Abhendra Pratap Singh, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology
An Open Access Journal

A Review on Big Data Analytics Server Machines for Cloud Computing

Abhendra Pratap Singh[1], Kanika[2], Nehal Sharma[3], Nandini Sharma[4]
[1]Department of Mechanical Engineering, HMRITM, Hamidpur, New Delhi, 110036, Delhi, India
[2,3]Department of Computer Science & Engineering and Data Science, HMRITM, Hamidpur, New Delhi, 110036, Delhi, India
[4]Department of Computer Science and Engineering, HMRITM, Hamidpur, New Delhi, 110036, Delhi, India

Abstract- In today's era, data is the most important element, and data generation is increasing day by day. Due to this increase in data, traditional servers cannot handle large amounts of information efficiently. Therefore, the present generation requires more advanced server infrastructure. In traditional systems, limited storage, low processing speed and lack of scalability are common issues. In big data systems, the 3V model (Volume, Velocity, Variety) is followed. Big data analytics server machines have high-performance processors. They also include parallel processing support, distributed architecture and cluster-based systems. The main role includes on-demand resources, virtualization, resource pooling and elastic scalability. Therefore, big data analytics server machines are a fundamental component of modern computing environments. Big data analytics plays a vital role in server machines and serves as a fundamental component of modern computing environments. This paper represents how this technology has transformed. This paper discusses big data features such as volume, velocity, and variety, along with evolving technologies such as Apache Hadoop, MapReduce, Spark, SaaS, PaaS, and IaaS. It also explains the differences between big data analytics, big data storage, and big data warehousing. Furthermore, it discusses cloud computing and software such as Kafka and its importance in cybersecurity. Security is a crucial component of cloud-based big data analytics, as it protects sensitive data. Cloud companies ensure protection through data encryption, access control, and compliance with security standards.

Keywords: Big Data, Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS).

I. INTRODUCTION

Big data can be referred to as a large and complex dataset that can be structured, unstructured, or semi-structured. It is produced at high speed from sources such as social media, sensors, and transactions, which traditional data processing software cannot manage efficiently. While some data analysts define big data as the process of extracting, transforming, and loading (ETL) enormous datasets, another well-known definition is based on three characteristics: volume, velocity, and variety (3Vs) [1]. Between 2020 and 2025, the volume of big data increased by more than 94%.

This data is digital and amounts to around 44 zettabytes (44×10^{21} bytes) every day [2]. Two types of systems can be used to explain data at rest: NoSQL systems for interactive environments that provide data, and systems for large-scale analytical processing such as Hadoop [3]. Frameworks like Apache Drill and Google's Dremel are examples of interactive analytical tools that analyze data in interactive environments and work with real-time data. S4 and Storm are the primary platforms used for information streaming [4].

A machine learning-assisted cloud computing model (ML-CCM), which may also help boost transmission rates, can enhance security by storing large amounts of data in cloud computing environments [5]. Meanwhile, big data has revolutionized numerous industries, from the healthcare sector's pursuit of improved diagnosis to the financial sector's search for data-driven strategies, providing stakeholders with operational efficiency and competitive advantages [6]. Using iSpot, a cost-effective temporary server provisioning framework designed to achieve predictable performance in the cloud by focusing on Spark as the representative directed acyclic graph (DAG)-style big data analytical workload, can help address this issue. However, deploying big data on unstable cloud transient servers may significantly harm job performance [7].

A group of independent systems that collaborate to accomplish a shared goal is called a System of Systems (SoS). In an SoS, each system functions autonomously while demonstrating emergent behavior, geographic dispersion, operational independence, managerial independence, and evolutionary growth. Due to these features, an SoS can provide unique capabilities that none of the component systems could achieve individually. Therefore, SoS is widely used to address complex systems engineering challenges in various industries, including defense, aerospace, healthcare, automotive, and IT sectors [8]. This paper discusses big data features such as volume, velocity, and variety. Along with evolving technologies, it also explores their advantages, benefits and uses.

II. LITERATURE REVIEW

Jain et al. [9] talk about how big data is an innovation to computing and has potential to structure, gather and analyze data across different sectors. Cloud computing has been very popular with the corporate world, with about 75% of Fortune 500 companies currently relying on cloud based solutions. According to IDC, nearly 80% of new enterprise applications are relying on the cloud. Hadoop clusters on unbelievably versatile and adaptable Infrastructure-as-a-Service (IaaS) platforms like Amazon Web Services and Rackspace were employed by early adopters of Big Data technologies in the cloud. These settings were primarily utilized for system development, testing, and dataset analysis. In this real time data is given to the processor which gives databases, processing and analytics this can all happen through cloud computing.

Saif et al. [10] claimed with substantial worldwide service availability, Amazon Web Services (AWS) provides a well unified platform in addition to providing advanced tools like TensorFlow for a range of machine learning applications. To make sure users are only charged for what they really use, Cloud Platform (GCP) uses a pay-per-second payment system. To make sure users are only charged for what they really use, Cloud Platform (GCP) uses a pay-per-second payment system. Microsoft Azure facilitates the deployment of Microsoft R Server-based models on large datasets and ensures a 99.99% service level agreement (SLA) for each virtual machine instance.

With a strong technical focus on meeting customer goals, IBM Cloud offers a complete solution. The features, services, applications, and use cases of these enterprise cloud platforms are all thoroughly checked in this survey. Additionally, it offers a comparative analysis that promotes innovation and the uptake of these new technologies while helping researchers, IT analysts, and business professionals in making quicker and more informed judgments. This study shows the difference between big data analytics, big data storage and big data warehouse and how it works in AWS, GCP, IBM and MS Azure as shown in Table 1.

Table.1. Representation of comparison between big data analytics, storage and warehouse [10]

S No	Framework Type and Features	AWS	GCP	IBM Cloud	MS Azure
1	Big Data Analytics	Amazon Elasticsearch Service	Google Cloud Dataproc	IBM Analytics Engine	Azure HDInsight
	Mode of Software	Open-Source	Open-Source	Open-Source	Open-Source
	Types of Data	Structured, semi-structured and unstructured	Structured, semi-structured and unstructured	Unstructured	Unstructured
	Data Sources	Amazon S3, Amazon Kinesis Firehose, and Amazon DynamoDB	Google Bigtable, Google Cloud Storage, and Google BigQuery	IBM Cloud Object Storage	Blob Storage
	Supported Operating System	CentOS, Ubuntu, and Amazon Linux	Debian 8	CentOS 7	Ubuntu 14, Ubuntu 16, and

					Windows Server 2012 R2
2	Big Data Storage	Amazon S3	Google Cloud Storage	IBM Cloud Object Storage	Azure Blob Storage
	Storage Type	Distributed Object Storage	Distributed Object Storage	Distributed Object Storage	Distributed Object Storage
	Object Types	Object	Object	Object	Append blobs, block blobs
	Deployment Unit	Bucket	Bucket	Bucket	Container
	Deployment Identifier	Globally Unique Key	Globally Unique Key	Universally Unique Identifier	Account-level Unique Key
	Storage Scaling	Yes, Automatic	Yes, Automatic	Yes, Manual	Yes, Manual
3	Big Data Warehouse	Amazon Redshift	Google BigQuery	IBM Db2 Warehouse on Cloud	Azure SQL Data Warehouse
	Deployment Unit	Cluster (EC2)	Fully Managed	Cluster	Massively Parallel Processor (MPP)
	Compute Scaling	Yes, Automatic	Yes, Automatic	Yes, Automatic	Yes, Manual
	Permissible range	1-128	NA, Fully	1-60 (1 head)	1-60 (excluding)

Goldin et al. [11] proposed a cloud-based big data processing architecture designed for real-time and non-real-time analytics using tools like Kafka, Spark, and cloud storage (S3). Figure 1 illustrates the data moving from local systems to a cloud service, where it is processed, stored, and analyzed for valuable insights. A local system (MEFOS Drive PC) generates measurement data and recommendation results from a data source. This data is sent to a service layer called MefosService.

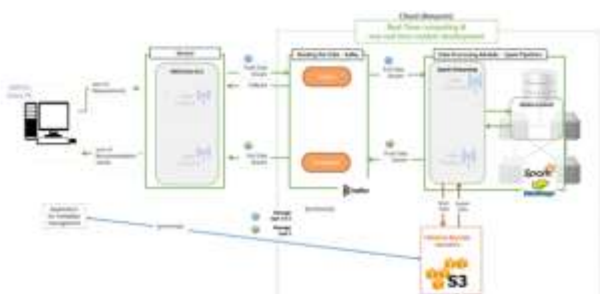

Fig.1. Simplified diagram of the proposed architecture [11]

The service layer acts as an intermediary between the local system and the cloud. It uses Kafka Producer and consumer to send data streams to the cloud and receive processed results back, which ensures smooth communication via data streams. Kafka is used as the central messaging system and handles push data streams and pull data streams, which enable real-time data transfer between components. Amazon Cloud is the core processing unit. It uses Spark Streaming, which continuously consumes data from Kafka and processes data in real time. Data can also be processed in non-real-time models, which are useful for deeper analytics and historical insights.

Kabanda et al. [12] mainly explains how big data is used for cyber security. The data can be computer-based data (logs, system activity, network traffic), mobile-based data (apps, GPS, user behavior), or physical user data (IoT devices, sensors, real-world interactions) The initial stage represents large, diverse, and unprocessed data collected from multiple sources. Big data analytics is where the raw data is processed using Data mining algorithms (to discover patterns). Further, various methods are applied for analysis such as Cyber Security methods like intrusion and malware detection, and Machine learning techniques such as vector-based models. After analysis, the system produces actionable findings that prioritize insights (what matters most) and suggest actions (what should be done next). The objective here is to extract significant patterns and identify abnormalities or dangers from enormous datasets. In this process, analysis is transformed into actionable choices for enterprises.

The forecasting capabilities, which employ predictive algorithms to foresee potential risks or trends, are highlighted at the bottom of the picture. Rather than only responding, the system assists in anticipating and averting issues before they arise.

Hwang et. [13] explain how to combine big data, cloud computing, IoT, and machine learning to build modern intelligent systems. It focuses on technologies like big data analytics, cloud computing, internet of things , machine learning and AI . It shows how these technologies work together to create smart and efficient computing systems where it uses tools like Hadoop, Spark, TensorFlow. Sharma et al. [14] claims that analytical tools and storage technologies are essential for the success of big data analytics. Nonetheless, it is evident that big data puts a strain on servers, networks, and storage, which has prompted companies and organizations to go to the cloud to take full use of the big data that is accessible. Consequently, it seems that machine learning is an ideal way of using the potential that big data offers.

Ji et al.[15] described data processing in the context of cloud computing through MapReduce optimization, popular parallel processing framework, cloud computing and storage architecture, and key applications. In order to obtain the necessary results almost instantly, it requires a distributed strategy and a scalable storage index. Big data will be complex and present throughout all major obstacles, which present users with so many opportunities. In the future, business and academia will need to address important issues. In order to ensure the long-term success of cloud computing and jointly investigate new territories. Purcell et al. [16] explained how cloud computing, small and medium-sized organizations may leverage big data technologies with minimal corporate resource commitment.

The processing capability of the big data model may provide new performance insights for the business. Innovation in business concepts, goods, and services, as well as assistance in decision-making. Saving money on processing and hardware as well as having the opportunity to test big data technology before committing a significant amount of corporate resources are advantages of using cloud computing to execute big data technology. Businesses can choose from a variety of cloud computing service models, each of which has trade-offs between the advantages of cost savings and the issues of data security and control loss.

Yassine et al. [17] present a system that integrates cloud computing, fog, IoT, and big data analytics technologies. The proposed system addresses the challenge of designing efficient solutions that are fast and can handle large volumes of unstructured smart home data. Fog computing offers quick, almost real-time analytics in this system, and computationally demanding applications are carried out using the cloud system's availability of computer and storage resources. The process involves incorporating data-in-motion from Internet of Things (IoT) sources, including individual smart meters, appliances, and devices, for incredibly intricate analytics processes that enable timely decision-making.

Lo’ai et al. [18] highlights the importance of cloud computing. They point out the increasing usage of cloud computing in the healthcare sector and discuss the reasoning for developing healthcare systems and apps. A cloud-based mobile cloud computing planning is commended for productively managing large data applications in healthcare. The study looks at many big data analytics tools, techniques, and applications linked to healthcare. In addition to requiring substantial processing and communication resources, healthcare applications typically demand dynamic access to enormous volumes of data both inside and outside of healthcare organizations.

The development of networked healthcare systems is driven by this need, which calls for the real-time analysis of unbearing data and other vital information which is something cloud and mobile cloud computing can do very well. One of the main trends for next-generation networked healthcare systems is the combination of healthcare with other smart city systems. This leads to sustainable undertaking systems that are unified into bigger municipal information and operations networks.

Ahmed et al.[19] describes big data performance using Hadoop. It includes performance testing, basic example programs, and the fundamentals of the Hadoop Distributed File System. The Hadoop MapReduce framework, key Hadoop tools, YARN applications, and how to use Apache Ambari to manage Hadoop are also explained. There are also basic Hadoop administration tasks available, and it also explains cloud computing. It describes the origins of cloud computing as well as its key ideas, aims, advantages, dangers, and difficulties. Additionally, it covers important concepts including cloud features, control, boundaries, delivery methods, and deployment models. Kune et al. [20] explained about the large amount of data being produced in many industries, big data computing has grown in importance. It eases the analysis and management of this data through the use of cutting-edge technologies, cloud systems, and innovative techniques, but it also presents storage, processing, and organization issues.

Talia et al. [21] divided cloud based data analytics into three parts as shown in the Table2. Software as a service (SaaS) for data analytics provides data extraction and application for urgent use. The main target of this service is those who are passionate about data and want to conduct research like data analysts, researchers or data managers. Platform as a service (PaaS) for data analysis provides users with a set of rules and a foundation for creating knowledgeable applications. It takes away the difficult, complex data and converts it into a simple one. Mainly, scientists, researchers, or programmers use this application for data analysis. Infrastructure service (IaaS) for data analytics uses this application for creating or manufacturing and analyzing the data. Data mining programmers, data management developers, and data mining researchers who need infrastructure control to customize analytics environments are its primary users.

Table.2. Representation of data analytics software, platform and infrastructure [21]

Cloud service model	Features	Users
Data analytics software as a service	A single and complete data mining application or task (including data sources) offered as a service	End users, analytics managers, data analysts
Data analytics platform as a service	A data analysis suite or framework for programming or developing high-level applications, hiding the cloud infrastructure and data storage	Data mining application developers, data scientists
Data analytics infrastructure as a service	A set of virtualized resources provided to a programmer or data mining researcher for developing, configuring, and running data analysis frameworks or applications	Data mining programmers, data management developers, data mining researchers

Muniswamaiah et al. [22] claimed cloud computing is mainly divided into three service models: Infrastructure as a Service, Software as a Service, and Platform as a Service. IaaS, the bottom layer, provides basic computer resources like networking, servers, and storage. Its users are connected and structured managers who supervise and preserve the digital and hardware layers. It requires digital physical resources that form the basis of higher-level services. Platform as a Service, or PaaS, is the middle layer whose main constituents are operating systems, application systems, and programming. It is mainly used by software developers who create, test, and execute software without looking at the supporting infrastructure. It enables developers to focus on coding by offering an instant platform that abstracts the complexities of infrastructure management. The top layer includes Software as a Service, or SaaS application stacks and full software programs, and related networks. It is for end users who do not need to bother about installation or upkeep in order to access and utilize software applications via the internet. It provides finished software solutions that are ready for use right away.

Aydin et al. [23] describe data classification based on different aspects, like data sources, which are web, social media, IoT, sensing devices, and business transactions. The data can be structured through content format using tables, videos, files, etc. Data type classifies the data based on its use, which can be transactional, historical, master data, or media data. Data stores tell where the data is stored. Different storage can be key-value, document, or wide column. The tools used to process the data are Hadoop, MapReduce, Spark, and Mahout. The data is analyzed in batch processing, which is used to process large data chunks. Some other processors used are real-time processing, private data center, and public cloud. The framework of big data is shown in the Figure2.

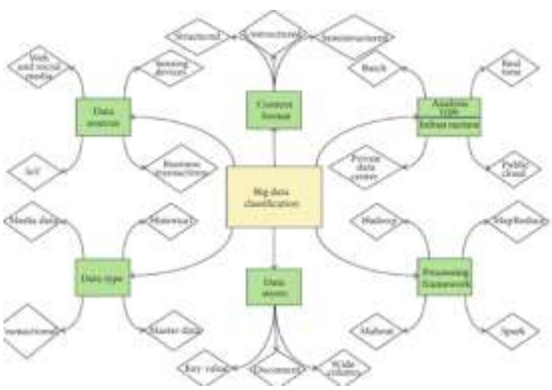

Fig.2. Big data classification framework [23]

Balachandran et al. [24] describe the passage's primary argument as that big data and cloud computing combine to improve business outcomes. Large amounts of storage and processing power are needed to handle big data, which is frequently costly and challenging for small and medium-sized businesses. This issue is resolved by cloud computing, which provides these resources at a reduced cost over the internet. This combination helps organizations make better decisions by making data analysis quicker, simpler, and more effective. Nevertheless, there are some hazards and difficulties associated with using cloud-based big data services that must be properly managed. Bahrami et al. [25] spoke about cloud computing, which offers complete storage and assistance, which is helpful for handling massive data. Key aspects of cloud computing are capacity expansion, adaptability, resource sharing, services on demand, and simple collaboration, which were discussed in the paper. The various cloud planning levels and their functions in managing large amounts of data were discussed. It was also described how business intelligence may employ cloud services to transform massive volumes of data into useful and usable information. Finally, the key problems with cloud computing need to be resolved for it to be a trustworthy big data solution.

III. CONCLUSION

Big data has evolved significantly around the world and now plays a central role in modern digital information societies, even among today's youth. The integration of big data, cloud computing, and machine learning plays a powerful role and provides a scalable computing environment. Such machine learning enhances data analysis, predictive modeling, and many other applications. This depicts both the advantages and challenges, which include data security and privacy risks, high operational and infrastructure costs, and network performance issues. These technologies have transformed the way data is stored, controlled, and processed in this era of computing systems.

This type of data system provides flexibility and allows institutions to control resources according to today's requirements. The combination of IoT, cloud computing, and big data enables smart applications in various sectors such as healthcare and business. Therefore, the interconnection of big data and machine learning defines the foundation of the next generation. The continuous increase in scalable big data will drive innovation and automation in the future. Without big data analytics, companies may feel blind, and it is the basic essence of any company. In conclusion, Big Data Analytics Server Machines are becoming a fundamental unit of the modern cloud computing environment and will continue to reshape future digital transformation in this era.

REFERENCES

1. Wu, C., Buyya, R. and Ramamohanarao, K., 2016. Big data analytics= machine learning+ cloud computing. arXiv preprint arXiv:1601.03115.
2. Berisha, B., Mëziu, E. and Shabani, I., 2022. Big data analytics in Cloud computing: an

boverview. Journal of cloud computing, 11(1), p.24.
3. Gupta, R., Gupta, H. and Mohania, M., 2012, December. Cloud computing and big data analytics: what is new from databases perspective?. In International conference on big data analytics (pp. 42-61). Berlin, Heidelberg: Springer Berlin Heidelberg.
4. Sandhu, A.K., 2021. Big data with cloud computing: Discussions and challenges. Big Data Mining and Analytics, 5(1), pp.32-40.
5. Mohammad, A.S. and Pradhan, M.R., 2021. Machine learning with big data analytics for cloud security. Computers & Electrical Engineering, 96, p.107527.
6. Demirbaga, Ü., Aujla, G.S., Jindal, A. and Kalyon, O., 2024. Cloud computing for big data analytics. In Big data analytics: Theory, techniques, platforms, and applications (pp. 43-77). Cham: Springer Nature Switzerland.
7. Xu, F., Zheng, H., Jiang, H., Shao, W., Liu, H. and Zhou, Z., 2018. Cost-effective cloud server provisioning for predictable performance of big data analytics. IEEE Transactions on Parallel and Distributed Systems, 30(5), pp.1036-1051.
8. Yetis, Y., Sara, R.G., Erol, B.A., Kaplan, H., Akuzum, A. and Jamshidi, M., 2016, July. Application of big data analytics via cloud computing. In 2016 world automation congress (wac) (pp. 1-5). IEEE.
9. Jain, V.K. and Kumar, S., 2015, May. Big data analytic using cloud computing. In 2015 Second International Conference on Advances in Computing and Communication Engineering (pp. 667-672). IEEE.
10. Saif, S. and Wazir, S., 2018. Performance analysis of big data and cloud computing techniques: a survey. Procedia computer science, 132, pp.118-127.
11. Goldin, E., Feldman, D., Georgoulas, G., Castaño, M. and Nikolakopoulos, G., 2017, July. Cloud computing for big data analytics in the process control industry. In 2017 25th Mediterranean Conference on Control and Automation (MED) (pp. 1373-1378). IEEE.
12. Kabanda, G., 2021. Performance of machine learning and big data analytics paradigms in cybersecurity and cloud computing platforms. Global Journal of Computer Science and Technology: G Interdisciplinary, 21(2), pp.1-25.
13. Hwang, K. and Chen, M., 2017. Big-data analytics for cloud, IoT and cognitive computing. John Wiley & Sons.
14. Sharma, N. and Shamkuwar, M., 2018. Big data analysis in cloud and machine learning. In Big data processing using spark in cloud (pp. 51-85). Singapore: Springer Singapore.
15. Ji, C., Li, Y., Qiu, W., Awada, U. and Li, K., 2012, December. Big data processing in cloud computing environments. In 2012 12th international symposium on pervasive systems, algorithms and networks (pp. 17-23). IEEE.
16. Purcell, B.M., 2014. Big data using cloud computing. Journal of Technology Research, 5, p.1.
17. Yassine, A., Singh, S., Hossain, M.S. and Muhammad, G., 2019. IoT big data analytics for smart homes with fog and cloud computing. Future Generation Computer Systems, 91, pp.563-573.
18. Lo'ai, A.T., Mehmood, R., Benkhlifa, E. and Song, H., 2016. Mobile cloud computing model and big data analysis for healthcare applications. IEEE Access, 4, pp.6171-6180.
19. Ahmed, S.T., Basha, S.M., Arumugam, S.R. and Patil, K.K., 2021. Big Data Analytics and Cloud Computing: A Beginner's Guide. MileStone Research Publications.
20. Kune, R., Konugurthi, P.K., Agarwal, A., Chillarige, R.R. and Buyya, R., 2016. The anatomy of big data computing. Software: Practice and Experience, 46(1), pp.79-105.
21. Talia, D., 2013. Clouds for scalable big data analytics. Computer, 46(5), pp.98-101.
22. Muniswamaiah, M., Agerwala, T. and Tappert, C., 2019. Big data in cloud computing review and opportunities. arXiv preprint arXiv:1912.10821.
23. Aydin, G., Hallac, I.R. and Karakus, B., 2015. Architecture and implementation of a scalable sensor data storage and analysis system using cloud computing and big data technologies. Journal of Sensors, 2015(1), p.834217.
24. Balachandran, B.M. and Prasad, S., 2017. Challenges and benefits of deploying big data analytics in the cloud for business intelligence. Procedia Computer Science, 112, pp.1112-1122.

25. Barhrami, M and Singhal, M, 2014. The role of cloud computing architecture in big data. In Information granularity, big data, and computational intelligence(pp. 275-295). Cham: Springer International Publishing.
26. Zanoon, N., Al-Haj, A. and Khwaldeh, S.M., 2017. Cloud computing and big data is there a relation between the two: a study. International Journal of Applied Engineering Research, 12(17), pp.6970-6982.
27. Agrawal, D., Das, S. and El Abbadi, A., 2011, March. Big data and cloud computing: current state and future opportunities. In Proceedings of the 14th international conference on extending database technology (pp. 530-533).
28. Matallah, H., Belalem, G. and Bouamrane, K., 2017. Towards a new model of storage and access to data in big data and cloud computing. International Journal of Ambient Computing and Intelligence (IJACI), 8(4), pp.31-44.
29. Buyya, R., Ramamohanarao, K., Leckie, C., Calheiros, R.N., Dastjerdi, A.V. and Versteeg, S., 2015. Big data analytics-enhanced cloud computing: Challenges, architectural elements, and future directions. arXiv preprint arXiv:1510.06486.
30. Abueid, A.I., 2024. Big data and cloud computing opportunities and application areas. Engineering, Technology & Applied Science Research, 14(3), pp.14509-14516.