ISSN No:-2456-2165

Role of Cloud Computing for Big Data

Farhan Aslam
University of the Cumberlands, KY, USA

Abstract:- Cloud computing has revolutionized the way businesses process, store and manage data. With the explosion of big data, cloud computing has become an integral tool in providing scalable and cost-effective solutions for handling large volumes of data. In this paper, we will explore the role of cloud computing in big data and its impact on businesses. We will start by defining cloud computing and its relation to big data. We will also examine the benefits of using cloud computing for big data processing and storage, as well as the challenges that come with implementing cloud computing for big data. Additionally, we will discuss the different cloud computing architecture models used in big data processing and storage and compare them to traditional on-premise computing architecture. We will analyze the advantages and disadvantages of each cloud computing architecture model for big data. Finally, we will explore the common applications of cloud computing for big data and the trends and future developments in cloud computing for big data, including the potential challenges and opportunities for using cloud computing for big data in the future.

Keywords:- *Cloud Computing, Big Data, Data processing, sql, query, Hadoop.*

I. INTRODUCTION TO CLOUD COMPUTING FOR BIG DATA

- *What is cloud computing, and how does it relate to big data?*

Cloud computing has become a critical technology for big data analytics because it enables scalable analysis of enormous amounts of data [1]. Cloud computing provides Platform as a Service (PaaS) to support querying and processing of big data, making it cost-effective and efficient to store, process, and analyze large data sets [1]. It is a distributed IT architecture where client data is processed at the periphery of the network, eliminating the need to maintain expensive computing hardware, dedicated space, and software [2][3]. The cloud computing paradigm delivers computing services over the Internet, providing flexible resources, faster innovation, and economies of scale [1]. Cloud computing has revolutionized the way computing infrastructure is abstracted and used. It has extended to include anything that can be considered as a service [1]. Service providers such as Amazon, Google, and Microsoft offer their own big data systems in a cost-efficient manner [1]. BigQuery is a Platform as a Service that supports querying using ANSI SQL and is a serverless data warehouse in GCP that enables scalable analysis over petabytes of data. Google Cloud Platform provides services for analyzing and processing big data, with built-in machine learning capabilities [1]. However, the required resources to manage big data in a cloud-IoT environment are still a big challenge [2].

Research has shown that cloud computing can significantly improve healthcare services and contribute to continuous and systematic innovation in a big data environment. A proposed model aims to optimize virtual machine selection in cloud-IoT health services applications to manage a large amount of data in integrated Industry 4.0 [2]. Cloud computing is also related to big data technologies, as it can be used to handle big data sets in biology and other fields [4]. Adopting a Sustainable Computational Cloud technology could provide an extra boost and competitiveness, offering many benefits to people in general, businesses, and Small and Medium Enterprises [2].

- *What are the benefits of using cloud computing for big data processing and storage?*

Cloud computing has become an increasingly popular solution for big data processing and storage, offering numerous benefits to businesses. One key advantage is the ability to avoid the need for costly computing hardware, dedicated space and software, as cloud computing provides hardware and software services through the Internet [5][2]. Cloud computing can also provide secure big data processing and storage, using tools such as Hadoop, which offers security for big data [5][21]. In addition to cost savings, cloud computing can enable big data to be controlled and distributed efficiently, reducing product return costs by predicting the likelihood of product returns [5][1]. Another significant advantage of using cloud computing for big data processing and storage is the availability of big data tools like Hadoop or Spark, which offers cost advantages for storing, processing and analyzing large amounts of data [1].

ISSN No:-2456-2165

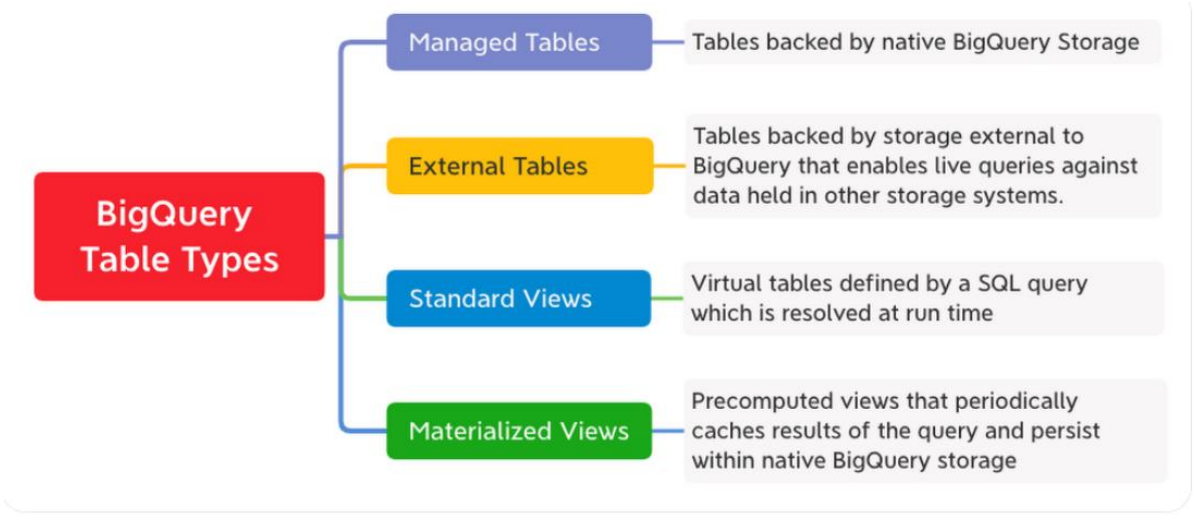

Fig 1.: Query Types for Big Data

Moreover, cloud computing provides various capabilities for storing and processing data in third-party data centers, such as BigQuery. BigQuery can help to reduce the load on relational databases, and offers pay-as-you-go services, where charges are made based on usage. Alternatively, BigQuery also provides a flat rate service with specific slot rates and charges in daily, monthly or yearly plans. In addition to offering different options and configurations to improve query performance, BigQuery is also suitable for scenarios where data does not change often as it has built-in cache [2][1]. Overall, cloud computing can provide cost-effective solutions for big data processing and storage, enabling businesses to leverage its benefits for better decision-making and more efficient operations[21].

- *What are the challenges of implementing cloud computing for big data?*

Despite the benefits of cloud computing for big data, there are still some practical challenges that need to be addressed. One major challenge is the lack of suitable tools and adequate computing resources required for implementing cloud computing for big data. This can be attributed to the fact that big data entails massive amounts of information, which makes it difficult to manage and process without the necessary resources and tools. Another critical challenge of implementing cloud computing for big data is the issue of security. Although Hadoop provides security to big data, there is still a risk of data breaches and unauthorized access. To address this challenge, symmetric key encryption algorithms such as AES are commonly used in cloud computing to ensure secure data transmission and storage. Hence, while cloud computing offers many advantages for managing big data, there are still several challenges that need to be overcome to ensure its effective implementation.

II. CLOUD COMPUTING ARCHITECTURE FOR BIG DATA

- *What are the different cloud computing architecture models used in big data processing and storage?*

Cloud computing systems have emerged as a crucial technology in managing big data in business intelligence models [6]. In particular, Cloud Computing Architecture has been proposed as a viable solution for dealing with the processing and storage challenges associated with big data [6]. However, there are different cloud service layers involved in big data processing and storage [6]. While the text does not delve into the details of various cloud computing architecture models used in big data processing and storage, it highlights Fog Computing as a new computing paradigm that extends Cloud Computing to the edge of the network. The hierarchical distributed architecture of Fog Computing is presented as a way to integrate massive numbers of infrastructure components and services in smart cities [7]. Despite Fog Computing's potential in this regard, CC remains the baseline technology for big data processing and storage. Therefore, it is important to understand the different cloud computing architecture models that are currently being utilized to provide effective solutions for big data processing and storage needs. Unfortunately, the text does not provide any information on the different cloud computing architecture models that are used in this context [2].

- *How does the cloud computing architecture differ from traditional on-premise computing architecture?*

One of the primary differences between cloud computing architecture and traditional on-premise computing architecture is the handling of real-time data processing. Cloud computing architecture offers real-time event processing, while traditional on-premise computing architecture may not have the processing power to handle real-time data processing [2]. Additionally, cloud computing architecture allows for authorized external users to access real-time data directly from IoT sensors, while traditional on-premise computing architecture may require access through internal networks, making it less accessible and more limited in its reach [2]. In cloud-driven IoT-based big

data environments, the cloud-based platform can be considered as a big data warehouse, while in traditional on-premise computing architecture, data is typically stored in traditional databases [2]. Moreover, in cloud computing architecture, a cloud-based platform is used to store data generated by IoT devices, while in traditional on-premise computing architecture, data is stored on physical servers located on the premises of the organization [2]. Another significant difference is scalability. Cloud computing architecture is highly scalable, while traditional on-premise computing architecture may require additional physical servers to accommodate increased data storage needs, which can be costly and time-consuming [2]. These differences highlight the advantages of cloud computing architecture in terms of accessibility, scalability and real-time data processing capabilities.

- *What are the advantages and disadvantages of each cloud computing architecture model for big data?*

Cloud computing offers numerous advantages and disadvantages when it comes to big data architecture. One of the primary advantages is that cloud computing provides a solution for big data through cloud services and open-source cloud software tools that make handling large amounts of data more manageable [6]. However, the hardware and software requirements of big data can be cost-prohibitive for small and medium-sized businesses [8]. There are different types of cloud computing architectures that can support big data integration, including distributed cloud, smart grid, and hybrid cloud [9]. A cloud-based architecture can also be used with big data to offer greater control over data distribution and management. The MetaCloudDataStorage security architecture is one example of a proposed solution to protect big data against intruders, in which data is stored in multiple cloud data centers [10]. Cloud computing platforms, databases, and storage solutions are also essential components of big data processing [11].

As big data tools continue to advance, cloud services are increasingly being used to implement big data applications. For example, the integration of services and cloud computing architectures has been proposed as a solution to handle big data in a digital environment [12]. However, there are also disadvantages to using cloud computing for big data, including concerns about security and privacy. Therefore, organizations must weigh the advantages and disadvantages of each cloud computing architecture model carefully when implementing big data solutions using cloud services [13][14].

III. APPLICATIONS AND FUTURE OF CLOUD COMPUTING FOR BIG DATA

- *What are the common applications of cloud computing for big data?*

Cloud computing has become increasingly important in the big data domain, offering a variety of applications and services for managing and processing large volumes of data. The integration of cloud computing with Industry 4.0 has allowed for precise and fast analysis of large data sets, which is critical for effective decision-making. In healthcare applications, cloud computing and IoT have the potential to significantly improve healthcare services by providing continuous and systematic innovation in a big data environment. Cloud-based infrastructure is commonly used as a backbone for big data storage and analytics, which can extract beneficial information to provide better insights and guide better decisions across various fields, including business, healthcare, sciences, and engineering [15][16][15]. Cloud computing provides the necessary computation, storage, applications, and networking required to support big data applications and task scheduling in cloud computing environments has gained considerable attention [16]. Additionally, scalable database management systems are critical for cloud infrastructure to support update-intensive application workloads and decision support systems for descriptive and deep analytics [17].

However, the use of cloud computing for big data also raises concerns regarding security and privacy, particularly regarding access control, authentication, and authorization requirements [15]. Overall, cloud computing is one of the common applications for big data processing and management, offering novel ideas and research outcomes on all aspects of Big Data, IoT and cloud computing [15].

- *What are the trends and future developments in cloud computing for big data?*

Cloud computing is a valuable tool for big data storage and analysis, but it does have its limitations. Thus, research is currently being conducted to examine machine learning-based offloading from the Internet of Things (IoT) to fog and fog to cloud for big data processing [16]. Fog computing has been found to have lower latency than cloud computing, which has significant implications for processing big data tasks [16]. Data processing for big data occurs at two levels: fog and cloud, with machine learning-based methods being used in both cases [16]. However, the need to monitor, analyze, and act upon these large volumes of data brings with it various issues related to data confidentiality, verification, authorization, mining, secure communication, and computation [15]. With the increasing attention given to the IoT in recent years, there has been a trend towards IoT generating enormous amounts of new types of data [15]. The IoT enables scalability, flexibility, agility, and ubiquity in fields of massive scale multimedia data processing, storage, access, and communications, which opens up exciting new possibilities for the future of cloud computing and big data analysis [15]. However, no specific information is provided about future developments in cloud computing for big data in this article.

- *What are the potential challenges and opportunities for using cloud computing for big data in the future?*

As the amount of data being generated continues to increase at an unprecedented pace, so do the challenges and opportunities for using cloud computing to manage and analyze big data. Efficiently managing workflow applications on cloud computing data centers is challenging due to the need to perform calculations on sensitive data [18]. However, cloud computing provides an opportunity for big data to be controlled and distributed efficiently, enabling

cost advantages for storing, processing and analyzing large amounts of data [19]. With the advent of internet of things (IoT) technology, such an application has all features of the 5Vs for Big Data and can be addressed by cloud computing [13]. To improve performance of big data applications and frameworks, researchers have proposed methods such as job scheduling and resource allocation algorithms [15]. Furthermore, cloud computing can facilitate innovation in healthcare services and contribute to its continuous and systematic innovation [16]. However, with the plethora of systems available and the varying application requirements, research challenges still exist in data management for the cloud [17]. Thus, it is important to carefully consider which options for cloud processing are suitable for a given application, with the goal of optimizing the potential of cloud computing as it relates to processing of Big Data [20].

IV. CONCLUSION

The role of cloud computing in big data management has become increasingly important in recent years. The scalability and cost-effectiveness of cloud computing has made it a critical technology for big data analytics, enabling the processing and analysis of enormous amounts of data. Cloud computing offers various benefits to businesses, including the flexibility of resources, faster innovation, and economies of scale. The use of cloud computing services such as BigQuery, Hadoop, and Spark has become increasingly popular in implementing big data applications. The cloud-based platform can be considered as a big data warehouse, while in traditional on-premise computing architecture, data is typically stored in traditional databases.

However, one of the significant challenges of implementing cloud computing for big data is the issue of security, which must be addressed to ensure safe and secure data processing and storage. Efficiently managing workflow applications on cloud computing data centers is also challenging due to the need to perform calculations on sensitive data. Future research should focus on addressing these challenges and exploring the potential of cloud computing for big data analytics in various fields, including healthcare, biology, and other industries.

Despite the numerous advantages of cloud computing, there are still gaps in our understanding of how it can be used to manage and analyze big data efficiently. Therefore, further research is needed to explore the full potential of cloud computing in big data management and to develop more effective and secure cloud-based solutions for big data analytics. Overall, the use of cloud computing for big data management is a rapidly evolving field, and future developments are likely to continue to revolutionize the way we analyze and manage big data.

REFERENCES

[1]. Berisha, B., Mëziu, E., Shabani, I. Big data analytics in Cloud computing: an overview. (n.d.) Retrieved August 11, 2023, from link.springer.com/article/10.1186/s13677-022-00301-w
[2]. Stergiou, C., Psannis, K., Gupta, B. Security, privacy & efficiency of sustainable cloud computing for big data & IoT. (n.d.) Retrieved August 11, 2023, from www.sciencedirect.com/science/article/pii/S2210537918300490
[3]. Hashem, I., Yaqoob, I., Anuar, N., Mokhtar, S., Gani, A. The rise of "big data" on cloud computing: Review and open research issues. (n.d.) Retrieved August 11, 2023, from www.sciencedirect.com/science/article/pii/S0306437914001288
[4]. O'Driscoll, A., Daugelaite, J., Sleator, R. Methodological Review 'Big data', Hadoop and cloud computing in genomics. (n.d.) Retrieved August 11, 2023, from www.sciencedirect.com/science/article/pii/S1532046413001007
[5]. Das, M., Dash, R. Role of Cloud Computing for Big Data: A Review. (n.d.) Retrieved August 11, 2023, from link.springer.com/chapter/10.1007/978-981-15-6202-0_18
[6]. Bahrami, M., Singhal, M. The role of cloud computing architecture in big data. (n.d.) Retrieved August 11, 2023, from link.springer.com/chapter/10.1007/978-3-319-08254-7_13
[7]. Tang, B., Chen, Z., Hefferman, G., Wei, T., He, H. A hierarchical distributed fog computing architecture for big data analysis in smart cities. (n.d.) Retrieved August 11, 2023, from dl.acm.org/doi/abs/10.1145/2818869.2818898
[8]. Purcell, B. [PDF][PDF] Big data using cloud computing. (n.d.) Retrieved August 11, 2023, from citeseerx.ist.psu.edu
[9]. Mayilvaganan, M., Sabitha, M. A cloud-based architecture for Big-Data analytics in smart grid: A proposal. (n.d.) Retrieved August 11, 2023, from ieeexplore.ieee.org/abstract/document/6724168/
[10]. Manogaran, G., Thota, C., Kumar, M. MetaCloudDataStorage architecture for big data security in cloud computing. (n.d.) Retrieved August 11, 2023, from www.sciencedirect.com/science/article/pii/S187705091630477X
[11]. Ji, C., Li, Y., Qiu, W., Awada, U., Li, K. Big data processing in cloud computing environments. (n.d.) Retrieved August 11, 2023, from ieeexplore.ieee.org/abstract/document/6428800/
[12]. Zimmermann, A., Pretz, M., Zimmermann, G. Towards service-oriented enterprise architectures for big data applications in the cloud. (n.d.) Retrieved August 11, 2023, from ieeexplore.ieee.org/abstract/document/6690543/

[13]. Yang, C., Huang, Q., Li, Z., Liu, K., Hu, F. Big Data and cloud computing: innovation opportunities and challenges. (n.d.) Retrieved August 11, 2023, from www.tandfonline.com/doi/abs/10.1080/17538947.2016.1239771
[14]. Schmidt, R., Möhring, M. Strategic alignment of cloud-based architectures for big data. (n.d.) Retrieved August 11, 2023, from ieeexplore.ieee.org/abstract/document/6690544/
[15]. Kobusińska, A., Leung, C., Hsu, C. Emerging trends, issues and challenges in Internet of Things, Big Data and cloud computing. (n.d.) Retrieved August 11, 2023, from www.sciencedirect.com/science/article/pii/S0167739X18311270
[16]. Elhoseny, M., Abdelaziz, A., Salama, A., Riad, A. A hybrid model of internet of things and cloud computing to manage big data in health services applications. (n.d.) Retrieved August 11, 2023, from www.sciencedirect.com/science/article/pii/S0167739X17322021
[17]. Agrawal, D., Das, S., El Abbadi, A. Big data and cloud computing: current state and future opportunities. (n.d.) Retrieved August 11, 2023, from dl.acm.org/doi/abs/10.1145/1951365.1951432
[18]. Chen, X., Metawa, N. Enterprise financial management information system based on cloud computing in big data environment. (n.d.) Retrieved August 11, 2023, from content.iospress.com
[19]. Gupta, B., Agrawal, D. [BOOK][B] Handbook of research on cloud computing and big data applications in IoT. (n.d.) Retrieved August 11, 2023, from books.google.com
[20]. Branch, R., Tjeerdsma, H., Wilson, C., Hurley, R. [HTML][HTML] Cloud computing and big data: a review of current service models and hardware perspectives. (n.d.) Retrieved August 11, 2023, from www.scirp.org/html/7-9301899_48106.htm
[21]. Aslam, F. . (2023). The Benefits and Challenges of Customization within SaaS Cloud Solutions. American Journal of Data, Information and Knowledge Management, 4(1), 14 - 22. https://doi.org/10.47672/ajdikm.1543