The Seybold Report

ISSN 1533-9211

DOI 10.17605/OSF.IO/ETUVP

A REVIEW ON BIG DATA ANALYTICS IN THE CLOUD: RESHAPING THE LANDSCAPE OF THE FUTURE

CHEERYL V. GARCIA

Assistant Professor, Iloilo Science and Technology University, Miagao Campus, Igtuba, Miagao, Iloilo, Philippines. Email: cheerylvgarcia@gmail.com, ORCID ID: https://orcid.org/0009-0002-5388-5300

Abstract

The cloud has emerged as the preferred platform for storage and processing massive quantities of data, and big data has become a crucial component of contemporary commercial operations. Big data processing now offers a higher degree of scalability, flexibility, and cost-effectiveness because of the use of internet-based services. This paper's objective is to look into the big data as utilized in the cloud and what impact it has on various industries. This paper utilized the research article and paper review which pertains to applications of "big data in the cloud". The author concentrated on articles that were open for review, summarizing the key findings and their relevance to the subject of study. The collected articles included the studies across different sections such as healthcare, banking, manufacturing, retail, and more. In the study, risks and other opportunities were identified. In addition, advantages and disadvantages was also enumerated.

Keywords: Big Data, Data in the Cloud, Future of Big Data, Narrative Literature Review

INTRODUCTION

The cloud has emerged as the preferred platform for storage and processing massive quantities of data, and big data has become a crucial component of contemporary commercial operations. Big data processing now offers a higher degree of scalability, flexibility, and cost-effectiveness because of the use of internet-based services. This paper's objective is to look into the big data as utilized in the cloud and what impact it has on various industries.

The term "big data" illustrates the immense and complex datasets generated by multiple sources, including sensors, social media, and transactional systems (Techtarget.com, n.d). The quantity, velocity, variety, and authenticity of the data in these datasets define them. A scalable and economical platform for storing and analyzing these huge datasets is provided by the cloud. Rather than just investing in expensive hardware and software infrastructure, cloud computing allows businesses to rent computing power as needed. Organizations can increase or reduce their computer resources in response to shifting business requirements thanks to this on-demand model. Big data processing has been revolutionized by this flexibility, which enables businesses to process enormous amounts of data fast and effectively.

As we live in the digital age, our data is everywhere, it could be on social media, in the applications that we use every day like payment channels and shopping apps, mobile phones, servers and sensors from the Internet of Things. Digitalization rate has increased enormously over the years and now over 94% of information is digital. Statistics show roughly 44 zettabytes (44 1021 bytes) of data are generated every day (Berisha et.al, 2022). Each person generates 1.7 MB of data every second in which data are taken from Statista.

ISSN 1533-9211

DOI 10.17605/OSF.IO/ETUVP

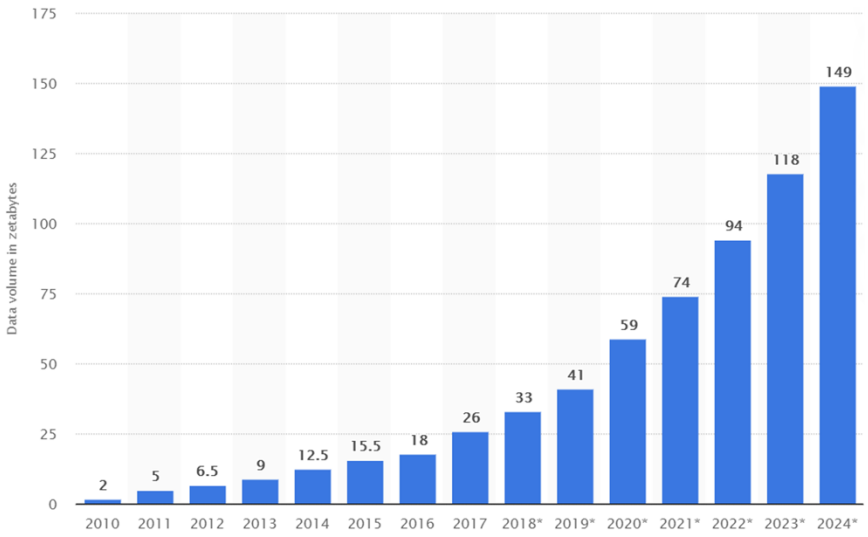

Figure 1: Data Volume Generated every day in Zettabyte as Projected from 2010 to 2024

Among the big data trends for the future, one of which is extended adoption of predictive analytics in which previous information is used to predict future results utilizing both machine learning and statistical methods (IBM, n.d.). In addition, organizations may improve their performance and make better decisions with the aid of predictive analytics. Predictive analytics, for instance, can be used to increase revenue, decrease customer attrition, and boost operational effectiveness, according to research.

In this paper, the author will look into how big data transforms the way different industries benefit the technology, look into different service platforms that cater to cloud computing, and identify the potential risks and challenges.

METHODS

This paper utilized a narrative literature review of journal article and papers which pertains to applications of "big data in the cloud". It involves gathering, critiquing and summarizing all the searched journal articles. The author concentrated on articles that were open for review, summarizing the key findings and their relevance to the subject of study. Figure 2 shows the approach on how the author conducts the study using the narrative literature review.

ISSN 1533-9211

DOI 10.17605/OSF.IO/ETUVP

Figure 2: Design for Conducting Narrative Literature Review

RESULTS

The following is a list of issues that are frequently talked up in different papers and articles that highlights the usefulness in utilizing cloud computing and big data.

Big Data: It's Use in Industry and Its Impact

In processing of massive data, higher scalability and performance are achieved via parallelization approaches and algorithms as found in Venkatesh et al (2015, as cited in S. Harini et al, 2018). There are various applications for big data in the cloud across many different sectors, including healthcare, banking, manufacturing, retail, and more. The following are the list of industry benefitting from the Big Data:

Healthcare: In this field, it is utilized in forecasting possible outbreaks of diseases, profiling of patients that are high-risk, and improve treatments' outcomes of patients. Cloud-based solutions allow healthcare providers to securely store and share patient data, collaborate with other healthcare providers, and analyze data in real-time. In many countries with the potential of big data that can do in healthcare, it can forecast and prevent outbreaks. It can diagnose and identify new diseases with less expense. With the generated data, it can treat and provide effective treatment models. Thorough reports which were transformed into useful, crucial insights can be enabled in with use of big data which can improve care (Berisha et.al, 2022).

Finance: Financial bodies can utilize Big Data analytics to uncover fraudulent transactions, assess credit risk, and analyze customer behavior. Cloud-based solutions allow financial institutions to get and perform data retrieval in processing large volumes of data quickly and efficiently, in not needing for expensive hardware. Investors may it be business or individuals may use the data trends and forecasting techniques to create decisions especially on investments (Ejrami, M. & Salehi, N., 2022 as cited in Al-Lozi et.al, 2022). Big data is utilized by corporate sectors to their manage resources and working with their capital effectively.

ISSN 1533-9211

DOI 10.17605/OSF.IO/ETUVP

Additionally, financial sectors, particularly the banking sector, produced vast amounts of data, including customer details, transaction details, and financial product details, which they may use for marketing efforts. Lui (2015), in addition, it can benefit the utilization of big data in creating suited application for their social media, can perform reliable risk management, and can provides security intelligence. Fraud detection and risk assessment are also now being utilized in the financial sector.

Manufacturing: Utilized to enhance quality assurance, streamline its manufacturing process, and also improve supply chain management. Cloud-based solutions enable manufacturers to share data across their supply chain partners, monitor equipment in real-time, and predict maintenance needs.

Retail: It helps the retailers understand more the consumer behavior, can create best pricing plans, and can have a better control with their inventories. Cloud-based solutions allow retailers to collect and examine data in a various source, which may include social media sites, online sales, at the same in-store transactions.

Cloud Computing

Muniswamaiah et al. (2019) with the utilization of internet, cloud computing offers quicker computing services such data retrieval from different servers, accessing data from storage across different locations, databases, networking, software analytics, and parallel data processing. It helps companies to concentrate on their primary operations by completely using a variety of prebuilt services. This includes storing, networking resource as by demand and abstracting processes.

Figure 3 presents the services for each computing facilities and how they differ with each other

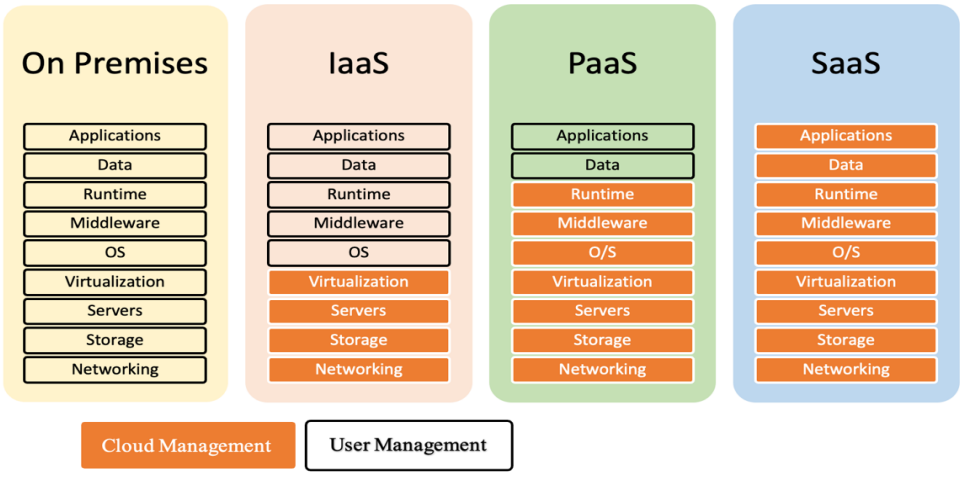

Figure 3: Differentiation of Every Computing Facilities

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/ETUVP

Software as a Service (SaaS)

It is remotely controlled and maintained in position of vendors as well as it distributed over the network. In a one-to-many strategy, the provider delivers software that is consumed at any time by all of the contracted clients via pay-per-use agreement or it may come with a contract based on their usage.

Platform as a Service (PaaS)

It provides the tools such as hardware and applications available online for programmers to use in creating unique apps. This service if utilized is cost-effective. Preforming the development, along with testing and deployment are made easy. Muniswamaiah et al, (2019) found that with the help of PaaS, operating system management, performing virtualization, configuring servers and networks, these enterprise operations enable industries to create and develop application that can be incorporated into this system.

Infrastructure as a Service (IaaS)

This service internet-based computing paradigm makes the use of virtualization technology to provide self-servicing platforms for businesses to navigate, track, and control distant data center infrastructures as enumerated: computing, storing, and networking services. Muniswamaiah et al. (2019) mentioned that IaaS consumers are in responsible of controlling of applications, data, overseeing the runtime, middleware and operating system management while vendors performs the management of virtualizations and infrastructure, redundancy of storing capacity, ang the management of network. This service offers adequate capability to data centers without the need for ongoing physical maintenance.

Advantages of Big Data in the Cloud

As previously indicated, businesses across the wide array of industry sectors are utilizing big data to promote decision-making based on data. El-Seoud et al. (2017, as cited in Yeboah, F., 2022), mentioned that together with technology sector, big data is now used and appreciated by other sectors too, namely government, retail, supply chain management, and education. Better efficiency for analysis of big data is a result in utilizing cloud computing by allowing resource allocation across multiple servers.

With the help of the extensive toolkit that the big data software supports, an individual is able to use the big data by evaluating the treats it possessed internally by planning out onto the full data presentation throughout the business (Venkatesh et al., 2015). El-Seoud et al. (2017, as cited in Yeboah, F., 2022) added that with the help of internet-based service and big data, costs could be decreased. Moreover, another advantage of using it over the cloud includes quicker and smarter decision-making. It provides an avenue to create new services and products, it can produce better product recommendations. Lastly, it can detect fraud activities faster.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/ETUVP

DISCUSSIONS

Mentioned Approaches Intended for the Big Data Security in a Cloud Computing Ecosystem

1. Encryption

Protecting the data could become a severe concern for any business or organization, therefore, different encryption techniques are necessary to keep the data of the user secured.

2. Nodes Authentication

Every time a node enters the cluster, authentication is required. Once there is an identified malicious cluster which entered in the node, it will not be authenticated and will not allow in the node.

3. Access Control and Security

One of the best measures of security is creating a different access control in the distributed environment. Ramya et al. (2023) states that by the using of Modules in Linux Security in the Linux kernels, can stop information from leaking, and contains a functionality that offers a framework for providing access control security policy. Jagani et al. (2021 as cited in Ejrami, M. and Salehi, N., 2022) mentions that it poses issues on vulnerabilities with regards to integrating the big data with cloud computing. In addition, it was also mentioned that implementation of data security polices with a structured data are not effective when handles through the use of conventional DBMS in a highly heterogenous with unstructured data. With this, there is a encourage the need to integrate in the new data management structures of storage. Thus, ensuring data privacy, availability and veracity turn out to be fundamental in this time for data owners has only limited control over their data.

4. Risk and Challenges on Big Data

With all the advantages mentioned with big data and cloud computing, it still poses risks and challenges that users need to be mindful when using big data in a cloud-based setting. One of which is the security on a cloud ecosystem (El-Seoud et al., 2017). Jain (2019) enumerated the list of security tools needed to secure the data such as user authentication, level of access control, encryptions of data, intrusion and detection including logging of events and monitoring. Additional challenge is the form of data and where it is located that the data are in different locations. It should be in consideration how the kind of processing should be done like in the manner of parallel processing or moving of the data to processing environment. In addition, data storage is another challenge posed by big data in the cloud. Jagani et al. (2021) noted that when designing a storage systems to become efficient in dealing analysis there must be a need in dealing with detailed attention so that it will produce the same output even that the data came from a variety of sources. Nasser and Tariq (2015) states that big data visualization can be reveled in hidden patterns and previously unidentified correlations can be used to facilitate decision-making. Often times, big data is heterogenous in form, organization and semantics and it makes the visualization critical to make sense.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/ETUVP

Alongside data analytics, it is said that cloud computing can makes the businesses competitive with benefits like polling of resources and on demand service in a cost-effective manner (Balachandran & Prasad, 2017 as cited in Binti, K & Salleh, A., 2021). Despite having advantages, there are also downsides with regards the latter such as privacy and security.

With this, the author can state that big data in the cloud are reshaping the approach every industry in how they operate, providing them with the ability in storing, processing and analyzing large data sets without the need for costly infrastructure. The author presented that the benefits of having a cloud-based big data solutions are clear which includes scalability, flexibility, cost-effectives, and real-time analytics.

Declaration of Interest

The author declares no potential conflicts of interest with respect to this research, authorship and publication.

About The Author

Cheeryl Villaren Garcia, a licensed Real Estate Broker, a licensed Financial Advisor and is currently an Assistant Professor at Iloilo Science and Technology University Miagao Campus, Iloilo, Philippines. At the time being, enrolled in Doctor of Information Technology with area of Specialization in Business Intelligence and Data Analytics at Angeles University Foundation. She completed her Master of Management major in Information Technology at Sultan Kudarat State University.

References

1) Balachandran, B. M., & Prasad, S. (2017). Challenges and Benefits of Deploying Big Data Analytics in the Cloud for Business Intelligence. Procedia Computer Science. Retrieved from https://uniselinus.education/sites/default/files/2022-12/frederick-twum-yeboah.pdf. https://doi.org/10.1016/j.procs.2017.08.138. Accessed March 3, 2023

2) Ejrami, M. & Salehi, N. (03 December 2022). The Effect of Big Data on Firm Value: The Mediation Roles of Innovation and Financial Performances. Retrieved in https://www.scienceopen.com/document_file/7983975a-9f75-4da4-b1ec-65d028ba3ff2/ScienceOpenPreprint/Paper.pdf

3) Gartner. (n.d.). Definition of Software as a Service (SaaS) - Gartner Information Technology Glossary. Retrieved from https://www.gartner.com/en/information-technology/glossary/software-as-a-service-saas. Accessed March 7, 2023

4) IBM. (n.d.). What is predictive analytics? Www.ibm.com. Retrieved from https://www.ibm.com/ph-en/topics/predictive-analytics. Accessed March 7, 2023

5) Jain, S. (2019). BigData Analysis on Cloud: Latest Trends and Challenges. Journal of Analysis and Computation.

6) Liu, Y. (2015). Big Data and Predictive Business Analytics. The Journal of Business Forecasting. Retrieved from https://www.semanticscholar.org/paper/Big-Data-and-Predictive-Business-Analytics-Liu/8b0bda12424d4373f7e6eab5276a91cfd0cb6200 . Accessed March 4, 2023

7) Ramya, S., et.al. (2023). Analyzing Big Data Challenges and Security Issues in Data Privacy. International Research Journal of Modernization in Engineering Technology and Science, 05. Retrieved from https://www.irjmets.com/uploadedfiles/paper/issue_1_january_2023/32834/final/fin_irjmets1673423558.pdf. Accessed March 4, 2023

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/ETUVP

8) S. Harini, K.Jothika & K.Jayashree. (2018). A Review of Big Data Computing ond Cloud. International Journal of Pure ond Applied Mathematics. Vol. 118 (18) 2018, 1847-1855

9) Techtarget.com. (n.d). Big Data Analytics. Retrieved from https://www.techtarget.com/searchbusinessanalytics/definition/big-data-analytics. Accessed March 7, 2023

10) Yeboah, F. (2022). Latest Trends in Big Data Utilization in Mobile and Cloud computing - Opportunities and Challenges [Dissertation, Selinus University of Science Literature. Retrieved from https://uniselinus.education/sites/default/files/2022-12/frederick-twum-yeboah.pdf

11) Zones.com. (n.d). Cloud Models. Retrieved from https://www.zones.com/site/statics/static_page.html?name=solutions%2Fcloud%2Fmodels.Accessed March 7, 2023