Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

36

Cloud Based Data Analytics: A Review

Arundhati Dhar

Student, MIT World Peace University,

arundhatidhar555@gmail.com

Mayank Tiwari

Student, MIT World Peace University,

mayank4958@gmail.com

Abstract-

Large volumes of data are generated every second from various sources like social networking platforms, IoT (Internet of Things), sensory devices, wireless communications services, ecommerce platforms, government agencies, to name a few. Regular data processing paradigms yield insignificant results while dealing with data of such large volumes, and are consequently labeled as Big Data. Big Data is a blanket terminology that deals with the storage, management, processing, and most importantly, the analyzing of such data. Cloud Computing has emerged as a technology of paramount importance to modern computing, and deals with providing the infrastructure and computing resources required for such processes in an efficient and cost-effective manner. Various sectors including healthcare, education, and government agencies, are leveraging Big Data to improve decision-making. For example, the medical industry is making use of Big Data to better understand their patients and develop personalized treatment plans, while government agencies are using it to track and prevent fraud, waste, and abuse. This paper presents an in-depth description of cloud computing and big data. We then delve into Big Data analytics where we discuss various Big Data paradigms, and introduce Big Data analytics in the

context of cloud computing. Lastly, we discuss the advantages of using Big Data analytics in Cloud Computing as well its limitations and future enhancements for this vast domain.

Keywords- Cloud Computing, Data Analytics, Big Data, Hadoop, MapReduce

I. INTRODUCTION

In the new age technology, a large amount of data is generated on a daily basis in which analyzing the clean data is extremely crucial [2]. The analyzed data thereafter needs a secured platform accessible for future use by companies, individuals or institutions for proper estimations and calculations.

Cloud-based data analytics has become a popular approach to dealing with large volumes of data in a scalable and cost-effective manner [1]. In recent years, there has been a significant increase in research on cloud-based data analytics, with a focus on developing new algorithms, tools, and frameworks to improve the performance, efficiency, and scalability of cloud-based data analytics systems. The massive amount of data generated from data warehouses, social media, IoT sensory devices, websites and applications [1] is needed to be stored and processed in order to analyze the trends and patterns of the different dimensions of the data. This process of computing the data requires efficient and cost-effective tools which harbor the data for proper accessibility and evaluations, and this is where cloud-based data analytics come into play.

One of the major advantages of cloud-based data analytics is the ability to leverage cloud resources to perform complex data processing tasks in real-time [5]. Due to this, new cloud-based data analytics tools and frameworks have been developed such as Apache Hadoop, Apache Spark, Snowflake, Databricks and Google BigQuery [9], which are built to handle large volumes of data in parallel across multiple nodes. These tools use distributed computing to enable the processing of massive datasets quickly and efficiently, making it possible to perform tasks such as predictive analytics and machine learning on large volumes of data [11].

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

II. BIG DATA

2.1 Feature characteristics [1]:

In this fast-paced world, which is primarily dominated by technology, massive sizes of data is generated everyday. Data is produced from numerous sources like websites, social media, forms, questionnaires, emails, surveys and many more such portals. Big data is basically referred to the huge amount of data that is exponentially generated then collected and stored. It is considered that Big Data has different feature-based attributes which are collectively called as the 5 V's [9].

The 5 V's stand for Volume, Variety, Velocity, Veracity and Value.

- Volume is a metric of how sizable the data is, that is it measures the amount of data accessible data is available in an organization.
- Variety means the different sources of data collection, for example the data for sales can be collected from different vendor's databases [11], therefore the total number of such sources can be referred to as the variety.
- Velocity refers at which rate the data is collected, for example the data from stock market is collated on a daily basis depending on its status, and hence the data collected in this field is collected at a higher rate [14].
- Veracity measures the accuracy and the reliability of the data as the authenticity of the data [10] for analysis is of the most crucial aspects of it.
- And lastly, Value means the final product that we receive from the data at our disposal, that is discovering the hidden patterns and conclusions from the cluster of acquired data [17].

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

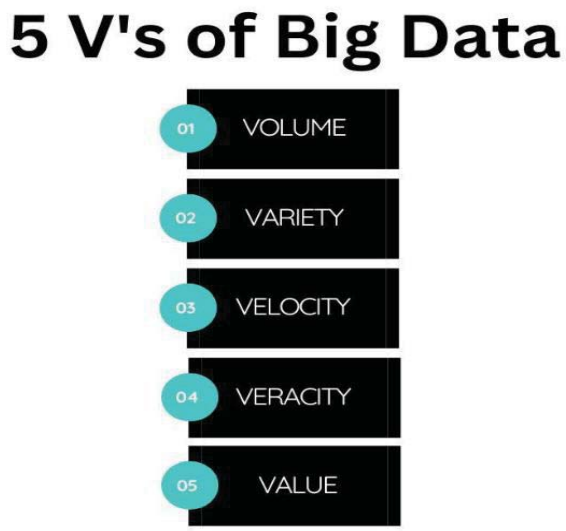

Fig. 1. The Five V's of Big Data

***2.2 Why Big data analysis*:**

Big data analysis benefits businesses, individuals and organizations in significant ways [7]. The customer behavior upon launching new products or making any change to the old existing products can be analyzed, moreover the results of the promotional offers implemented by companies on their products to increase their sales numbers can be obtained through big data analysis [3]. Subscription based streaming service companies keep a track of their viewers habits within the national or the international arena, based on the results of which the company further decides about adding or removing the features from their platform. Customers' needs, preferences and purchase behavior are some of the vital information that any company needs to keep in check in order to upkeep their business according to the demands [5, 7, 8].

2.3 The BDA Cycle:

The unstructured data that is gathered which includes emails, blogs, twitter, facebook posts, images and videos, company data etc. needs to be cleansed, wrangled and analyzed to be understood by the common man. Data analysis comes into the picture here, where the data accumulated is processed to identify trends, connected patterns, unknown facts and to pinpoint the key insights.

The huge chunks of raw data sizes up to Terabytes and petabytes [16], hence making a calculation or handling such a huge database becomes extremely strenuous and a complicated process in local computers. Organizations have confidential data in the form of passwords and keywords which can be under the threat of accessibility by hackers [17] if digital security is compromised. Moreover, the more data we get access to, the more infrastructure is required, for example a big amount of data needs a storage space and hence more servers need to be built in order for storage, which eventually increases the costs of the entire procedure of data analysis. Cloud computing plays a key role in such scenarios. Cloud computing provides us with computational resources as services like storage, servers, networking, analytics provisioned with minimal management in a pay per use system which is also known as Infrastructure on Demand (IoD).

III. CLOUD COMPUTING

This chapter provides an overview of cloud computing and its related terminologies.

3.1. Cloud Computing:

Cloud Computing has completely transformed the way we use and manage technology [1, 6]. Cloud Computing is the delivery of computer services and resources through a system of remote servers connected to the Internet that can be availed in an on-demand fashion, thereby enabling customers to avail the services and resources like general storage, databases, softwares and applications. This enables customers, organizations as well as individuals both, this enables customers to pay only for what they use and helps reduce the IT costs and overhead of buying and maintaining physical data centers and servers [8]. Being at the forefront of the current IT landscape, Cloud Computing has been adopted by companies of every scale and every type, and industry for the general IT paradigms, like backup and recovery, virtual desktops, software development and testing, big data analytics, etc. Due to its extensive list of advantages, Cloud Computing has made its way into every major industry like healthcare, finance, education, government, etc. [6, 7, 29].

Cloud Computing has completely changed the IT landscape. Here are some reasons why it's a big thing:

- Scalability and Flexibility: Cloud computing has enabled businesses to scale their IT

and computing infrastructure with ease, without having to pay for any additional and expensive hardware or software [13]. This has made Cloud Computing a boon for businesses with periodic spikes, or unpredictable growth.

- Cost-Effective: Since computing resources and services like servers, storage, databases, software and applications, etc., are delivered over the Internet, it eliminates the need for businesses to invest in costly computing resources and their maintenance. This has made Cloud computing a cost-effective solution for customers, as they only have to pay for the required services on a pay-as-you-go model [12, 14].
- Accessibility and Availability: Cloud Computing enables remote access to data, applications and any services on any device with an active internet connection.
- Security: Due to the security measures offered by Cloud providers like firewalls, intrusion detection and prevention, data encryption, and data backup and recovery, Cloud Computing has proven to have more secirity compared to traditional IT infrastructure [28, 31].
- Innovation: One of the major reasons for adoption of Cloud Computing is that it enables businesses to innovate by providing effortless access to latest technologies like AI, ML, Big Data analytics, and IoT, as compared to traditional IT systems. This allows businesses to innovate efficiently by developing cutting-edge technologies.

3.2. Types of Cloud:

Public, private, and hybrid clouds are widely considered to be the primary types of cloud computing. The following section discusses these three environments [7].

1. Public: A public cloud is a type of computing where computing resources and services are offered and maintained by third-party CSPs (Cloud Service Providers) via the internet and used by organizations or individuals who want to purchase them [1, 3, 4]. Users, both organizations and individuals, can avail these resources in a pay-as-you-go basis, thereby reducing the overhead cost of additional resources. Some popular CSPs are Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) [11].

2. Private: Private Cloud is a type of computing environment wherein the resources and services are provided by the CSPs to a single customer only. Private cloud environments have proven to offer better security measures for their customers. These types of computing environments are managed as well as maintained by the customer's own in-house IT teams [25]. Many organizations opt to avail private cloud instead of public cloud because private clouds make it easier for the organizations to manage as well as satisy their internal benchmarks. The preferred mode of hosting private clouds is on-premise in the customer's own data center.
3. Hybrid: It is a type of cloud computing environment that mixes certain aspects of both public and private clouds, enabling organizations to capitalize on the advantages of both the computing environments [17]. In this type of cloud, some resources are provided by a public cloud, and others are provided by a private cloud. Hybrid cloud computing has now become the mainstream and preferred approach since customers do not want to rely solely on single public cloud [2]. The two clouds are connected through a secure and scalable network, allowing data and applications to be seamlessly transferred between them.

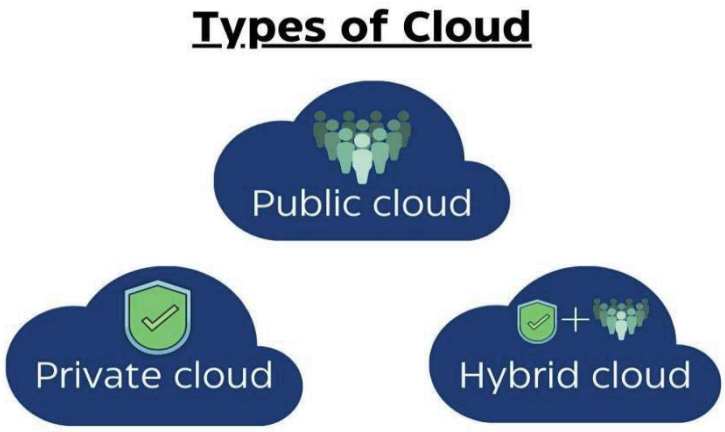

Fig. 2. Types of Cloud

Cloud computing can further be classified into three service models to cater to different business requirements: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [27, 31].

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

1. Infrastructure-as-a-service (IaaS): The infrastructure resources offered in this model include virtual machines, storage systems, and networking to users [14]. It allows businesses to use and manage the upper level of their systems like the operating system, softwares and data, without having to worry about the implementation or maintenance of the underlying infrastructure. Some of the major IaaS providers include Amazon Web Services (AWS) and Microsoft Azure.
2. Platform-as-a-Service (PaaS): This model offers a complete development and deployment platform to users, including application development frameworks, databases, and middleware [14], among others. This gives developers the freedom to build applications as they would on a normal system without having to make any considerations towards the underlying infrastructure. Some major PaaS providers are: Google App Engine and Heroku.
3. Software-as-a-Service (SaaS): This model offers a complete software application to users, which is accessed over the internet. The cloud provider is responsible for maintaining the entire stack, including infrastructure, middleware, and application software. To access the softwares available through SaaS models, customers have to pay a subscription fee and an active internet connection. Salesforce and Microsoft 365 are some of the major SaaS providers.

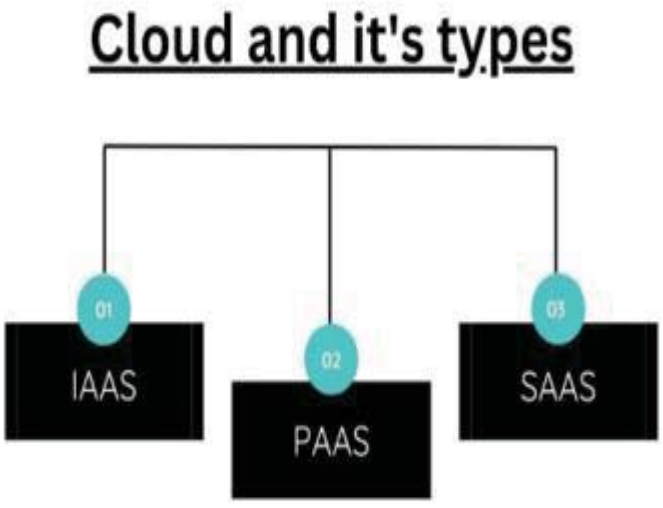

Fig. 3. Cloud and its types

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

3.4 Data services of cloud:

Data-as-a-service:

Data as a service (DaaS) is a cloud computing model that provides users with an interface to access and use data on demand. DaaS allows organizations to outsource the technical components of data management and focus on leveraging data to drive business decisions. DaaS providers traditionally offer a plethora of services such as data warehousing, data integration, data analytics, data visualization, and data security. Some popular DaaS providers include AWS Data Exchange, Google Cloud BigQuery, Microsoft Azure Data Marketplace, and IBM Cloud Data Services. DaaS is becoming increasingly popular as a cost-effective and efficient way to manage and analyze data. It allows businesses to scale their data operations without needing to invest in expensive infrastructure, tools, and personnel. With DaaS, businesses can access the data they need in real-time, helping them make better and faster business decisions.

Big Data as a Service:

Big data as a service (BDaaS) is a cloud-based solution that provides access to big data applications, storage, and analytics capabilities. BDaaS enables organizations to store, process, and analyze large amounts of structured and unstructured data in a scalable and cost-effective manner without having to invest in expensive infrastructure [1, 30]. BDaaS providers offer a plethora of BigData services such as Hadoop cluster management [10], data warehousing, data analytics, data visualization, and machine learning. Some popular BDaaS providers include Amazon Web Services EMR, Google Cloud Big Data, Microsoft Azure HDInsight, and IBM Analytics Engine. BDaaS is becoming increasingly popular as it enables businesses to derive insights from large data sets without having to worry about the technical aspects of managing their big data infrastructure. With BDaaS, businesses can rapidly scale their big data operations, reduce costs, and focus on their core business operations.

In summary, BDaaS has revolutionized the way businesses can leverage big data by enabling them to access, process, and analyze large data sets efficiently and cost-effectively.

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

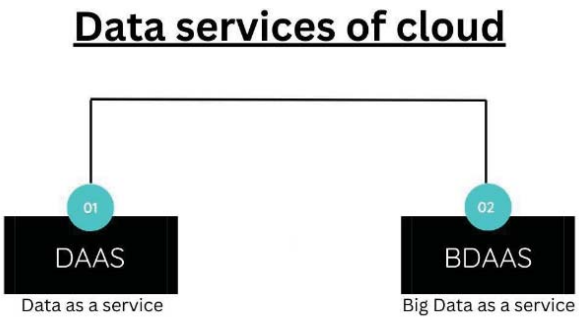

Fig. 4. Data Services of Cloud

Cloud computing has revolutionized the way businesses operate, and data analytics is no exception. Cloud computing offers several advantages for data analytics, however, that is not without its drawbacks. The following section describes some of the advantages as well as disadvantages to using cloud computing for data analytics.

3.5.1 Advantages of Cloud Computing for Data Analytics:

Cloud computing has made it possible to store Big Data, thereby increasing the effectiveness of using cloud-based solutions for data analytics. Some of these are:

1 Scalability: One of the primary advantages of cloud computing is its scalability. With cloud computing, organizations can easily scale their computing resources up or down as needed, depending on their data analytics needs. This is particularly useful for organizations that experience seasonal spikes in their data analytics needs or for those that are rapidly expanding.

2 Flexibility: Cloud computing is highly flexible, allowing organizations to access computing resources from any location and on any device [15]. This feature is especially more useful for organizations with remote or distributed teams who need to collaborate on data analytics projects.

3 Easy access to computing resources [18]: With cloud computing, organizations can easily access computing resources, such as storage and processing power, without having to worry about maintaining their own infrastructure. Since the management of

the computing resources will be looked after by the CSP, this allows organizations to focus on their core functionalities only.

***3.5.2. Disadvantages* [26, 27, 28]:**

1. Security risks [26]: Cloud computing revolves around storing and processing sensitive data in servers located remotely, which can expose the data to security risks such as hacking, data breaches, and unauthorized access [31]. This can be a particular concern for organizations in highly regulated industries, such as finance or healthcare.
2. Limited control over data [28]: When enterprises opt for the usage of cloud computing, they essentially are investing their resources and trust in a third-party service provider [28, 31]. This means that they have limited authority over how the data is stored, processed, and secured.

IV. CLOUD BASED DATA ANALYTICS

Cloud-based data analytics (CBDA) has emerged as an efficient solution for organizations to handle large-scale data processing and analytics [16, 17]. With the ever-evolving volume and the increase in complexity of data, cloud computing offers a flexible and scalable solution for data processing, storage, and analytics. In this paper, we will provide an overview of CBDA, its architecture, along with the data services offered in the cloud. CBDA involves the use of cloud computing resources to store, process, and analyze large datasets. The primary advantage of CBDA is the ability to access computing resources on-demand, without the need for local hardware and infrastructure [19]. Cloud-based data analytics can be used for various purposes, including business intelligence, predictive analytics, machine learning, and data mining [19, 22]. The architecture of CBDA involves multiple layers of cloud computing resources that work together to store, process, and analyze data [17]. The first layer is the storage layer, which provides scalable and reliable storage for data. The second layer is the processing layer, which involves the use of cloud-based processing resources, such as virtual machines or containers. The third layer is the analytics layer, which includes various tools and services for data analytics, such as data visualization, machine learning algorithms, and predictive modeling [18, 19].

Cloud computing platforms offer a wide range of data services that can be used for CBDA, such as Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure. These services include storage services, such as Amazon S3 and Google Cloud Storage, database services, such as Amazon RDS and Google Cloud SQL, and analytics services, such as Amazon Athena and Google BigQuery.

4.1 Data Lakehouse:

The word "Data Lake" was coined by James Dixon who was a Chief technology officer at Pentaho [20]. Data lake revolves around cloud environments and acts like a logical storage space where enterprises can store large data sets which distribute accessibility to the data sets to get the maximum throughput. Both structured and unstructured data can be pooled in a Data Lake.

The idea of data warehouses existed for the longest time which provided feasibility whilst handling distributed data for parallel computing [27]. In a data lakehouse, instead of storing data in different harbors, the idea of a single hub of data is treated as the main repository of information [23]. Hadoop technologies or Cloud computing services can be used to build a data lake [22]. The Cloud services provider handles the various infrastructure, applications and security needs, thus minimizing the workload of the IT professionals who can now focus more on analyzing, processing and managing the data. As more public clouds emerge, companies and institutions both private and public have found an efficient way to store and process which most importantly acts as a reliable data repository. Data lakes assists with data security, and by supporting the workloads by maintaining storage systems and by providing access to the multiple users all across the globe for quicker processing needs [24]. Real time can be dealt with in Data Lake house which is a huge bonus feature for data analyzing as well as business decisions for improving customer experience and for meeting up with the demands.

4.2 Hadoop:

Doug Cutting and Mike Cafarella developed Hadoop in 2006 at Yahoo, where the entire idea for development was based on Google's Map Reduce and Google File System (GFS) technologies [22]. It is primarily written in Java and is maintained by the Apache Software

Foundation. Hadoop framework has been serving the purpose of log file analysis in cloud by various domains by organizations and researchers alike [21].

Hadoop is an open-source distributed processing framework designed to efficiently store and process large datasets with commodity hardware [26, 27, 28]. It is a highly scalable and fault-tolerant system that makes it possible to process huge datasets in parallel.

The Hadoop ecosystem includes several different components, each serving a unique purpose [9, 20]. The two primary components are:

- Hadoop Distributed File System (HDFS): a distributed file system that stores and manages datasets across a cluster of machines [20].
- MapReduce: a programming model for processing large datasets in parallel across a distributed system [16, 25, 28].

Other Hadoop ecosystem components include [30]:

- YARN (Yet Another Resource Negotiator): a resource manager that schedules tasks across a cluster of machines [16, 18].
- Hive: Another popular data warehouse, Hive allows processing of large datasets using its own mirrored version of SQL, called HiveQL. It converts HiveQL queries into MapReduce [20].
- Pig [20, 25]: A dataflow platform/system, Pig gives users the freedom to describe how the data would be processed, while also capitalizing on Hadoop MapReduce by giving the user an engine for parallel execution of operations. This is achieved by writing scripts using the Pig Latin language.
- Spark: an in-memory data processing and analytics engine that runs on top of Hadoop [21, 24].

***4.3 MapReduce*:**

MapReduce is a component of the framework of Apache Hadoop software. It is also a programming module which is used for operating on large datasets and processing them. The Apache Hadoop software is also known for handling distributed processing of big data.

The MapReduce model expects input data to be in the form of tuples, and breaks down processing into two phases: the Map phase and the Reduce phase. In the first phase, parallel processing of the input data across multiple nodes takes place by splitting the input data into independent blocks. Every node in the cluster will process its own block of data and produce tuples of key-value pairs [10, 11]. In the second and final phase, the key-value pairs thus generated in the first phase are collected, sorted by key and processed to generate the final output. The key idea behind MapReduce is to simplify the process of processing large datasets in a highly parallel and distributed fashion by breaking the data processing into smaller chunks that can be processed independently by separate compute nodes [10].

The MapReduce programming model has been highly successful in handling batch processing of big data [23, 27] such as log processing, analyzing social media data, and processing large datasets in a variety of industries such as healthcare, finance, retail, and e-commerce. While Hadoop sparked the initial development and adoption of MapReduce, other distributed computing systems such as Apache Spark have also adopted the model, provided additional functionality and performance improvements while maintained compatibility with existing MapReduce code [11, 21].

V. CONCLUSION

The emergence of cloud computing has transformed the way big data is processed and analyzed. This literature review explored various aspects of big data analytics in cloud computing, including its features, advantages, classification, and the BDA cycle. Additionally, we discussed the different types and advantages of cloud computing, as well as data services in cloud environments. Furthermore, the review delved into popular big data analytics tools used in cloud computing such as data lakehouses, Hadoop, and MapReduce. We then highlighted the advantages of cloud-based data analytics, including its scalability, cost-effectiveness, and ease of use. Our findings conclude that the merger of cloud computing with big data analytics has introduced new possibilities for organizations to harness the power of data for business insights and innovation.

REFERENCES

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

[2] Subia Saif, Samar Wazir, *Performance Analysis of Big Data and Cloud Computing Techniques: A Survey*, Procedia Computer Science, Volume 132, 2018, Pages 118-127, ISSN 1877-0509.

[3] Sangeetha, K. & Prakash, Parvathy. (2015), *Big Data and Cloud: A Survey*, 10.1007/978-81-322-2135-7_81.

[4] Ahmed, N., Barczak, A.L.C., Susnjak, T. et al, *A comprehensive performance analysis of Apache Hadoop and Apache Spark for large scale data sets using HiBench. J Big Data,* 7, 110 (2020).

[5] R. Buyya, K. Ramamohanarao, C. Leckie, R. N. Calheiros, A. V. Dastjerdi and S. Versteeg, "Big Data Analytics-Enhanced Cloud Computing: Challenges, Architectural Elements, and Future Directions," 2015 IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS), Melbourne, VIC, Australia, 2015, pp. 75-84, doi: 10.1109/ICPADS.2015.18.

[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] 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.

[8] A. K. Sandhu, "Big data with cloud computing: Discussions and challenges," in Big Data Mining and Analytics, vol. 5, no. 1, pp. 32-40, March 2022, doi: 10.26599/BDMA.2021.9020016.

[9] Gupta, R., Gupta, H., Mohania, M. (2012), "Cloud Computing and Big Data Analytics: What Is New from Databases Perspective?", In: Srinivasa, S., Bhatnagar, V. (eds) Big Data Analytics. BDA 2012. Lecture Notes in Computer Science, vol

7678. Springer, Berlin, Heidelberg.

[10] A. K. Manekar and G. Pradeepini, "Cloud Based Big Data Analytics a Review," 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, 2015, pp. 785-788, doi: 10.1109/CICN.2015.160.

[11] Zanoon, Dr. Nabeel & Alhaj, Abdullah & Khwaldeh, Sufian. (2017), "Cloud Computing and Big Data is there a Relation between the Two: A Study. International Journal of Applied Engineering Research", 12. 6970-6982.

[12] Ying Liu, Anthony Soroka, Liangxiu Han, Jin Jian, Min Tang, "Cloud-based big data analytics for customer insight-driven design innovation in SMEs, International Journal of Information Management", Volume 51, 2020, 102034, ISSN 0268-4012.

[13] Khan, S., Shakil, K.A., Alam, M. (2018), "Cloud-Based Big Data Analytics—A Survey of Current Research and Future Directions", In: Aggarwal, V., Bhatnagar, V., Mishra, D. (eds) Big Data Analytics. Advances in Intelligent Systems and Computing, vol 654. Springer, Singapore.

[14] Marino S, Zhao Y, Zhou N, Zhou Y, Toga AW, Zhao L, et al. (2020), "Compressive Big Data Analytics: An ensemble meta-algorithm for high-dimensional multisource datasets", PLoS ONE 15(8): e0228520.

[15] Ajimoko, O. J., 2018, "Considerations for the Adoption of Cloud-based Big Data Analytics in Small Business Enterprises", The Electronic Journal Information Systems Evaluation, 21(2), pp. 63-79.

[16] Shingyu Kim, Junghee Won, Hyuck Han, Hyeonsang Eom, and Heon Y. Yeom. 2011, "Improving Hadoop performance in intercloud environments. SIGMETRICS Perform. Eval", Rev. 39, 3 (December 2011), 107–109.

[17] Depeige, A., Doyencourt, D, "Actionable Knowledge as A Service (AKAAS): Leveraging big data analytics in cloud computing environments", Journal of Big Data 2, 12 (2015).

[18] Carretero Pérez, Jesús; et.al. (eds.), (2015) Proceedings of the Second International Workshop on Sustainable Ultrascale Computing Systems (NESUS 2015): Krakow,

Poland. Universidad Carlos III de Madrid, pp. 51-62. ISBN: 978-84-608-2581-4.

[19] Naga Raju Hari Manikyam and Dr. S. Mohan Kumar, "Methods and Techniques To Deal with Big Data Analytics and Challenges In Cloud Computing Environment", International Journal of Civil Engineering and Technology, 8(4), 2017, pp. 669-678.

[20] Rai, Ibrahim. (2018), "Performance Analysis of Big Data Tools. International Journal of Advances in Computer Science and Technology", 7. 43-48. 10.30534/ijacst/2018/05762018.

[21] Ilias Mavridis, Helen Karatza, "Performance evaluation of cloud-based log file analysis with Apache Hadoop and Apache Spark", Journal of Systems and Software, Volume 125, 2017, Pages 133-151, ISSN 0164-1212.

[22] Apache Hadoop, available at https://hadoop.apache.org

[23] C. -H. Lin, J. -C. Liu and T. -C. Peng, "Performance evaluation of cluster algorithms for Big Data analysis on cloud," 2017 International Conference on Applied System Innovation (ICASI), Sapporo, Japan, 2017, pp. 1434-1437, doi: 10.1109/ICASI.2017.7988182.

[24] K. Wang and M. M. H. Khan, "Performance Prediction for Apache Spark Platform," 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems, New York, NY, USA, 2015, pp. 166-173, doi: 10.1109/HPCC-CSS-ICESS.2015.246.

[25] Yadav, Saneh & Sohal, Asha. (2017), "Review Paper on Big Data Analytics in Cloud Computing.".

[26] Inukollu, Venkata & Arsi, Sailaja & Ravuri, Srinivasa. (2014), "Security Issues Associated with Big Data in Cloud Computing", International Journal of Network Security & Its Applications. 6. 45-56. 10.5121/ijnsa.2014.6304.

[27] Sudhir Allam. (2018), "USAGE OF HADOOP AND MICROSOFT CLOUD IN BIG DATA ANALYTICS: AN EXPLORATORY STUDY", International Journal of

Innovations in Engineering Research and Technology, 5(10), 27–32.

[28] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, Samee Ullah Khan, "The rise of "big data" on cloud computing: Review and open research issues", Information Systems, Volume 47, 2015, Pages 98-115, ISSN 0306-4379.

[29] Bahrami, M., Singhal, M. (2015), "The Role of Cloud Computing Architecture in Big Data", In: Pedrycz, W., Chen, SM. (eds) Information Granularity, Big Data, and Computational Intelligence. Studies in Big Data, vol 8. Springer, Cham.

[30] Zulkernine, Farhana & Martin, Patrick & Zou, Ying & Bauer, Michael & gwadry-sridhar, Femida & Aboulnaga, Ashraf. (2013), *Towards Cloud-Based Analytics-as-a-Service (CLAaaS) for Big Data Analytics in the Cloud*, 62-69. 10.1109/BigData. Congress. -2013.18.

[31] M. B. Nirmala, "WAN Optimization Tools, Techniques and Research Issues for Cloud-Based Big Data Analytics," 2014 World Congress on Computing and Communication Technologies, Trichirappalli, India, 2014, pp. 280-285, doi: 10.1109/WCCCT.2014.72.