Title- Scalability

Students

Gianni Pessotto- 13199219

Daniel Boustani- 13207509

Yasin Celik- 99112486

Abstract

The popularity of cloud computing has risen dramatically as technology has evolved as the performance of it has successfully improved dramatically. Cloud computing and a reliance on it in the future is becoming inevitable and will be faced with numerous issues. However, many positives exist to support cloud computing. These positives and negatives are the same as they ultimately present a challenge to the future of cloud computing, challenges that must be solved. One important challenge is the scalability of cloud computing. Scalability is the ability of cloud computing to instantly provide necessary resources when requested and to cease providing those same resources when requested or is no longer needed. In this report we will conclusively analyse the existing opinions towards scalability in cloud computing and previous results on the issue.

Keywords

Virtual Machine

Scalability

Cloud Computing

Overview

Scalability of cloud computing, is the primary reason cloud computing is being adopted more by businesses. Without scalability, the usefulness of cloud computing would be diminished. Scalability allows cloud computing to be as efficient as possible by providing numerous resources in many different forms, such as virtual machines, to fulfill requests made by businesses. At request a business can also scale back on the number of virtual resources provided. Scalability allows cloud computer providers to provide enough resources to any large or small scale request in a rapid time frame. The ability to scale-up or scale-down ensures

businesses can save upfront costs by switching from investments in capital expenses, to operational expenses. Operational expenses also reduce in value due to achieving economies of scale. The scalability also works in favour of any business size, as the amount of scalability and resources used is what is paid for. Scalability has the additional benefit of auto-managing, ensuring only necessary resources are used by the business and instances are self-managing. This autonomous managing allows the customer or business to focus on core activities and reduces time cost as well as resource costs.

A common example of companies making use of scalability is creating instances of VM OS accounts to allow thousands of employees to access their accounts on any company pc. This is also true for students at universities and high schools. Since cloud computing is highly scalable, you can have ten accounts or ten thousand. Another benefit of cloud computing scalability is the fact that it is mostly risk free. Traditionally, companies needed to look for suitable servers to buy. Then, get them shipped safely. Finally, you will need to find a suitable place to store the servers. Even after receiving the servers, you will need to manually prepare and maintain them permanently. The traditional way also didn't allow for scaling down. If you needed to scale down, the best option was to sell the server for a fraction of the price. Thanks to cloud computing, this can all be done in minutes through cloud service providers online with many advanced options that allow for scaling. Scalability is also one of the main reasons why cloud computing exists today.

Summaries

The rise of cloud computing has increased the attention its challenges receive, especially scaling of cloud computing resources. As cloud computing is the future, many notable theorists from many different journal publishers have proposed many theories and viewpoints to solve the scalability issue in cloud computing. These approaches are generally designed to specifically counter an aspect of the scalability challenge. For example, Yamoah et al. (2015) proposes a specific architecture to ensure stable support for system engineers. This model would "employ the main benefits" by utilising the Open Services for Life-cycle Collaboration (OSLC) technology. This technology allows tool interoperability, consequently enabling the support of complex engineering steps (Yamoah) as well as the creation of a reactive middleware, which directs information to stakeholders regarding the change of development artefacts (Yamoah).

The supposed artefacts are stored in a repository and this repository acts as a platform to ensure formal model decomposition and work decomposed models follows a certain protocol. Furthermore, the architecture supposedly contains components to guarantee the system engineering process is dependable. This approach from Yamoah et al. (2015) is designed to provide cloud computing an architecture structure that is flexible and friendly to cloud creators to ease program complexity.

Cloud computing scalability has issues that will only increase over time. An example of this is the increased adoption of cloud computing may result in very congested network traffic, which would slow down overall system performance dramatically. This is due to two main reasons, as experiments performed by Tan et al. (2012) highlighted that host disk input and output as well as host capacity are the two most prominent factors when solving this issue of high level traffic. Host input and output is limited as many gigabytes make up a virtual machine and when many VMs are being provided by a single host, the input/output becomes a bottleneck. The cloud system and the amount of VMs that can be provided is the amount of capacity the host has. With this constraint in mind, new VM requests will be delayed or dropped entirely if host capacity or input/output has insufficient room or is congested. Tan et al. (2012) and their newly proposed model accompanied by detailed analysis may provide the appropriate solution to this issue. Tan highlights the use of a queue system to contend with heavy use of cloud computing. Through analysis using a newly proposed ‘Jackson model’, Tan et al. (2012) was able to simulate a large cloud network with heavy traffic and determine that simply adding more host servers close to critical points would improve the systems performance. The purpose of the improvement from Tan et al. (2012) is to ensure cloud applications are easier to access and are more efficient in their usage of space/capacity, especially since cloud computing will become more relevant as we continue.

Cloud computing upon creation and delivery, consists of environments and applications to deliver a final VM to the customer who requested the resources. As mentioned previously, scalability has issues such as size and capacity and must act according to these constraints. Kachelle and Hauck (2013) has pioneered a new system to aid cloud computing scalability by utilising a new method of scalability. This new method is based on the component structure as it

can “be exploited”. The proposed PaaS (platform as a service) divides applications into components. This system automatically manages the resources and deployment of the component based applications to ensure the packages are seamlessly delivered and running without exceeding capacity or size. It automates scaling in and out as well as the distribution revolving around the various cluster nodes. The newly designed system is quite agile and allows for the possibility in real-time, to make changes to applications to adapt to differing loads, infrastructures and objective functions. To ensure this scaling and new method correctly performs, two services are used to ensure smooth transition and re-allocation of resources. Kachelle and Hauck (2013) system is very agile, fine and highly automated. Similarly to other mentioned viewpoints, with cloud computing increasing, solving scalability remains a high priority.

Furthermore, in relation to the topic of scalability and how scalability correlates with cloud computing. Cloud computing as spoken above has 4 main centres of attention which include: delivery, secure, fault-tolerance and scalable infrastructures for hosting internet based application services. What’s cloud computing citizens might ask, well quickly it's a technology which produces and consumes high amounts of data every single day. What does this mean well this means that the cloud essentially stores tons of these applications. The replication of several huge amounts of data in the cloud made it much more efficient and simpler which has numerous success in performance in terms of load balancing, response and availability (HShahapure and Jayarekha, 2015). In relation to scalability, Data must be effectively increased so that it becomes easily accessible which will assist the load balance in the system. Subsequently, the use of scalability and availability will gradually improve the performance of the cloud (HShahapure and Jayarekha, 2015). As mentioned above, reliability, fault-tolerance, scalability, and security are all essential parts that are linked directly to each other which will ensure that the system is being used correctly and successfully (HShahapure and Jayarekha, 2015).

Cloud computing is heavily reliant on the progression of technology. As mentioned before, IO operations have always been a problem when it came to heavy loads on cloud computing when it was first introduced due to HDDs struggling with read and write operations. Since, this issue is based on the actual hardware the only answer to increase throughput was to increase the number of HDDs in a VM. Another hardware that can also store information was a device called Solid

state drive (SSD). SSDs have a much greater I/O and throughput compared to HDDs but are always too low on storage. With the improvements to storage technology SSDs can now store up to 10tb which allows SSDs to replace HDDs when it comes to high input and output systems (Micheloni, Marelli and Eshghi, 2012).

Findings/reflections

The above solutions all have the potential to solve the challenge of scalability. All three solutions consist of rearranging architecture, creating more in-depth tools for developers and redesigning how VM and cloud resources can be delivered without causing issues. These solutions all aim to target excess cloud resources and how cloud resources are delivered to the customer. The solutions have potential to solve the issue completely however, without proper implementation, cloud computing will have the issue of scalability. By implementing these solutions, scalability can be partially solved as customers will have more access to a higher quality product, provided in a better manner by cloud providers and have improved design by developers. In essence, these solutions will improve cloud computing and the experience it offers to all customers, but the issues remain as they are mainly hardware limitations and large software clusters.

Hardware, being the greatest limitation of cloud computing can only be solved by time and progression of technology. Some examples of hardware limitations are internet speeds, storage, processing etc. Upgrades to the mentioned technologies will always positively increase the capabilities of cloud computing and reduce costs, pair that with efficient systems and helpful tools, we can have an extremely reliable and flexible system without any upfront cost.

References

Tan, J. Feng, H. Meng, X. Zhang, L. 2012, 'Heavy-traffic analysis of cloud provisioning;, *Proceedings of the 24th International Teletraffic Congress*, pp. 1-8, viewed 24 March 2021, <https://dl-acm-org.ezproxy.lib.uts.edu.au/doi/10.5555/2414276.2414309>.

Yamoah, D. Romanovsky, A. Iliasov, A. 2015, 'A Reactive Architecture for Cloud-Based System Engineering', *International Conference on Systems and Press*, pp. 77-81, viewed 23 March 2021, <https://dl-acm-org.ezproxy.lib.uts.edu.au/doi/10.1145/2785592.2785611>.

Kachelle, S. Hauck, F. 2013, 'Component-based scalability for cloud applications', *Proceedings of the 3rd International Workshop on Cloud Data and Platforms*, pp. 19-24, viewed 25 March 2021, <https://dl-acm-org.ezproxy.lib.uts.edu.au/doi/10.1145/2460756.2460760 >.

HShahapure, N. and Jayarekha, P., 2015. Replication: A Technique for Scalability in Cloud Computing. *International Journal of Computer Applications*, 122(5), pp.13-18, viewed 16/04/2021.

Blanchard, J., 2021. When Did Cloud Computing Start? The History of the Cloud. [online] Blog.servermania.com. Available at: <https://blog.servermania.com/the-history-of-cloud-computing> [Accessed 14 April 2021].

Micheloni, R., Marelli, A. and Eshghi, K., 2012. Inside Solid State Drives. New York: Springer.