ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

CLOUD COMPUTING IN BIG DATA ANALYTICS: AN EMPIRICAL STUDY

[1]Dr. N. Shanmuga Priya, [2]Ms. T. Sneha, [3]Mr. L. Shreesh

[1]Associate Professor and Head, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

[2]PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

[3]PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

Abstract

A cloud framework refers to the aggregation of components like development tools, middleware and database services, needed for cloud computing, which aids in developing, deploying and managing cloud based applications strenuously, consequently making it an efficacious paradigm for massive scaling of dynamically allocated resources and their complex computing. Big Data is a concept that deals with storing, processing and analyzing large amounts of data. Cloud computing on the other hand is about offering the infrastructure to enable such processes in a cost-effective and efficient manner. Many sectors, including among others businesses (small or large), healthcare, education, etc. are trying to leverage the power of Big Data. In healthcare, for example, Big Data is being used to reduce costs of treatment, predict outbreaks of pandemics, prevent diseases etc. This paper describes how cloud and big data technologies are converging to offer a cost-effective delivery model for cloud-based big data analytics. Cloud computing is a powerful technology to perform massive-scale and complex computing. It eliminates the need to maintain expensive computing hardware, dedicated space, and software. Massive growth in the scale of data or big data generated through cloud computing has been observed. Addressing big data is a challenging and time demanding task that requires a large computational infrastructure to ensure successful data processing and analysis. In this paper the relationship between big data and cloud computing, the classification of big data and the scope of big data analytics are discussed.

Keywords: *- Big Data, Cloud Computing, Infrastructure as a Service (IaaS), Internet of Things (IoT), Platform as a Service (PaaS), and Software as a Service (SaaS)*

1. INTRODUCTION

While cloud is all about delivering pay-as-apply, on demand, flexible and scalable services, BDA emphases on revolutionizing its information assets represented by 3 V's symbolizing Volume, Velocity and Variety, into another V symbolizing Value (to organizations' business) [2]. Illustration of each of these big data dimensions is given below. [9]

- Volume: The amount of data propagated.
- Velocity: The speed with which the data is propagated.
- Variety: The heterogeneity of data type propagated.

Cloud Computing offers the possibility of accommodating a massive volume of data over the internet through hardware virtualization, thus, adding to the availability, scalability and accessibility of Big Data [3]. Moreover, cloud computing also delivers exclusive statistical tools for resourceful processing and analyses of big data through a service termed as Big Data as a Service (BDaaS) [12]. Subsequently, both big data and cloud unify together to bring in value to enterprises by enhancing the agility, elasticity, accessibility and the ease of processing of cloud based big data, and, by reducing its cost of ownership and implementation complexity of big data solutions [4]..

Cloud computing is one of the most significant shifts in modern ICT and service for enterprise applications and has become a powerful architecture to perform large-scale and complex computing. The advantages of cloud computing include virtualized resources, parallel processing, security, and data service integration with scalable data storage. Cloud computing can not only minimize the cost and restriction for automation and computerization by individuals and enterprises but can also provide reduced infrastructure maintenance cost, efficient management, and user access [2]. As a result of the said advantages, a number of applications that leverage various cloud platforms have been developed and resulted in a tremendous increase in the scale of data generated and consumed by such applications.

The cloud and big data analytics are often used together. This is because big data requires huge computational power and storage. Cloud computing offers on-demand storage, computation resources, and tools to store and analyze big data. Hence, big data cloud computing and big data cloud analytics are becoming increasingly popular. The rise of big data on cloud computing has made the process of analyzing big data more efficient. Businesses can choose from three types of cloud computing services, IaaS, PaaS, and SaaS, for cloud-based big data analytics [5]. These services are available on a pay-per-use or subscription basis, which means users only pay for the services they use. Cloud analytics essentially means storing and analyzing data in a big data cloud instead of on-premises systems of the organization. This includes any type of data analytics that is performed on systems hosted in the cloud, including big data analytics.

1.1. What Is Big Data?

Recently, the term of Big Data has been coined referring to those challenges and advantages derived from collecting and processing vast amounts of data. The sources of huge quantity of information are those applications that gather data from click streams, transaction histories, sensors, and elsewhere. However, the first problem for the correct definition of 'Big Data' is the name itself, as we might think that it is just related to the data volume. The heterogeneous structure, diverse dimensionality, and variety of the data representation, also have significance in the big data.

Big data is mainly available in two forms, structured and unstructured [6].

- **Structured data** comprises the data already available in the organization's databases and spreadsheets. This type of data is mostly numeric.
- **Unstructured data** is usually human-driven data, such as the information collected from customers' comments on social media, product purchases, questionnaires, personal apps, etc. Since this type of data is unorganized and available in different formats, it needs modern tools for proper processing.

There are several benefits of big data in the cloud and big data analytics cloud [7]:

- **Scalability -** Cloud computing for big data offers flexible, on-demand capabilities. With big data cloud technology, organizations can scale up or scale down as per their needs. For example, organizations can ask cloud-based big data solutions providers to increase cloud storage as the volume of their data increases. Businesses can also add data analysis capacity as needed. Big data cloud server's help businesses respond to customer demands more efficiently.
- **Higher Efficiency -** Cloud computing for big data analytics provides incredible processing power. This makes big data processing in cloud computing environments more efficient compared to on-premise systems.
- **Cost Reductions -** When it comes to big data on-premise vs. cloud, another major difference is cost. In comparing big data cloud vs. on-premise, on-premises systems involve different costs, such as power consumption costs, purchasing and maintaining hardware and servers, replacing the hardware, etc.

- However, with cloud and big data cloud technologies, there are no such costs because the cloud service providers are responsible for everything. Additionally, cloud services are based on a pay-per-use model, which further reduces the cost.
- **Disaster Recovery -** Data of any size is a valuable asset for organizations, so it's important not to lose it. However, cyber-attacks, equipment failure, and power outages can result in data loss, especially if you're using an on-premise system. On the other hand, a big data cloud service replicates data to ensure high availability and security. Hence, cloud computing for big data helps organizations recover from disasters faster.

1.1.1. Classification of Big Data

Big data are classified into different categories to better understand their characteristics. Fig. 1.1 shows the numerous categories of big data. The classification is important because of large-scale data in the cloud. The classification is based on five aspects: (i) Data sources, (ii) Content format, (iii) Data stores, (iv) Data staging, and (v) Data processing.

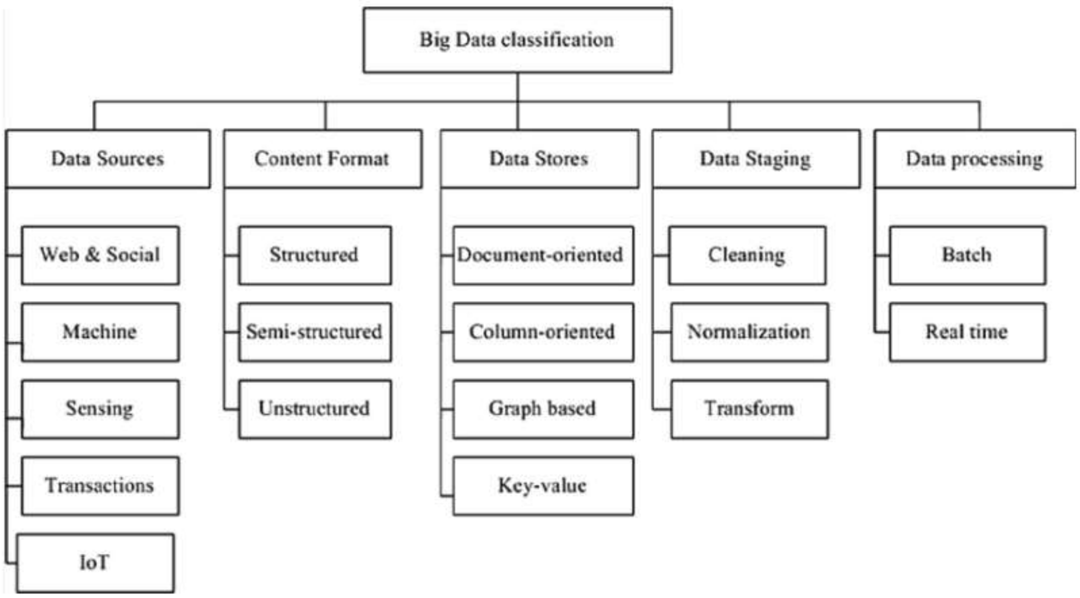

Fig. 1.1: Classification of Big Data

2. CLOUD COMPUTING

Cloud computing is a fast-growing technology that has established itself in the next generation of IT industry and business. Cloud computing promises reliable software, hardware, and IaaS delivered over the Internet and remote data centers [8]. Cloud services have become a powerful architecture to perform complex large-scale computing tasks and span a range of IT functions from storage and computation to database and application services. The need to store, process, and analyze large amounts of datasets has driven many organizations and individuals to adopt cloud computing [10]. A large number of scientific applications for extensive experiments are currently deployed in the cloud and may continue to increase because of the lack of available computing facilities in local servers, reduced capital costs, and increasing volume of data produced and consumed by the experiments. In addition, cloud service providers have begun to integrate frameworks for parallel data processing in their services to help user's access cloud resources and deploy their programs [6].

2.1. Relationship between Cloud Computing and Big Data

Cloud computing and big data are conjoined. Big data provides users the ability to use commodity computing to process distributed queries across multiple datasets and return resultant sets in a timely manner. Cloud computing provides the underlying engine through the use of Hadoop, a class of distributed data-processing platforms. The use of cloud computing in big data is shown in Fig. 3. Large data sources from the cloud and Web are stored in a distributed fault-tolerant database and processed through a programming model for large datasets with a parallel distributed algorithm in a cluster [11]. The main purpose of data visualization, as shown in Figure 2.1, is to view analytical results presented visually through different graphs for decision making.

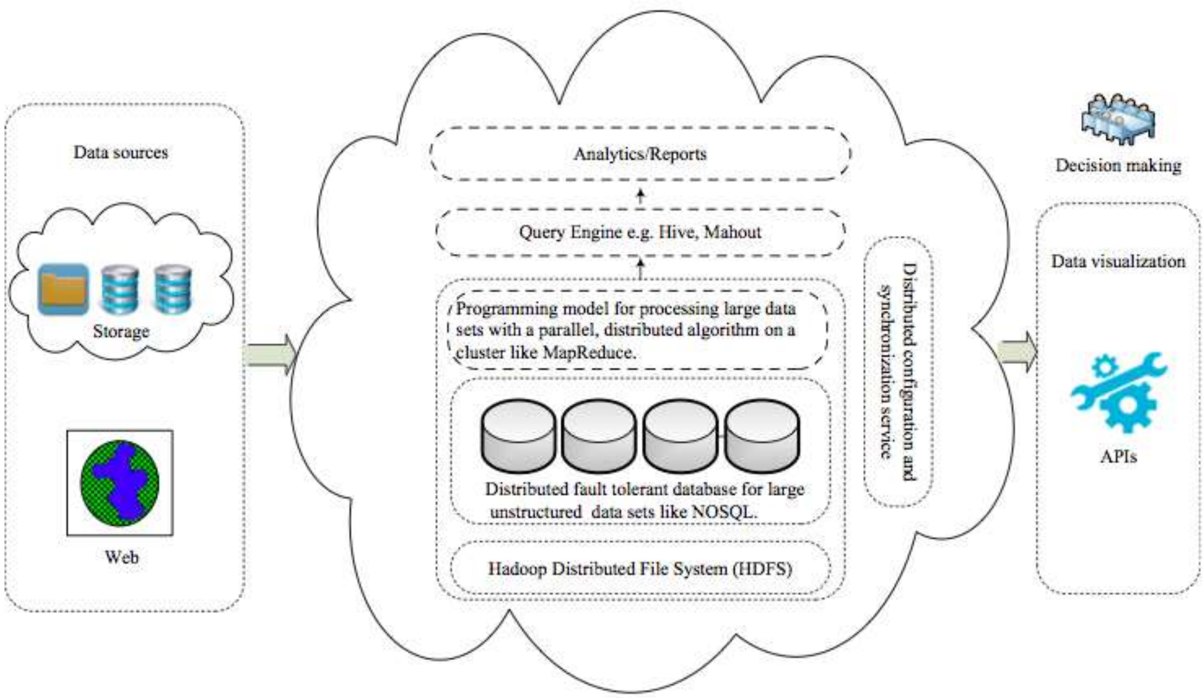

Fig. 2.1: Cloud computing usage in big data

Big data utilizes distributed storage technology based on cloud computing rather than local storage attached to a computer or electronic device. Big data evaluation is driven by fast-growing cloud-based applications developed using virtualized technologies. Therefore, cloud computing not only provides facilities for the computation and processing of big data but also serves as a service model.

2.1.1. Cloud Computing Environments for Big Data

Cloud Computing is an environment based on using and providing services [13]. There are different categories in which the service-oriented systems can be clustered. One of the most used criteria to group these systems is the abstraction level that is offered to the system user. In this way, three different levels are often distinguished [9]; Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) as we can observe in Figure 2.2.

Cloud Computing offers scalability with respect to the use of resources, low administration effort, flexibility in the pricing model and mobility for the software user. In particular, a common Big Data analytics framework [10] is depicted in Figure 2.2.

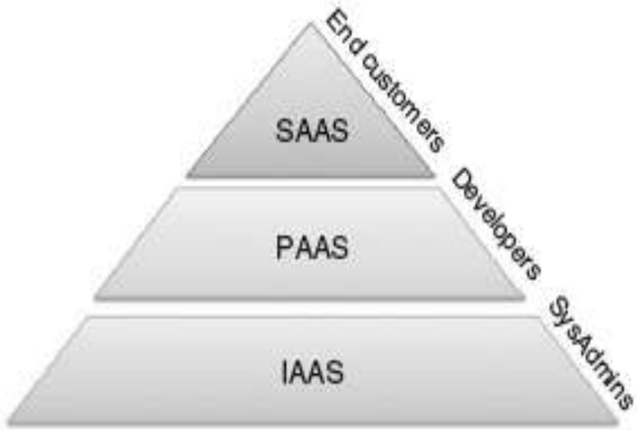

Fig. 2.2: Illustration of the layers for the Service-Oriented Architecture

A file system for the storage of Big Data, i.e., a wide amount of archives of large size. This layer is implemented within the IaaS level as it defines the basic architecture organization for the remaining tiers [15].

- An execution tool to distribute the computational load among the computers of the cloud. This layer is clearly related with PaaS, as it is kind of a 'software API' for the codification of the Big Data and BI applications.
- A query system for the knowledge and information extraction required by the system's users, which is in between the PaaS and SaaS layers.

2.1.2. The Scope of Big Data Analytics

Early interest in big data analytics focused primarily on business and social data sources, such as e-mail, videos, tweets, Facebook posts, reviews, and Web behavior. The scope of interest in big data analytics is growing to include data from intelligent systems, such as in-vehicle infotainment, kiosks, smart meters, and many others, and device sensors at the edge of networks—some of the largest-volume, fastest-streaming, and most complex big data [14]. Ubiquitous connectivity and the growth of sensors and intelligent systems have opened up a whole new storehouse of valuable information. Interest in applying big data analytics to data from sensors and intelligent systems continues to increase as businesses seek to gain faster, richer insight more cost-effectively than in the past, enhance machine-based decision making, and personalize customer experiences.

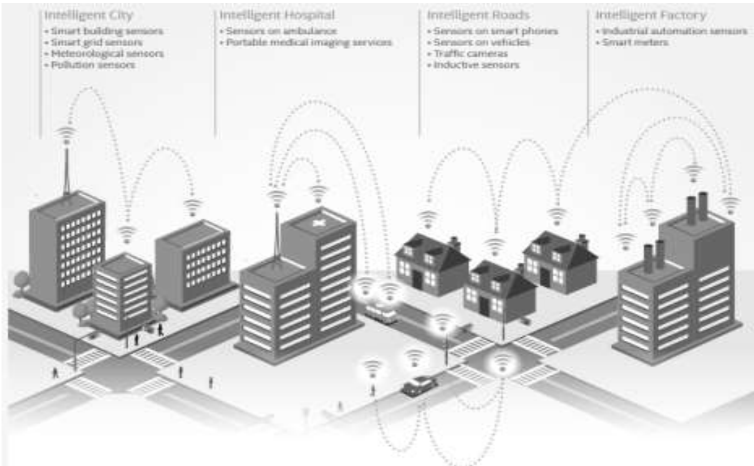

Fig. 2.3: Big Data in Context: Smart City Example

For big data analytics in cloud computing, the data (both structured and unstructured) is gathered from different sources, such as smart devices, websites, social media, etc. The next step involves cleaning and storing this large amount of data. Companies then use big data cloud tools by big data cloud providers to process this data for analysis [16].

The big data cloud architecture below will help you understand cloud big data, cloud computing big data, and how cloud computing and big data are used together:

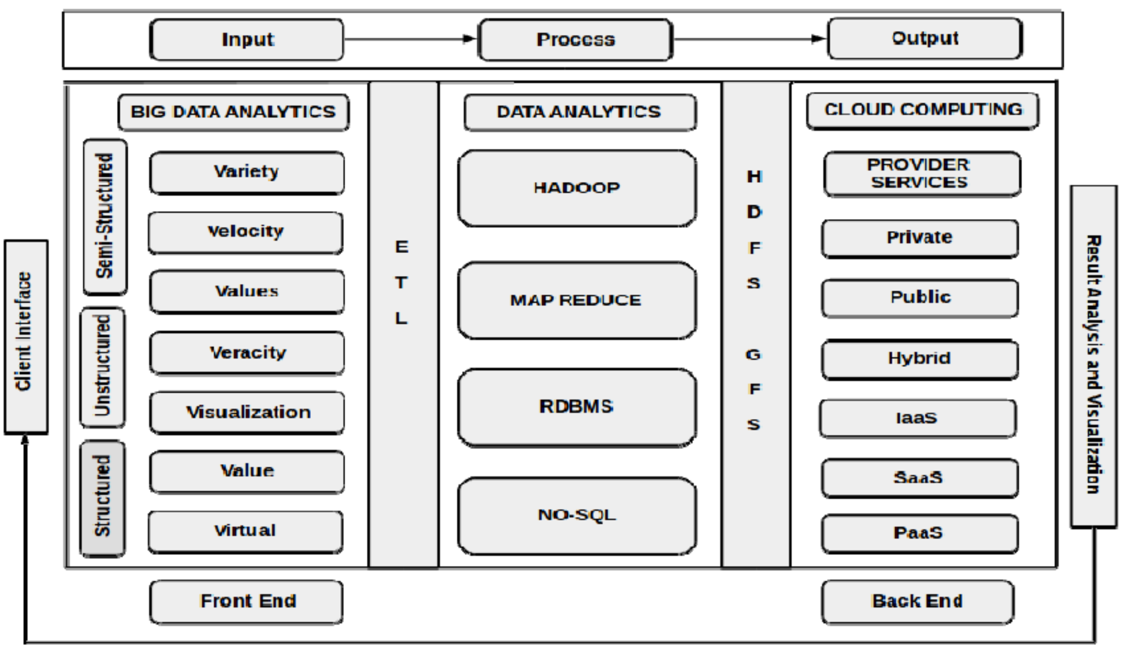

Fig. 2.4: Big Data Analysis in Cloud Computing

One of the most common cloud computing platforms for big data processing and analysis is AaaS. AaaS or Analytics as a service refers to a big data cloud solution that provides analytics software and procedures. It provides efficient business intelligence (BI) solutions that help organize, analyze, and present big data so that it is easy to interpret.

AaaS involves advanced data analytics technologies, such as machine learning algorithms, AI, predictive analytics, data mining, etc., to analyze data and show trends.

3. CONCLUSION

The size of data at present is huge and continues to increase every day. The variety of data being generated is also expanding. The velocity of data generation and growth is increasing because of the proliferation of mobile devices and other device sensors connected to the Internet. These data provide opportunities that allow businesses across all industries to gain real-time business insights. The use of cloud services to store, process, and analyze data has been available for some time; it has changed the context of information technology and has turned the promises of the on-demand service model into reality. In this study, we presented a review on the rise of big data in cloud computing. Here, proposed a classification for big data, a conceptual view of big data, and a cloud services model. In the future, significant challenges and issues must be addressed by the academia and industry. Researchers, practitioners, and social science scholars should collaborate to ensure the long-term success of data management in a cloud computing environment and to collectively explore new territories.

REFERENCES

[1] Berisha, B., Mëziu, E., & Shabani, I. (2022). Big data analytics in Cloud computing: an overview. Journal of Cloud Computing, 11(1), 24.

[2] Lee, R. (Ed.). (2020). Big Data, Cloud Computing, and Data Science Engineering. Springer.

[3] Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. International Journal of Digital Earth, 10(1), 13-53.

[4] Yetis, Y., Sara, R. G., Erol, B. A., Kaplan, H., Akuzum, A., & Jamshidi, M. (2016, July). Application of big data analytics via cloud computing. In 2016 world automation congress (wac) (pp. 1-5). IEEE.

[5] Skourletopoulos, G., Mavromoustakis, C. X., Mastorakis, G., Batalla, J. M., Dobre, C., Panagiotakis, S., & Pallis, E. (2017). Big data and cloud computing: a survey of the state-of-the-art and research challenges. Advances in mobile cloud computing and big data in the 5G Era, 23-41.

[6] Sandhu, A. K. (2021). Big data with cloud computing: Discussions and challenges. Big Data Mining and Analytics, 5(1), 32-40.

[7] Das, M., & Dash, R. (2021). Role of cloud computing for big data: A review. Intelligent and Cloud Computing: Proceedings of ICICC 2019, Volume 2, 171-179.

[8] Huttunen, J. E. N. N. I. F. E. R., Jauhiainen, J., Lehti, L. A. U. R. A., Nylund, A. N. N. I. N. A., Martikainen, M., & Lehner, O. M. (2019). Big data, cloud computing and data science applications in finance and accounting. ACRN Journal of Finance and Risk Perspectives, 8, 16-30.

[9] Suhasini, N. S., & Puli, S. (2021, November). Big Data Analytics In Cloud Computing. In 2021 Sixth International Conference on Image Information Processing (ICIIP) (Vol. 6, pp. 320-325). IEEE.

[10] Dzulhikam, D., & Rana, M. E. (2022, March). A critical review of cloud computing environment for big data analytics. In 2022 International Conference on Decision Aid Sciences and Applications (DASA) (pp. 76-81). IEEE.

[11] Saif, S., & Wazir, S. (2018). Performance analysis of big data and cloud computing techniques: a survey. Procedia computer science, 132, 118-127.

[12] Zbakh, M., Essaaidi, M., Manneback, P., & Rong, C. (Eds.). (2019). Cloud Computing and Big Data: Technologies, Applications and Security. Springer.

[13] Al-Shawakfa, E., & Alsghaier, H. (2018). An empirical study of cloud computing and big data analytics. International Journal of Innovative Computing and Applications, 9(3), 180-188.

[14] Neves, P. C., Schmerl, B., Cámara, J., & Bernardino, J. (2016, April). Big Data in Cloud Computing: features and issues. In International Conference on Internet of Things and Big Data (Vol. 2, pp. 307-314). SCITEPRESS.

[15] Dasoriya, R. (2017, October). A review of big data analytics over cloud. In 2017 IEEE international conference on consumer electronics-Asia (ICCE-Asia) (pp. 1-6). IEEE.

[16] Rani, S., Bhambri, P., & Kataria, A. (2023). Integration of IoT, Big Data, and Cloud Computing Technologies: Trend of the Era. In Big Data, Cloud Computing and IoT (pp. 1-21). Chapman and Hall/CRC.