Running Head: Big Data and Cloud Computing

History of Cloud Computing and Big Data

Tyler S. Unsworth

Tyler.Unsworth@cwu.edu

Central Washington University

February 24th, 2021

Abstract

This study investigates the relationship between cloud computing and big data. Cloud computing is a growing infrastructure ideology that is revolutionizing the way the modern world stores its big data. Research has indicated that cloud computing is on the rise because of its low cost relative to its physical storage counterpart. Its ability to process vast amounts of structured, unstructured, and semi-structured data while providing swift access to any of its users across the world just by having an internet connection. Now with the rise of different cloud computing services, there is a new calling to see how it may revolutionize other industries.

Previous research funded by NASA and carried out by Harvard University state that a field that could greatly benefit would be research fields like earth science. The results of this study suggested that to emulate what the cloud, a person would need 1,152 CPU cores to be comparable to using the cloud (Tawalbeh et al,. 2015). The merging of this branch of science with cloud computing would allow organizations to run smoother, cost-effective simulations without being dependent on third parties. Similar findings have been obtained from research on the topic of integrating the cloud with mobile devices. Joining these two technologies have been predicted to lead to even more benefits that can be translated to cloud heath care, vastly more advanced and capable mobile devices. It would also aid and introducing historically impoverished countries to the same cloud computing power as first-world nations.

Keywords: Big Data, Cloud Computing, Internet of Things, virtualization*,* cloud scalability*,* Mobile Cloud Computing

Contents

[Abstract 2](#_Toc66361424)

[Keywords 2](#_Toc66361425)

[Introduction 4](#_Toc66361426)

[History of Cloud Computing and Big Data 4](#_Toc66361427)

[Cloud Computing Advantages 5](#_Toc66361428)

[Cloud Computing Drawbacks 6](#_Toc66361429)

[The Future of Big Data 8](#_Toc66361430)

[Cloud Computing in Mobile Devices 9](#_Toc66361431)

[Cloud Computing in Earth Science 9](#_Toc66361432)

[Conclusion 10](#_Toc66361433)

[References 12](#_Toc66361434)

[Appendix A 14](#_Toc66361435)

[Supplemental Graphics 14](#_Toc66361436)

**Table of Tables**

[Figure 1 How Cloud Computing Analyzes Big Data 7](#_Toc66360373)

[Figure 2 Four Vs of big data 9](#_Toc66360374)

[Figure 3 Time-cost tradeoff 13](#_Toc66360375)

[Figure 4 The Cloud To University Cloudlets 13](#_Toc66360376)

Introduction

In the ever-growing world of technology, there have been transformative effects over the last ten years. Cloud computing has revolutionized the way companies and organizations manage their information technology infrastructure while eliminating the need for them to maintain expensive computing hardware (Hashem et al,. 2015). Cloud computing has proven to have numerous advantages over traditional storage hardware and provides a dramatic reduction in overhead costs, increased reliability, speed, and agility.

Investigating the relationship between big data and cloud computing could prove beneficial in understanding other technology sectors that this infrastructure could prove to be useful. Due to the expected growth of big data in the future, storing and analyzing it is a time-demanding challenge requiring extensive computational infrastructure to ensure success (Bakhouya, Zbakh, & Manneback 2020). This literature review examines the relationship between big data and cloud computing to show how this new technology will impact business and science in the future. Additionally, it illustrates the benefits and drawbacks of cloud computing in terms of cost and performance and other areas in which cloud computing implementation could be incredibly beneficial.

History of Cloud Computing and Big Data

Cloud computing has been around since the 1960s, which is believed to be invented by Joseph Carl Robnett Licklider with the vision to connect people and data from anywhere at any time. The etymology of cloud computing traces back to early network diagrams representing overlapping server icons in the shape of a cloud. The term cloud is just a metaphor that means any segment of networks that can be accessed anytime, anywhere through any number of devices connected to the internet. By the 1990s, the term was so popular that Time Magazine introduced it by calling it “The strange new world of the internet.” This period was then marked by the founding of many internet companies referred to as dot coms. By 2008, the internet of things provided a fast and innovative connection so that anyone could access it from nearly anywhere see Figure 1.

Big data is the term used to define large amounts of data that can be processed to reveal patterns, trends, and associations. Roger Mougalas coined this term from O’Reilly media in 2005. Big data can be classified into three classifications: structured, unstructured, and semi-structured (Coro, Panichi, Scarponi & Pagano 2017). The first appearance of using data to anticipate business behavior was approximately seven thousand years ago in Mesopotamia through evidence of accounting. Big data has paved the way for developing other technologies, and as big data continues to grow, so will the computer’s ability to process and analyze it.

### Cloud Computing Advantages

Cloud computing is one of the most noteworthy shifts in modern computing times. This service has proven to be one of the most formidable architectures to perform large-scale complex computing (see Figure 1). The cloud does not just radically reduce overhead by allowing companies to move away from physical storage devices and grant ease of accessing the data across many locations. But other primary benefits include virtualization. Virtualization is required to access and store data in big data environments, increasing a given computer's resource utilization, efficiency, and scalability (Bahhouya, Zbakh, & Manneback 2020). Virtualization is required to access and store data in big data environments.

Diagram

Description automatically generated

Source: (Hashem et al,. 2015)

Figure 1 How Cloud Computing Analyzes Big Data

Cloud computing services come at a dramatically lowered price which can significantly benefit many starting companies. This can be done by providing nearly instant virtual hardware resources to any business, which will allow it to be on a competitive edge with other much larger budgeted corporations (Zhou & Zhao, 2020). This benefit also translates to helping underdeveloped countries that have historically lacked the resources to install commercial-grade IT services. This gives the competitive edge for other countries to pave their way into this IT sector without having to risk so much financial starting capital. Cloud computing is magnitudes cheaper in startup costs than any other comparable infrastructure.

Due to cloud computing’s architectures ability to perform large-scale complex computing, it vastly changes an organization's difficulty to scale their data. This is because of how the cloud manages its resources through software instead of hardware. That way, providing updates is a swift process that helps resolve issues quickly and efficiently. Lastly, cloud computing's very nature is to dynamically scale its data with almost no service provider interaction.

### Cloud Computing Drawbacks

While cloud computing comes with many advantages such as reduced overhead, accessible data from anywhere, and quick processing speed, there are also some drawbacks with the current state of the technology. This infrastructure's current disadvantages are lack of control, Network connection dependency resulting in reduced access, the cost for additional business-related features, and security. Loss of control refers to the trust that the user of this cloud computing service must give to a third party. As a customer to cloud computing, there could be no guarantee that the service providers can maintain their data centers and servers to the client's standard. Therefore, the customer would have to place their best faith that the data they upload to the cloud is both physically and virtually secure. This type of loss of control could be a burden to the client, and they may attribute this absence of security or control a primary drawback.

Network connection dependency refers to the need for a strong day in day out internet connection. Since cloud computing is trusting the files and data uploaded online, a company must have full internet access throughout their daily operations. There is no way around this, so areas that experience outages may find this use unfavorable. Any loss in the network could lead to more significant time delays in the retrieval of this information (Bakhouya, Zbakh, & Manneback 2020). If any outages happen on the service provider's side, the client would be at the mercy of retrieving any data.

Many service providers offer cloud computing packages, and not all are created equally. When choosing cloud computing providers, only a limited few have unlimited bandwidth for a company’s data. Therefore, many of the current service providers start with a free package but then charge for their more relevant premium packages. So many companies may have to add extra computing or storage expenses as their business grows. This could prove to be a drawback for many starting companies wanting to store their data in the cloud.

There has been evidence of cloud computing hacking cases proving that all cloud service providers are not all secure. Since many businesses may upload sensitive data, many cannot afford to fall victim to hackers. This could be a prime example of cloud computing’s current drawbacks. There also is the case for any technical issues experienced by the client. Since all a company’s data would be uploaded to a third-party server, there would be no way for the client to help remedy it but instead would be required to call the service provider's technical support. Some service providers also do not host twenty-four-hour technical support.

# The Future of Big Data

With the advancement of information technology in the last decade, there has been a call for new and innovative systems to change the way people manipulate and store data. The volume of data in 2017 alone was predicted to be approximately 2.7 trillion gigabytes and is expected to grow exponentially in the future (Coro, Panichi, Scarponi & Pagano 2017). As stated above, big data is the term used to define large amounts of data that can be processed to reveal patterns, trends, and associations. Professors have done research at the Alessandro Faedo Institute of Information Science and Technologies in this area. They found the data can be classified to have four attributes referred to as the four Vs. The four Vs are Volume, Variety, Velocity, and Value see Figure 2.

Chart, diagram, bubble chart

Description automatically generated

Source: (Hashem et al,. 2015)

Figure 2 Four Vs of big data

This definition was investigated and concluded by the Alessandro Faedo Institute of Information Science and Technologies by compiling over one hundred peer-reviewed articles, journals, and conferences. The authors found an inconsistent definition of the term big data and tried to define it in terms of its use and characteristics in big data computation.

### Cloud Computing in Mobile Devices

With many recent advances in information technology, including those in cloud computing, the internet of things is shrinking the world into a small neighborhood of things. In the past few years, computing ideologies have been far-reaching into various fields of business and science. Some of those are now crossing into mobile devices. Recent advances in mobile devices are bringing phones from 3G to 4G and now to 5G. Mobile devices themselves are already integrated into our current lifestyle to help us perform various tasks, stay connected, and help people manage our data.

The mobility of mobile phones has forever changed the way humans use technology. Now that there is current research going on in Umm Al-Qura University in Mecca, Saudi Arabia, and King Abdulaziz University. Researchers, there are on the front line investigating mobile cloud computing’s capabilities and benefits. These two combined technologies are gaining momentum as Mobile Cloud Computing and are set to benefit many technology sectors such as cloud-healthcare systems (Tawalbeh, Mehmood, Benkhlifa & Song 2016). These institutions' research on mobile cloud computing indicates that mobile devices can overcome their limited memory, CPU power, and battery life see figure 5 to see how the cloud would reach devices. Integrating mobile devices with cloud computing allows those devices to utilize the cloud’s unlimited services while maximizing the devices’ ability to reduce the management and cost.

### Cloud Computing in Earth Science

Cloud computing has a wide range of fields that it can complement due to its ability to handle large-scale big data computations for both a reasonable time and price (Zhuang et al., 2020). Another field in which cloud computing could complement in Earth science. NASA in 2018 funded a two yearlong study to investigate how cloud computing could aid in running Earth science simulations see figure 4 to see the time-cost tradeoff. Earth science simulations are computationally expensive and usually require using a supercomputer that is typically owned by Universities and Laboratories (Zhuang et al., 2020). With that, there are fees and learning curves that come with using other institutions' equipment.

Harvard University completed its investigation by running micro benches to test hardware components. These tests gave the researchers the peak performance of a CPU and then compared the values to the performance running the cloud simulations. Their investigation revealed that it would take approximately 1,152 CPU cores (Zhuang et al., 2020) to be comparable to using the cloud. They, therefore, found that cloud computing is a much more efficient alternative to running complex earth science simulations.

# Conclusion

Through the investigation of the relationship between cloud computing and big data, this literature review has outlined the plethora of its uses and how it may benefit other sectors of

business and science. One of its most significant advantages is its reduced cost, ability to process large data sets, and its ability to help underprivileged countries get the connection and data processes they need. Through the investigation, it has also been found that cloud computing would be able to help earth scientists run more efficient and less costly simulations. It also explored the clouds use in mobile devices and how the integration could prove useful. The integration of cloud computing in mobile phones would drastically increase the productivity of its users by offering them better battery life, more processing speed and ability to obtain data from anywhere

Some suggestions that could be made for further research in this field. Would be a more concise definition for cloud computing and big data. There have been numerous publications that define differently the ways to characterize big data by what is known as the big Vs. More research can be invested in that area, so there could be some common ground amongst cloud computing professionals. Lastly, from further research, a suggestion could be made to invest more resources in cloud computing security as there have been numerous occasions uploaded data has been hijacked.

References

Bakhouya, M., Zbakh, M., Essaaidi, M., & Manneback, P. (2020). Cloud computing, IoT, and big data: Technologies and applications. *Concurrency and Computation Practice and Experience,* 32

Buyya, R., Srirama, S. N., Casale, G., Calheiros, R., Simmhan, Y., Varghese, B., Gelenbe, E., Javadi, B., Vaquero, L. M., & Netto, M. A. S. (2017). A Manifesto for Future Generation Cloud Computing: Research Directions for the Next Decade. ACM Computing Surveys, 51

Coro, G., Panichi, G., Scarponi, P., & Pagano, P. (2017). Cloud computing in a distributed e-infrastructure using the web processing service standard. *Concurrency and Computation Practice and Experience*, 29

Chang, V., Kuo, Y.-H., & Ramachandran, M. (2016). Cloud computing adoption framework: A security framework for business clouds. *Future Generation Computer Systems*, 57, 24–41

Hashem, I. A., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115

Saif, S. & Wazir, S. (2018). Performance Analysis of Big Data and Cloud Computing Techniques: A Survey. *Procedia Computer Science*, 132

Tawalbeh, L., Mehmood, R., Benkhlifa, E., & Song, H. (2016). Mobile Cloud Computing Model and Big Data Analysis for Healthcare Applications. *IEEE Access*, 4, 6171–6180

Villalpando, L. E. B., April, A., & Abran, A. (2014). Performance analysis model for big data applications in cloud computing. *Journal of Cloud Computing*, 3

Zhuang, J., Jacob, D. J., Lin, H., Lundgren, E. W., Yantosca, R. M., Gaya, J. F., & Sulprizio, M. P. (2020). Enabling High-Performance Cloud Computing for Earth Science Modeling on Over a Thousand Cores: Application to the GEOS-Chem Atmospheric ChemistryModel. *James Journal of Advances in Modeling Earth Systems*, 12

Zhou, Z. & Zhao, L. (2020). Cloud computing model for big data processing and performance optimization of multimedia

a communication. *Cloud Communications*, 160, 326–332

# Appendix A

### Supplemental Graphics

Chart, scatter chart

Description automatically generated

Source: (Zhuang et al., 2020)

Figure 3 Time-cost tradeoff

Diagram

Description automatically generated

Source: (Tawalbeh, Mehmood, Benkhlifa & Song 2016)

Figure 4 The Cloud To University Cloudlets