISSN: 0471491101.

Corporation, Intel (Jan. 2020). *An Overview of Cloud Deployment Models*. Accessed on 21-10-2020. URL: https://www.intel.com/content/www/us/en/cloud-computing/deployment-models.html.

Doelitzscher, Frank et al. (Jan. 2011). "Private cloud for collaboration and e-Learning services: from IaaS to SaaS". In: *Computing* 91.1. Accessed on 20-10-2020, pp. 23–42. ISSN: 1436-5057. DOI: 10.1007/s00607-010-0106-z. URL: https://doi.org/10.1007/s00607-010-0106-z.

Gigler, Björn-Sören, Alberto Casorati, and Arnold Verbeek (2018). *Financing the Future of Supercomputing: How to Increase Investment in High Performance Computing in Europe*. Accessed on 01-11-2020. European Investment Bank.

Google (Oct. 2020). *Machine types | Compute Engine Documentation | Google Cloud*. Accessed on 21-10-2020. URL: https://cloud.google.com/compute/docs/machine-types.

Kumar, Adesh (Jan. 2020). "Research Issues in Virtualization in Cloud Computing". In: *International Journal of New Innovations in Engineering and Technology* 12. Accessed on 20-10-2020. ISSN: 2319-6319. URL: http://www.ijniet.org/wp-content/uploads/2020/10/24.pdf.

Rouse, Margaret, Kate Brush, and Stephen Bigelow (July 2020). *What is Platform as a Service (PaaS)?* Accessed on 20-10-2020. URL: https://tinyurl.com/y4g9fdye.

Short, Taylor (Sept. 2020). *What is SaaS? 10 FAQs About Software as a Service*. Accessed on 21-10-2020. URL: https://www.softwareadvice.com/resources/saas-10-faqs-software-service/.

Verma, Aman (2020). "Cloud Platform Optimization for HPC". In: *Supercomputing Frontiers*. Ed. by Dhabaleswar K. Panda. Accessed on 01-11-2020. Cham: Springer International Publishing, pp. 55–64. ISBN: 978-3-030-48842-0. URL: https://link.springer.com/chapter/10.1007/978-3-030-48842-0_4.

List of Figures