International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 3 ǁ 2022 ǁ PP. 62-67

Evolution of Big Data Analytics in Cloud Computing Environment

B Kiruthika[1], Dr. B. Srinivasan[2], Dr. P. Prabhusundhar[3]
[1]*Research Scholar in Computer Science*
[2]*Associate Professor in Computer Science,*
[3]*Assistant Professor in Computer Science,*
Gobi Arts & Science College (Autonomous), Gobichettipalayam.

ABSTRACT
Big data are increasing massively with growing diversity and challenges for the data management lifecycle to store, transport, process, mine and serve the data. Most of the largest-volume, fastest-streaming, and most complex data can be source of big data from multitude of sources. The processing of these data is complicated in conventional data processing software. Also it demands large computational infrastructure for data analysis to ensure successful data processing and analysis. Cloud computing is an influential technology to perform extremely large and complex computing. Cloud can provide a reliable, fault-tolerant, scalable environment, infrastructure and software resources as services to big data distributed management systems. And therefore, it eliminates the need to maintain expensive computing hardware, dedicated space, and software.

KEYWORDS
Big Data, Cloud Computing, Big data Analytics, Data Analytics as a Service, Software as a Service, Data Aggregation, Virtualization, Hadoop, Apache

---

Date of Submission: 09-03-2022 Date of acceptance: 25-03-2022

I. INTRODUCTION

Big Data and Cloud Computing as two mainstream technologies are at the center of concern in the IT field. The term big data can be defined as a dataset, which is too large and complex. Big Data is used in decision making process to gain useful insights hidden in the data for business and engineering. At the same time it presents challenges in processing, cloud computing has helped in advancement of big data by providing computational, networking and storage capacity. During the same time frame, cloud computing emerged to provide crucial computing support to address these challenges. The two go hand-in-hand, with several public cloud services performing big data analytics. With Software as a Service **(SaaS)** becoming ever more popular, keeping state of the art with cloud infrastructure best practices and the types of data that can be stored in huge quantities is essential. It is important to form a clear distinction between cloud computing and big data, the relationship between them, and why the two are a perfect match, bringing us lots of new, innovative technologies, such as artificial intelligence.

II. BASICS OF CLOUD COMPUTING

As a new model for computing, cloud computing delivers scalable, on-demand, pay-as-you-go access to a group of computing resources. Basically, cloud computing aims to increase the utilization rate of physical resources and provide virtual resources to support applications and services. The cloud computing services include Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS).

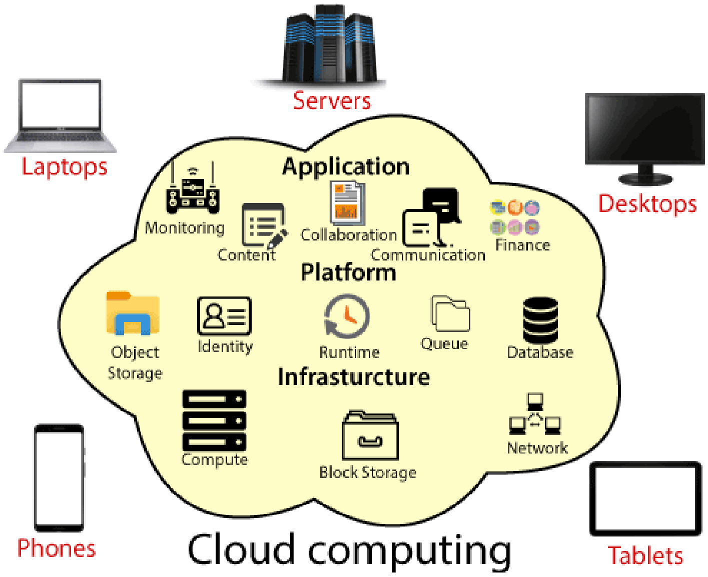

Fig 2.1 Cloud Computing

Most importantly the small to medium sized businesses use cloud computing for big data technology implementation for three main reasons – (i) hardware cost reduction, (ii) processing cost reduction, and (iii) ability to test the value of big data. The major concerns regarding cloud computing are security and loss of control.

2.1 Types of Cloud Computing

The three types of cloud computing are **public** cloud, the **private** cloud, and the **hybrid** cloud.

Public Cloud: A public cloud is the pay-as-you-go services. The cloud resources that are owned and operated by a third-party cloud service provider are termed as public clouds. It delivers computing resources such as servers, software, and storage over the internet

Private Cloud: A private cloud is internal data center of a business not available to the general public but based on cloud structure. The cloud computing resources that are exclusively used inside a single business or organization are termed as a private cloud.

Hybrid Cloud: It is the combination of public and private clouds, which is bounded together by technology that allows data applications to be shared between them. Hybrid cloud provides flexibility and more deployment options to the business.

III. COMPARISON OF BIG DATA & CLOUD COMPUTING

"Big Data" refers to the large sets of data collected, while "Cloud Computing" refers to the mechanism that remotely takes this data in and performs any operations specified on that data.

Big data is a data analysis methodology that contributes in the rapid growth of various applications using in daily life like social network analysis, semantic web analysis and bioinformatics network analysis. Every day a huge amount of data is produced from different sources. This data is so big in size that traditional processing tools are unable to deal with them. Besides being big, this data moves fast and has a lot of variety. Big Data is a concept that deals with storing, processing and analyzing large amounts of data. According to the National Institute of Standards and Technology (NIST), "*Big Data is a term used to describe the large amount of data in the networked, digitized, sensor-laden, information-driven world*".

Cloud computing on the other hand is about offering the infrastructure to enable such processes in a cost-effective and efficient manner. This refers to the *processing* of anything, including Big Data Analytics, on the "cloud". The "cloud" is just a set of high-powered servers from one of many providers. They can often view and query large data sets much more quickly than a standard computer could. Many sectors, including among others businesses (small of large), healthcare, education, etc. are trying to leverage the power of Big Data. The application of cloud computing is a recent trend to settle and handle the relevant issues of big data. It also provides security to big data through Hadoop.

IV. UNIFIED APPROACH OF CLOUD COMPUTING FOR BIG DATA

One of the best qualities of cloud is sharing of resources and data into data centers on internet. In today's world Cloud is using big data processing technology to enhance application aggregation, data aggregation and data utilization. Cloud computing relies on several technologies including virtualization, network security, and high availability to provide services over the network. These technologies make it easier, more efficient, and more economical to set up architecture for big data analysis.

Virtualization is the fundamental technology for cloud computing, which abstracts an application, operating system, or data store from the underlying hardware or software. Virtualization creates a "layer of abstraction" between the physical systems and a virtual environment in the virtualization process (Big Data Virtualization). Server virtualization optimizes the use of redundant computing and storage resources by virtualizing distributed computer resources (e.g., CPU, RAM, Network, and Disk) and managing them in the same resource pool. With virtualization, cloud computing can provide on-demand big data services and support big data technologies including big data storage, process, analysis, visualization, and remote collaboration (Fig. 4.1). Virtualizing big data resources as a pool serves as a user-friendly interface and makes big data analytics accessible to end users.

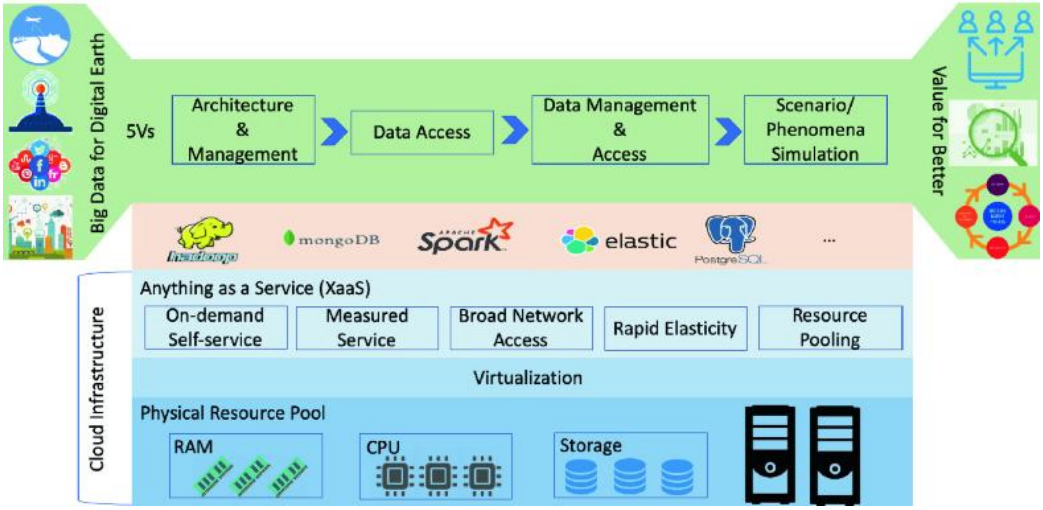

Fig. 4.1 Cloud computing for big data analysis

Cloud computing provides fundamental support to address the challenges with shared computing resources including computing, storage, networking and analytical software; the application of these resources has promoted impressive Big Data advancements.

4.1 Cloud Storage for Big Data Storage

The high volume big data lead to challenges for data storage. Cloud computing is capable of providing unlimited storage support helps to solve the volume challenge of big data, as the cloud provides virtually customizable storage with the flexibility to expand and reduce the size. An alternative solution is Data Storage as a Service (DSaaS) enabled by block storage, which has the potential of adding external storages as "blocks". With block storage, enlarging the storage is possible without physically loading hard drives. Virtually unlimited scalable storage offered by cloud computing grants users the capability of dynamic adjustment to satisfy the storage requirements of data with high volume and velocity. The adaptable virtual resource offers effortless data sharing within production environments by allowing for an external data block to be detached and remounted from one machine to another. External data storage can be automatically backed up to prevent users from losing data, and backups that are securely saved at the back-end server can be easily transferred and restored. In addition, information security is guaranteed because the physical location cannot be obtained from the disk drive.

4.2 Big Data Processing by Cloud Computing

Processing large volumes of data requires dedicated computing resources, e.g., faster CPUs and networks and larger disks and RAMs. Cloud computing provides on-demand resources and delivers configurable resources including mountable external storage spaces, computing resources (CPU, RAM), and network services. Traditionally, a computer uses approximately two-thirds of the power of a busy computer, and cloud computing has the potential to provide on-demand computing resources. Isolated virtual structures have been created for big data systems to enhance system stabilities, which can be easily managed in different file systems and replicated through backup images to provide fast configuration recovery. The ability to replicate environments automates the expansion of compute nodes in virtual machine clusters, thereby efficiently utilizing resource pools to support big data analytics. With the foundational support of storage for big data, data processing inherits the advantages of fast data acquisition and relocation.

Although cloud computing could serve as an excellent infrastructure option for big data processing, several aspects should be considered to minimize the bottleneck effect for the general processing speed, such as the choice of cloud volume type according to I/O demand and cloud bandwidth selection according to application requirements.

4.3 Big Data Analytics by Cloud Computing

Big data analytical platforms such as Apache and Hadoop are installed on physical machine clusters, which results in a waste of computing resources due to hardware redundancy (CPU and RAM). Distributed analytical platforms can be migrated to the virtual clusters provided by cloud computing through virtualization technology from physical machine clusters, and therefore the usage of computing resources is optimized in an efficient manner.

With the aid of autoscaling and load balancing, deploying on-demand and scalable big data analytical platforms could easily provide flexible analytical frameworks and minimize the waste of computing resources. Autoscaling supports parallel algorithms on distributed systems and architectures for scalability. It allows for the expanded resources to function when the algorithms or programs are enabled with parallel computing capability. Without it, public cloud providers such as AWS could not offer automatic scalability. The load balancer distributes workloads among virtual clusters and triggers autoscaling functions when analytics require higher computing configurations. The virtual system as a whole could dynamically fit higher computing requirements by launching more virtual duplications as needed. The load balancer acts as a virtual network traffic distributor and can be optimized to better allocate overall resources.

4.4 Big Data Sharing and Remote Collaboration by Cloud Computing

Conventional deployment of big data systems requires complex settings and efforts to share data assets. It lacks access control and often leads to data security and data privacy issues. Cloud computing enhances the sharing of information by applying advanced analytical tools and managing controlled access and security as well. Virtualization enables different parties to share data assets to achieve various goals and objectives under a centralized management system. With the support of cloud computing, it is possible to flexibly share data and remotely collaborate, which involve inter-disciplinary collaborations and superior workflows. Though data sharing, computational resource sharing, and production environment sharing, cloud computing can potentially be used to build an intuitive environment to support various businesses and applications. Unfortunately, workflow sharing remains challenging due to domain limitations.

V. RELATIONSHIP BETWEEN BIG DATA & CLOUD COMPUTING

In the last decade, cloud-adoption in the Big Data space has been rising exponentially. In turn, Big Data is often generated by large, network-based systems. It can be in either a standard or non-standard format. If the data is in a non-standard format, artificial intelligence from the Cloud Computing provider may be used in addition to machine learning to standardize the data. From there, the data can be harnessed through the Cloud Computing platform and utilized in a variety of ways. For example, it can be searched, edited, and used for future insights.

Cloud computing and big data go together, as big data is concerned with storage capacity in the cloud system, cloud computing uses huge computing and storage resources. Thus, by providing big data application with computing capability, big data stimulate and accelerate the development of cloud computing. Cloud Computing providers often utilize a “Software as a Service” model to allow customers to easily process data. Typically, a console that can take in specialized commands and parameters is available, but everything can also be done from the site’s user interface. Some products that are usually part of the package include database management systems, cloud-based virtual machines and containers, identity management systems, machine learning capabilities, and more.

This cloud infrastructure allows for real-time processing of Big Data. It can take huge “blasts” of data from intensive systems and interpret it in real-time. Another common relationship between Big Data and Cloud Computing is that the power of the cloud allows Big Data analytics to occur in a fraction of the time it used to.

VI. BIG DATA & CLOUD COMPUTING: A PERFECT MATCH

The combination of Big Data and Cloud Computing enables spatio-temporal thinking that leads to science discoveries and various application developments with new requirements. Cloud computing provides the source for finding technical and theoretical solutions for Big Data processing. Essentially, there are infinite possibilities when we combine Big Data and Cloud Computing. If we just had Big Data alone, we would have huge data sets that have a huge amount of potential value just sitting there. Using the computers to analyze them would be either impossible or impractical due to the amount of time it would take. However, Cloud Computing allows us to use state-of-the-art infrastructure and also the development which is fueled by Big Data. There would be very fewer cloud-based applications without big data and the data too are collected by cloud-based applications.

To be precise, Cloud Computing services mainly exist because of Big Data. Simlarly, the only reason that we collect Big Data is because we have services that are capable of taking it in and decoding or interpreting it in a matter of seconds. These two are a perfect match, provided neither would exist without the other!

V. CONCLUSION

Big Data gives significant value to organizations who implement cloud in their infrastructure. Merging Big data and cloud technology is an intelligent option to make big data analytics in clouds. The key idea of big data is to collect, handle, visualize, and evaluate the huge amount of data, which is achieved by collaboration with cloud computing. Cloud for Big Data Analytics is also known as Data Analytics as a Service (DAaaS). Finally, it’s important to note that both Big Data and Cloud Computing play a massive role in our digital society. The two linked together allow people with great ideas but limited resources a chance at business success. They also allow established businesses to utilize data that they collect but previously had no way of analyzing.

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”, Februry 2021

[4]. Big data and cloud computing – challenges and opportunities: https://bigdata-madesimple.com/big-data-and-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/data-analytics/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", nternational 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.

[27]. Azhaharasan V, Ramalingam M, Prabhusundhar P, Prabhahari R & K. Yuvaraj, “Classifications of Biometric Security Authentication and Cluster Based Intrusion Detection, Prevention Technique in Mobile Adhoc Network”, International Journal of Scientific Research in Computer Science Applications and Management Studies, ISSN 2319 – 1953, Vol. 7, No. 4, July – 2018.

[28]. Payal Patel & Riddhi Patel, "The Correlation of Cloud Computing and Big data: Review & Benefits", © 2021 JETIR July 2021, Volume 8, Issue 7 (ISSN-2349-5162)

[29]. Big Data & Cloud Computing: The Roles & Relationships: https://www.computer.org/publications/tech-news/trends/big-data-and-cloud-computing

[30]. P. Prabhusundhar & B. Srinivasan, “Multimodal Biometric Based Intrusion Detection System for Clustered Mobile Ad Hoc Network Using POMDP Algorithm”, KASMERA Journal (Science Citation Indexed Journal), ISSN: 0075-5222, Vol. 44, No. 1, June-2016, pp. 2 - 9.

[31]. B. Srinivasan & P.Prabhusundhar, “Stretch and Shrink Method for Security in MANET Using Biometric Based Intrusion Detection System”, International Journal of Scientific Research in Science, Engineering and Technology, ISSN No.2395-1990 (print), Vol. 4, No. 4, April - 2018, pp. 455 – 462.

[32]. T.P. Senthilkumar and P. Prabhusundhar, “An Effective Feature Selection and Incremental Classifier Design for Rice Disease Prediction”, Journal of Advanced Research in Dynamical and Control Systems, (SCOPUS Journal), ISSN NO. 1943 – 023X, Vol. 12, No. 4, March – 2020.