IT technologies: Clouds, services, and servers

What does it do?

Cloud computing is the term used to describe the enabling of various computing resources to be conveniently utilised over the internet. These computing resources involve the use of services, servers, storage, networks and applications that can be rapidly accessed and allow the user to scale the amount of computing resources to their current and future needs.

As shown in figure 1, cloud computing utilises virtualisation which is the process of allowing multiple users to share a single computer (the server) – which fundamentally consists of a processor, memory, storage and network capacity -- at the same time without affecting each other; this contrasts to a single person using a single piece of hardware such as a laptop (Sclater, 2020). Virtualisation is possible through software, called a hypervisor, which is loaded onto the server and allows multiple instances of an operating system to be run at the same time (Arabnia, Deligiannidis & Tinetti 2019, pg. 3). The importance of sharing a single piece of hardware is that often the amount of computing resources, such as processing power and storage, required are less than what the user needs which results in inefficient use of hardware and power consumption (Ruparelia, NB 2016, pg. 1).

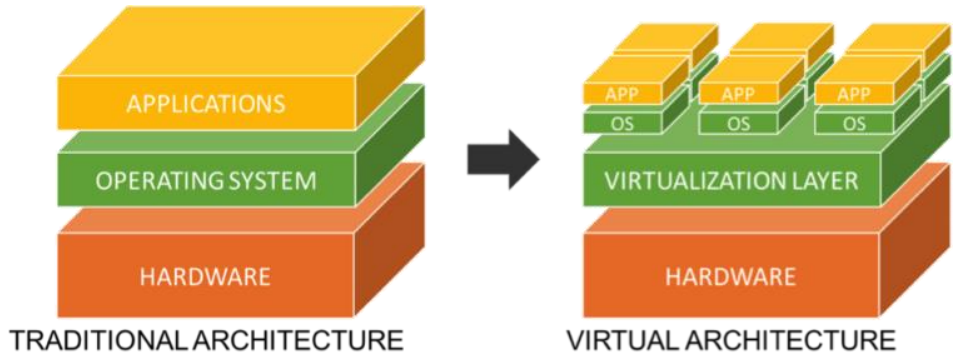

Figure 1: *Traditional Architecture Virtual Architecture* by Emily Sclater, 2020.

Although it is possible to setup a private network that utilises virtualisation to reduce power and hardware inefficiencies, there are still many remaining issues such as managing hardware, security, and data. Providers of cloud computing, such as Amazon Web Services (AWS), offer to manage these issues and provide benefits which include applications and rapid access to these computing resources depending on user needs (Amazon Web Services 2021, Benefits of cloud computing). This allows the user to have limited upfront costs of purchasing hardware, hiring staff to manage it, whilst also allowing for flexibility in the amount of resources needed in the future.

Fox & Hao (2017, p. 477-478) describes the three main methods that cloud computing services offer as Infrastructure as a Service (IaaS), Platform as a service (PaaS), and Software as a Service (SaaS). As shown in figure 2, these services are often called a stack because they start from a more basic service (IaaS) and increase to a more complex service (SaaS) (Fu, 2017). IaaS is the most basic and offers hardware infrastructure which provides computing resources such as processing power,

memory, storage and network capability. When choosing PaaS, in addition to what IaaS offers, the user will also be provided with necessary infrastructure such as operating systems and databases. Finally, in addition to what both IaaS and PaaS provide, SaaS will provide the entire package with the addition of software and application hosting. A popular example of SaaS use is Google workspace, Dropbox, and salesforce.

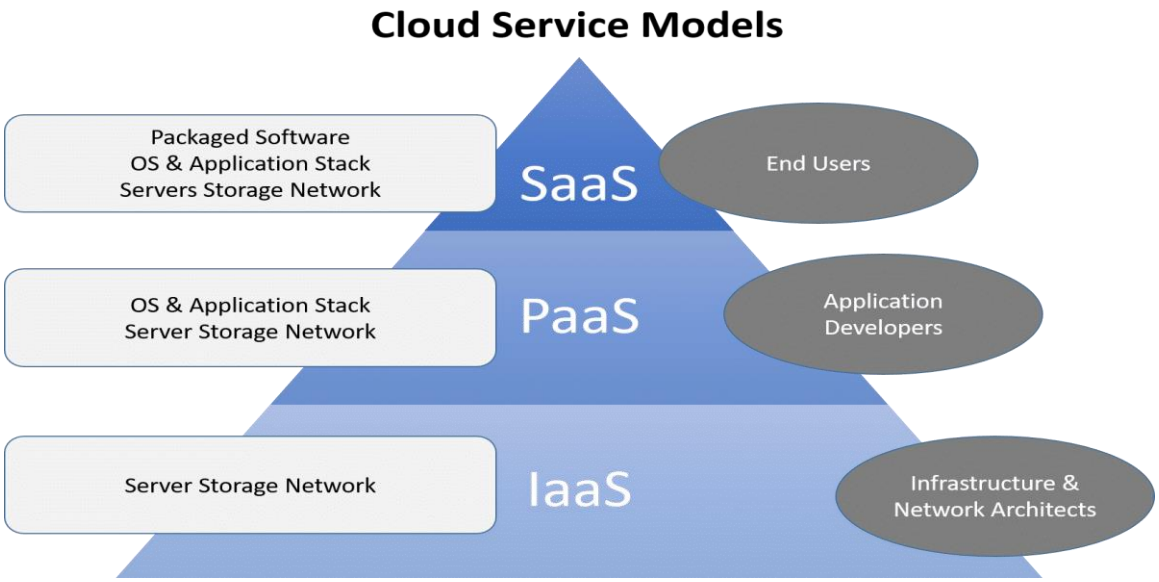

Figure 2: *Cloud Service Models by Arron Fu, 2017.*

There are four deployment models for cloud computing which are public, private, hybrid, and community (Fox & Hao 2017, p. 478). The public cloud provides services to anyone with internet access but have poor security. Private clouds are restricted to one organisation and are not shared. These clouds can be internal and managed by the owning organisation or external and managed by a third party. The difference being greater security for internal, but cheaper fees for external. Hybrid clouds use a combination of private and public clouds which allow the flexibility of using the cheaper public cloud to store nonconfidential information when the private cloud is at capacity. The community cloud involves a range of parties with similar interests that share the same cloud with the purpose of retaining similar levels of security as the private cloud.

What is the state of the art of this new technology? What can be done now?

Current cloud technology allows the user to purchase only the computing resources needed. This scalability is a major feature of the cloud and improvements through machine learning also allow the services to predict and change the computing resources required by the user. An example of this service is Amazon EC2 which recently released an auto scaling warm pools feature that allows applications to start faster and stop processes sooner allows you to increase and decrease the resources within only a few minutes (Amazon Web Services 2021, Amazon EC2 Auto Scaling introduces Warm Pools to accelerate scale out while saving money).

What is likely to be able to do be done soon (say in the next 3 years)?

One of the expanding technologies relating to cloud computing is fog computing which is expected to increase in usage over the coming years. This technology deploys computing resources (edge devices), such as mobile base stations, closer to the data sources to improve response times and network latency (Buyya et al. 2019, p. 3). This responsiveness allows users to operate and offload their high bandwidth applications and computational tasks, such as virtual reality and security, with lower performance issues than connecting directly to the main cloud server.

What technological or other developments make this possible?

With edge devices offering reduces latency and response times, it is desirable to cater to increasing bandwidth usage by users. Users on mobile devices will be able to take advantage of these cloud advancements with the assistance of wireless technologies such as 5G which may leverage a thousand-fold improvement over the previous generation (Ren et al. 2020, p.2).

What is the likely impact?

The use of edge devices for their improvements to responsiveness and bandwidth efficiencies is likely to impact everyone who uses the cloud. This includes the individuals who use their phones for private uses, such as video games, to larger organisations that utilise a great amount of bandwidth for surveillance recordings.

Individuals concerned with performance can benefit greatly from edge devices used by the cloud. Buyya et al. (2019, p. 11-12) describes the utilisation of edge device servers to take the burden of heavy processing whilst providing improved latency and response times. Interacting with edge devices through the fog computing concept, users of devices such as smartphones and health tracking watches can utilise the energy-saving features of the cloud whilst retaining their responsiveness.

Larger organisations often use large amounts of bandwidth, but also require more responsiveness than the main cloud server can provide. (Ren et al. (2020, p.2) describes the bandwidth usage concern for an airport that deals with thousands of surveillance cameras and the accumulating data that results from their recordings. With the use of edge devices such as dedicated smaller servers located near the data source (the airport) it is possible for these edge devices to filter through the video recordings before uploading them on to the main cloud servers. This filtering process allows for the reduction of otherwise large files to be more efficiency transmitted to the main cloud server.

Will this create, replace or make redundant any current jobs or technologies?

With the growth of the cloud and it being used as a substitute for storage and computation of localised devices, it will reduce the pressure of technology manufacturers of devices, such as smartphones and laptops, to expand the capabilities of them. Although advancements in these devices are desirable, especially the power consumption of smartphones, the increase in offloading of these demands onto the cloud may result the redundancy of large local storage capacity or other previously desired features.

How will this affect you?

With the cloud being present in so many parts of daily life it is difficult not to be affected. Daily I use products of the cloud such as online banking, Emails, media streaming services, and gaming. With the continuous expansion of edge computing servers, it will allow me to enjoy these services with less response time issues. This is especially important for online gaming for which response times and latency impact the activity significantly.

As someone who frequently uses devices in power saving modes, the increase in low latency cloud services through edge devices will allow me to use my smartphone with less worries of battery life. With the offloading of computing processes to the cloud I will be able to worry less about my phone's battery draining and being stuck in important situations that require its use.

In addition to less power consumption worries, I no longer need to be as concerned with my smartphone having limited storage capacity. Previously if more storage was required, I would often have to purchase additional storage, but now cloud services often offer free storage.

What will be different for you?

Although as a user I most likely will not notice a significant decrease using cloud and edge computing, it is certainly easy to appreciate reduced annoyance from high response times of older cloud technologies. As I continue down the path of utilising more technology as a software developer, I am sure the daily benefits of cloud computing will become more appealing.

How might this affect members of your family or your friends?

Much of my family and friends have already enjoyed the benefits of cloud computing through its backup capabilities. Recently my family lost a lot of localised data relating to videos and pictures but were fortunate enough to have uploaded copies of this information to the cloud. Without the use of the cloud, they would not have been able to restore the data and nostalgic pictures would be lost forever.

References

Amazon Web Services 2021, *What is cloud computing?* Amazon Web Services, viewed 19 April 2021, <https://aws.amazon.com/what-is-cloud-computing/>.

Amazon Web Services 2021, Amazon EC2 Auto Scaling introduces Warm Pools to accelerate scale out while saving money, Amazon Web Services, viewed 19 April 2021, <https://aws.amazon.com/about-aws/whats-new/2021/04/amazon-ec2-auto-scaling-introduces-warm-pools-accelerate-scale-out-while-saving-money/>.

Arabnia, HR, Deligiannidis, L & Tinetti, FG (eds) 2019, *Grid, Cloud, and Cluster Computing*, C.S.R.E.A, ProQuest Ebook Central.

Buyya, R, Srirama, S, Casale, G, Calheiros, R, Simmhan, Y, Varghese, B, Gelenbe, E, Javadi, B, Vaquero, L, Netto, M, Toosi, A, Rodriguez, M, Llorente, I, Vimercati, S, Samarati, P, Milojicic, D, Varela, C,

Bahsoon, R, Assuncao, MD, Rana, O, Zhou, W, Jin, H, Gentzsch, W ,Zomaya, A & Shen, H 2019, 'A Manifesto for Future Generation Cloud Computing: Research Directions for the Next Decade', *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–38, viewed 19 April 2021, ACM digital library.

Fox, R, & Hao, W 2017, *Internet Infrastructure : Networking, Web Services, and Cloud Computing*, Taylor & Francis Group, ProQuest Ebook Central.

Fu, A 2017, *Cloud services models,* viewed 23 April 2021, https://www.uniprint.net/en/7-types-cloud-computing-structures/>.

Ren, J, Zhang, D, He, S, Zhang, Y., & Li, T. 2020, 'A Survey on End-Edge-Cloud Orchestrated Network Computing Paradigms: Transparent Computing, Mobile Edge Computing, Fog Computing, and Cloudlet', *ACM Computing Surveys*, vol.*52*, no. 6, pp. 1–36, viewed 19 April 2021, ACM digital library.

Ruparelia, NB 2016, *Cloud Computing*, MIT Press, ProQuest Ebook Central.

Sclater, E 2020, *Traditional Architecture Virtual Architecture, viewed on 23 April 2021,* <https://www.linkedin.com/pulse/what-5-types-virtualisation-emily-sclater/?trackingId=5O55ZNkNSnSbgfdmtL32Qw%3D%3D>.