



DATA ANALYTICS IN CLOUD COMPUTING

Dr. AparnaRajesh A, Dr. Amiya Sahoo, Dr. Kiran G M, Department of Computer Science & Engineering, Aryan Institute of Engineering & Technology, Bhubaneswar

ABSTRACT: The exponential growth of big data presents a multitude of challenges in managing the data lifecycle, encompassing storage, transportation, processing, mining, and serving of data. The diverse nature of big data, originating from various sources, includes vast volumes, high velocity streams, and intricate complexities. Traditional data processing software struggles to handle the processing of such data effectively. Addressing these challenges requires substantial computational resources for data analysis to ensure efficient processing and analysis. Cloud computing emerges as a pivotal technology for handling extensive and intricate computations. By offering a dependable, fault-tolerant, and scalable environment, cloud computing provides infrastructure and software resources as services for distributed big data management systems. Consequently, it obviates the necessity for maintaining costly computing hardware, dedicated physical space, and software resources.

KEYWORDS: Big Data, Cloud Computing, Big data Analytics

I. INTRODUCTION

Big Data and Cloud Computing are key technologies in the IT industry. Big data refers to large, complex datasets used for decision-making in business and engineering, offering valuable insights. While big data poses processing challenges, cloud computing has significantly advanced its capabilities by providing computational, networking, and storage resources. Cloud computing emerged concurrently to support big data challenges, with public cloud services offering analytics. The rise of Software as a Service (SaaS) underscores the importance of adhering to cloud infrastructure best practices and managing vast data types effectively. It is crucial to differentiate between cloud computing and big data, understand their symbiotic relationship, and recognize how they complement each other, fostering innovation like artificial intelligence.

II. BASICS OF CLOUD COMPUTING

Cloud computing is a revolutionary technology that enables users to access and utilize computing resources over the Internet. It offers scalability, flexibility, cost-effectiveness, and security for managing IT operations. Cloud computing operates on a pay-as-you-go model, providing on-demand availability of resources without direct user management. Examples of cloud computing services include Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). Cloud deployment models include public, private, hybrid, and community clouds, each with distinct advantages. The cloud allows users to access data and applications from anywhere, eliminating the need for physical proximity to data storage devices. While cloud computing enhances accessibility and ease of use, it also raises concerns about data security and privacy. Understanding the different types of cloud services, deployment models, and the benefits of cloud computing is essential for individuals and organizations looking to leverage this technology for efficient and cost-effective IT solutions.

III. COMPARISON OF BIG DATA & CLOUD COMPUTING

Big Data and Cloud Computing are two essential technologies that often work in tandem to provide innovative solutions across various sectors. Big Data involves managing and analyzing large, complex datasets, focusing on the volume, velocity, and variety of data. It enables organizations to extract valuable insights for decision-making processes. On the other hand, Cloud Computing offers on-demand access to computing resources over the internet, providing scalability, flexibility, and cost-



effectiveness. It includes service models like Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

When comparing Big Data and Cloud Computing:

- Focus: Big Data concentrates on processing vast amounts of data to derive patterns for decision-making, while Cloud Computing emphasizes providing insights about data to enhance decision-making processes.
- Usage: Big Data stores and analyzes large datasets, while Cloud Computing processes data efficiently for various applications.
- Deployment: Big Data can utilize Cloud Computing resources for scalability and fault tolerance, enhancing data processing capabilities.

The synergy between Big Data and Cloud Computing enables organizations to harness the power of data for informed decision-making in a scalable and cost-effective manner. By leveraging Cloud Computing infrastructure, businesses can efficiently store, process, and analyze Big Data, leading to improved operational efficiency and strategic insights. This collaboration between Big Data analytics and Cloud Computing technologies opens up new possibilities for businesses across different industries by enabling them to manage and derive value from large datasets effectively.

IV. UNIFIED APPROACH OF CLOUD COMPUTING FOR BIG DATA

The unified approach of cloud computing for big data involves leveraging cloud infrastructure to enhance the storage, processing, and analysis of large datasets efficiently[1-5]. Cloud computing provides on-demand access to computing resources over the internet, offering scalability, flexibility, and cost-effectiveness. By utilizing cloud services like Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS), organizations can store, process, and analyze big data effectively.

Cloud computing enables organizations to:

- Store Data: Large volumes of data can be stored remotely on cloud servers.
- Process Data: Cloud computing facilitates the processing of vast amounts of data for decision-making.
- Analyze Data: By harnessing cloud resources, organizations can derive insights from big data analytics.

The unified approach of cloud computing for big data involves integrating cloud services with big data analytics tools to optimize operations, gain valuable insights, and predict future outcomes. This synergy allows for real-time processing of big data, enabling organizations to make informed decisions based on the analysis of extensive datasets[6-7]. By combining the capabilities of cloud computing with big data analytics, businesses can unlock the full potential of their data assets and drive innovation across various industries.

1. Cloud Storage for Big Data Storage

Cloud storage plays a crucial role in managing and working with big data effectively. It provides a scalable and cost-efficient solution for storing, processing, and analyzing large datasets. Here are some key points based on the provided sources regarding cloud storage for big data storage:

- Google Cloud Storage: Google Cloud Storage is an integral part of storing and working with big data on Google Cloud. It allows for loading data into BigQuery, holding staging files for Dataflow pipelines, and integrating with Dataproc for running Apache Hadoop or Apache Spark jobs directly on data stored in Cloud Storage[7].
- Best Cloud Storage Options:
 - Sync.com: Known for private encryption, no file size limits, and unlimited storage with its Pro Teams plan, making it ideal for large files[8].
 - MASV: Offers direct ingest integrations with major cloud storage providers, allowing for sending and receiving up to 15 TB per file for virtually unlimited uploads[9].



- Considerations[10]:
 - File Size: Ensure the cloud storage provider supports the file sizes you work with, especially for large files like 4K videos or business project files exceeding 50GB.
 - Transfer Speeds: Look for providers that offer faster transfer speeds to handle large data efficiently.
 - Security: Prioritize providers with robust security features like encryption and access control to protect your data.

2. Big data processing by cloud computing involves utilizing cloud infrastructure to store, process, and analyze large and complex datasets efficiently. Cloud computing offers on-demand access to computing resources over the internet, providing scalability, flexibility, and cost-effectiveness for handling big data. By utilizing cloud services like AaaS and deploying big data in public or private clouds, organizations can effectively manage and derive valuable insights from their big data initiatives while ensuring scalability, security, and accessibility.

Here are some key points from the provided sources regarding big data processing by cloud computing:

- **Cloud Computing Benefits:** Cloud computing revolutionizes how computing infrastructure is abstracted and used, offering benefits like elasticity, pay-as-you-go models, low upfront investment, and scalability for big data storage, management, and analytics[11].
- **Analytics as a Service (AaaS):** AaaS is a common cloud computing platform for big data processing and analysis. It provides analytics software and procedures that help organize, analyze, and present big data efficiently for informed decision-making[12].
- **Relationship with IoT:** Cloud computing plays a crucial role in handling the large volume of real-time data collected by Internet of Things (IoT) devices. IoT devices communicate via the internet without human intervention, generating vast amounts of data that require substantial storage space and computational capacity provided by cloud computing[12].
- **Big Data Deployment Options:**
 - **Public Cloud:** Multiple organizations share a public cloud environment where each company's data is isolated. It is a common choice for storing and analyzing big data.
 - **Private Cloud:** In a private cloud model, computing resources are dedicated to a single customer, offering more control over data security and management.

3. Big Data Analytics by Cloud Computing

Big Data Analytics in Cloud Computing is a pivotal process that harnesses the capabilities of cloud infrastructure to efficiently store, process, and analyze vast and intricate datasets. Cloud Computing, exemplified by giants like Amazon and Google, offers a scalable and cost-effective solution for managing Big Data across various sectors such as healthcare, education, and businesses. The synergy between Big Data and Cloud Computing addresses the challenges posed by the sheer volume, velocity, and variety of data generated daily, which traditional tools struggle to handle. Big Data Analytics, a critical component in numerous fields, involves an analytics cycle that extracts valuable insights from data, leading to operational efficiency improvements, cost optimization, and innovation [13-15]. By transitioning from the traditional ETL (Extract, Transform, Load) to the ELT (Extract, Load, Transform) paradigm, organizations can leverage Big Data analytics in the Cloud to derive actionable insights for informed decision-making.

Google's BigQuery, a fully-managed, serverless data warehouse, serves as a prominent case study in this domain, enabling scalable analysis over petabytes of data. The platform supports ANSI SQL queries and boasts built-in machine learning capabilities, making it a popular choice for data analytics among large enterprises[16]. The amalgamation of Big Data Analytics and Cloud Computing empowers organizations to efficiently manage and derive valuable insights from large datasets, leading to cost savings, operational enhancements, innovation, and data-driven decision-making across diverse industries. This symbiotic relationship offers scalable and flexible solutions for handling and analyzing Big Data initiatives effectively.



4. Big Data Sharing and Remote Collaboration by Cloud Computing

Big Data Sharing and Remote Collaboration facilitated by Cloud Computing revolutionize data management and collaboration across various sectors. Cloud Computing, offering scalable and cost-effective solutions, enables real-time collaboration without geographical constraints. The convergence of Big Data and Cloud Computing optimizes data sharing and analysis, enhancing operational efficiency and innovation. Traditional deployment of Big Data systems often involves complex settings and lacks efficient data sharing mechanisms. Cloud Computing, through virtualization technology, eliminates hardware redundancy and optimizes computing resources, enabling distributed analytical platforms to migrate to virtual clusters seamlessly. Autoscaling and load balancing in the cloud support parallel algorithms, ensuring scalability and efficient resource utilization for Big Data analytics [17-19]. Cloud Computing enhances data sharing by applying modern analytical tools, managing controlled access, and enabling remote collaboration. Virtualization facilitates secure data sharing among different parties, fostering interdisciplinary collaborations and advanced workflows. Despite challenges like domain boundaries hindering workflow sharing, Cloud Computing's flexibility allows for data and computational resource sharing, creating a conducive environment for diverse businesses and applications.

The synergy between Big Data and Cloud Computing offers scalable, cost-effective, and efficient solutions for managing and analyzing large datasets. This collaboration empowers organizations to make data-driven decisions, drive innovation, and improve operational processes across industries. By leveraging the capabilities of Cloud Computing for Big Data sharing and remote collaboration, businesses can build a collaborative environment that supports diverse goals and objectives, enhancing productivity and decision-making processes.

V. RELATIONSHIP BETWEEN BIG DATA & CLOUD COMPUTING

The relationship between Big Data and Cloud Computing is symbiotic, with cloud computing enabling Big Data analytics to occur in a fraction of the time it would take on traditional computing resources. Cloud computing offers the infrastructure and resources necessary for processing and analyzing large datasets, which is essential for Big Data analytics. The cloud infrastructure allows for real-time processing of Big Data, which is crucial for gaining insights from the data in a timely manner. Cloud computing providers often utilize a Software as a Service (SaaS) model, which allows customers to easily process data through a console or the site's user interface. This model offers a range of services, including database management systems, cloud-based virtual machines and containers, identity management systems, and machine learning capabilities, all of which are essential for Big Data analytics [20].

Moreover, cloud computing services are largely driven by Big Data, as they provide the necessary infrastructure for storing, processing, and analyzing large datasets. In turn, Big Data is often collected by cloud-based applications, which would not be possible without the cloud infrastructure. Cloud computing has revolutionized the way computing infrastructure is abstracted and used, offering benefits such as elasticity, pay-as-you-go or pay-per-use models, and low upfront investment [21-25]. These features make cloud computing a viable and desirable choice for Big Data storage, management, and analytics [26]. The relationship between Big Data and Cloud Computing is essential for the efficient processing and analysis of large datasets. Cloud computing provides the necessary infrastructure and resources for Big Data analytics, while Big Data drives the demand for cloud computing services. The combination of the two allows for the development of innovative technologies such as artificial intelligence, which can further enhance the capabilities of both Big Data and Cloud Computing.

VI. CONCLUSION

In conclusion, the integration of Big Data analytics in Cloud Computing offers a transformative approach to data management and analysis, revolutionizing how organizations handle and derive



insights from large datasets. Cloud Computing provides the necessary infrastructure and resources for processing and analyzing Big Data, enabling scalability, cost-effectiveness, and real-time processing capabilities. The benefits of this synergy include enhanced data governance, scalability, cost-effectiveness, agility, accessibility, and security. However, organizations must also address challenges such as data governance complexity, vendor lock-in risks, latency issues, complexity in implementation, and potential cost overruns. By implementing best practices like establishing clear data governance policies, avoiding vendor lock-in, monitoring performance, building expertise, and managing costs effectively, organizations can harness the full potential of Big Data analytics in the cloud. This approach allows businesses to leverage cloud-based analytics solutions to gain valuable insights, drive innovation, and stay competitive in today's data-driven world. Ultimately, the collaboration between Big Data analytics and Cloud Computing presents a powerful solution for organizations seeking to unlock the value of their data, make informed decisions, and propel growth in the digital era.

REFERENCES

- [1]. Qi, Lianyong & Khosravi, Mohammad & Xu, Xiaolong & Zhang, Yiwen & Menon, Varun (2021), Cloud Computing. 10.1007/978-3-030-69992-5.
- [2]. T.Sri Harsha, "Big Data Analytics in Cloud Computing Environment", International Journal of Scientific & Engineering Research Volume 8, Issue 8, August-2017 ISSN 2229-5518
- [3]. Blend Berisha & Endrit M'eziu, "Big Data Analytics in Cloud Computing: An overview", A seminar paper in the subject "Cloud Computing", February 2021
- [4]. Big data and cloud computing – challenges and opportunities: <https://bigdata-madesimple.com/big-dataand-cloud-computing-challenges-and-opportunities/>
- [5]. F.C.P, Muhtaroglu, Demir S, Obali M, and Girgin C. "Business model canvas perspective on big data applications." Big Data, 2013 IEEE International Conference, Silicon Valley, CA, Oct 6-9, 2013.
- [6]. Chaowei Yang, Qunying Huang, Zhenlong Li, Kai Liu & Fei Hu (2017) Big Data and cloud computing: innovation opportunities and challenges, International Journal of Digital Earth,10:1, 13-53, DOI: 10.1080/17538947.2016.1239771
- [7]. Samir A. El-Seoud, Hosam F. El-Sofany, Mohamed Abdelfattah & Reham Mohamed "Big Data and Cloud Computing: Trends and Challenges", International Journal of Interactive Mobile Technologies (IJIM) · April 2017
- [8]. The Evolution of Cloud Adoption in Big Data: <https://www.coforge.com/salesforce/blog/dataanalytics/the-evolution-of-cloud-adoption-in-big-data/>
- [9]. Majhi, S.K. & Shial, G., 2015. Challenges in Big Data Cloud Computing And Future Research Prospects: A Review. The Smart Computing Review, 5(4), pp.340–345.
- [10]. Neelay Jagani, Parthil Jagani & Suril Shah, "Big Data In Cloud Computing: A Literature Review", Vol. 5, Issue 11, ISSN No. 2455-2143, Pages 185-191 Published Online March 2021 in IJEAST (<http://www.ijeast.com>)
- [11]. Brevini, Benedetta. (2015). Book Review: To the Cloud: Big Data in a Turbulent World. Media, Culture & Society. 37. 1111-1113. 10.1177/0163443715596318a.
- [12]. Marchiori, Massimo. (2017). Learning the way to the cloud: Big Data Park. Concurrency and Computation: Practice and Experience. 31. 10.1002/cpe.4234.
- [13]. Wang, Lizhe & von Laszewski, Gregor & Younge, Andrew & He, Xi & Kunze, Marcel & Tao, Jie & Fu, Cheng. (2010). Cloud Computing: A Perspective Study. New Generation Comput..28. 137-146. 10.1007/s00354-008-0081-5.
- [14]. Ibrahim Abaker Targio Hashem, Nor Badrul Anuar, Salimah Mokhtar (2014) Information Systems 47:98- 115 "The rise of the "Big Data" on cloud computing: Review an open research issues". DOI:10.1016/j.is.2014.07.006
- [15]. Book-Big Data and Cloud Computing: https://link.springer.com/chapter/10.1007/978-981-32-9915-3_9



- [16]. Yadav S., Sohal A. (2017) “Review Paper on Big Data Analytics in Cloud Computing” in International Journal of Computer Trends and Technology (IJCTT) V49(3):156-160, July 2017. ISSN:2231-2803.
- [17]. Alsghaier, Hiba & Al-Shawakfa, Emad. (2018). An empirical study of cloud computing and big data analytics. International Journal of Innovative Computing and Applications. 9. 180. 10.1504/IJICA.2018.10014870.
- [18]. M. Ramalingam, P. Prabhusundhar & V. Thiagarasu, “Biometric Based Intrusion Detection System using Dempster-Shafer Theory for Mobile Ad hoc Network Security”, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169, Volume: 5 Issue: 7 pp 384 – 391.
- [19]. Manoj Muniswamaiah, Tilak Agerwala & Charles Tappert, “Big Data In Cloud Computing Review And Opportunities”, International Journal of Computer Science & Information Technology (IJCSIT) Vol 11, No 4, August 2019.
- [20]. Bernice M. Purcell, “Big data using cloud computing”, Journal Technology Research, January 2013.
- [21]. Nabeel Zanoon, Abdullah Al-Haj & Sufian M Khwaldeh, "Cloud Computing and Big Data is there a Relation between the Two: A Study", International Journal of Applied Engineering Research ISSN 0973- 4562 Volume 12, Number 17 (2017) pp. 6970-6982
- [22]. P.Prabhusundhar & B. Srinivasan, “Detection Mechanism Against the Attacks in MANET and Biometric Classification”, International Journal of Computer Science and Engineering, ISSN No. 2347 – 2693, Vol. 6, No. 5, May – 2018, pp. 986 – 990.
- [23]. Affreen Ara & Aftab Ara, "Cloud for Big Data Analytics Trends", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 5, Ver. IV (Sep. - Oct. 2016), PP 01-06
- [24]. Dr. Harjinder Kaur & Dr. Major Singh Goraya, "Role of Big Data in Cloud Computing: A Review", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 8 Issue 07, July-2019
- [25]. Venkatesh H, Shrivatsa D Perur & Nivedita Jalihal, "A Study on Use of Big Data in Cloud Computing Environment", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3), 2015, 2076-2078
- [26]. Ramalingam M, Prabhusundhar P, Azhaharasan V & Prabhahari R, “Distributed Storage Architecture and Hadoop Framework Tools for Crunching Big Data in Clustered Processing System”, Test Engineering & Management (SCOPUS Journal), ISSN: 0193 -4120 Page No. 4659-4664.