



JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CLOUD COMPUTING CHALLENGE IN BIG DATA ENVIRONMENT

^[1]Dr. Shailja Sharma^[2]Nusrat Perween^[3] Sanjeev Kumar

^[1]Associate Professor^[2]Research Scholar^[3]Research Scholar

^[1]shailja.sharma@aisectuniversity.ac.in ^[2]shailja.sharma@aisectuniversity.ac.in

^[3]shailja1901@rediffmail.com

Abstract- Communicating by using information technology in various ways produces big amounts of data. Such data requires processing and storage. The cloud is an online storage model where data is stored on multiple virtual servers. Big data processing represents a new challenge in computing, especially in cloud computing. Data processing involves data acquisition, storage and analysis. In this respect, there are many questions including, what is the relationship between big data and cloud computing? And how is big data processed in cloud computing? The answer to these questions will be discussed in this paper, where the big data and cloud computing will be studied, in addition to getting acquainted with the relationship between them in terms of safety and challenges. We have suggested a term for big data, and a model that illustrates the relationship between big data and cloud computing.

Keywords: big data, Hadoop, Cloud, MapReduce, resources, Five (Vs).

I. INTRODUCTION

Data is the uncooked material for Data before sorting, arranging and processing. It cannot be used in its primary shape prior to processing. Information represents records after processing and analysis [1]. The technology has been evolved and used in all factors of lifestyles, increasing the call for storing and processing greater information. As an end result, several systems were developed consisting of cloud computing that assist huge records. at the same time as large information is liable for records storage and processing, the cloud gives a reliable, handy, and scalable surroundings for huge records systems to characteristic [2]. large records is defined as the amount of digital data made out of different resources of era for instance, sensors, digitizers, scanners, numerical modeling, mobile phones, net, movies, e-mails and social networks. The facts types include texts, geometries, photos, motion pictures, sounds and combos of each. Such facts may be without delay or in a roundabout way related to geospatial facts [3].

II. BIG DATA

Big Data comes and is composed through electronics operations from a couple of assets. It calls for proper processing strength and high talents for evaluation [9]. The importance of big Data lies inside the analytical use which can help generate a knowledgeable selection to offer better and quicker services [4].

The term Big Data is known as on the massive quantity of excessive-speed massive information of various kinds; these records cannot be processed and stored in normal computer systems. the main traits of massive data, known as V's five As in parent 1, may be summed up inside the truth that the issue isn't always simplest approximately the volume of Data, other dimensions of large records, referred to as 'five Vs', are as follows:

1. Volume: It represents the quantity of records made out of multiple sources which show the huge records in numbers by means of zeta bytes. The extent is most evident dimension in what concerns to massive information.

2. variety: It represents Data sorts, with, growing the wide variety of net customers everywhere, clever telephones and social networks customers, the familiar form of facts has modified from established facts in databases to unstructured data that consists of a big range of formats consisting of photos, audio and videos, SMS, and GPS data [5].

3. pace: It represents the rate of facts frequency from one of a kind assets, that is, the velocity of data manufacturing which include Twitter and facebook. The massive increase in facts extent and their frequency dictates the want for a device that ensures high-quality-pace Data evaluation.

4. Veracity: It represents the best of the data, it shows the accuracy of the records and the self-assurance in the data content. The exceptional of the information captured can range substantially, which affects the accuracy of evaluation. despite the fact that there may be wide agreement on the ability cost of massive information, the records is almost worthless if it isn't correct [6].

5. Fee: It represents the fee of massive data, i.e. it suggests the importance of information after analysis. that is because of the fact that the information on its own is sort of worthless. The value lies in careful analysis of the precise Data, the facts and thoughts it offers. The value is the final level that comes after processing quantity, pace, variety, assessment, validity and visualization.

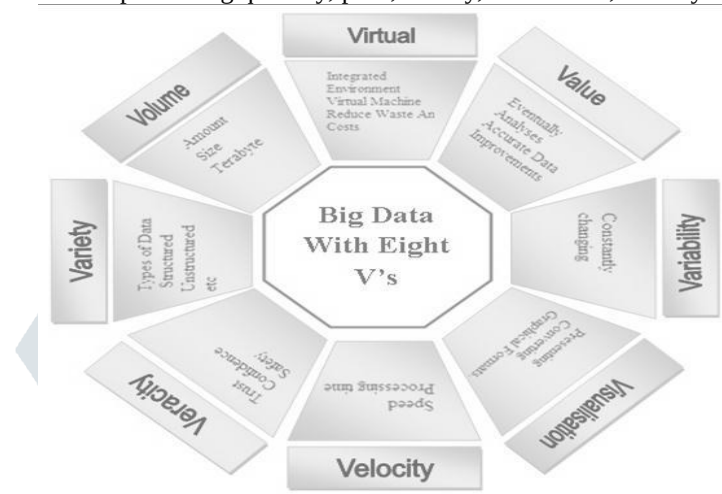


Figure 1. Characteristics of Big Data

There had been several revisions to the big information till they reached (7 v) . on this paper, based on the relationship between cloud computing and large data, will endorse a brand new time period, virtualization, which simply represents The statistics structure is through default. The virtualization of huge information is a technique that focuses on creating digital systems for massive statistics structures. Virtualization technology is the key generation used to help cloud computing handle large amounts of information flexibly and facilitate the procedure of coping with massive statistics.

III. Difference between traditional data and big data

In general, the data in the world of technology is a set of letters, words, numbers, symbols or images, but with the evolution of multitasking technology tools the data has become different in content and source[7]. In light of this, big data emerged which differs from traditional data. Differences between traditional data and big data are shown in Table 1:

Table 1 Comparison between traditional and big data[18]

	Traditional Data	Big Data
Volume	MB and GB	PBs And ZBs
Data Generation Rate	Long periods of time	More rapid
Data Type	Structure	Sim-Structure , Unstructured
Data sources	Centralized	multiple sources, and distributed
Data Store	RDBMS	HDFS, No SQL

IV. Cloud Storage

The concept of cloud storage is the same as that of storing files on a remote server to retrieve them from multiple devices at any time we need. Cloud storage is basically a system that allows storing data on the internet. Examples of this system are Google Drive, Dropbox, etc. [8]. Cloud storage, it is stored data while cloud computing is used to complete the specified digital tasks. In most cloud computing applications, data is sent to remote processors over the internet for complete operation, and the resulting data is sent back where you can use the program interface but the bulk of the program activity is remote instead of the computer. Cloud computing is usually more useful for companies than individuals in most cloud computing applications. It is a set of technologies hosting a cloud, and giving resources to hire and consume on demand over the internet on the basis of pay-per-user. Among the best known cloud computing providers are Amazon, Google, and Microsoft.

The increasing amount of data requires equipment to store them. The cloud provides storage units, making it easier to navigate without having to carry physical storage equipment while on the move. Limited storage space is a real concern for both consumers and businesses. The storage of data in the cloud is done through a cloud service provider (CSP) in a set of cloud servers where the user interacts with the cloud servers via CSP to access or retrieve its data. Since they no longer have their data locally, it is important to assure users that their data is properly stored and maintained. This means that users should be provided with security means so that they can ensure that their stored data is consistently maintained even without local copies.

V. DATABASE MANAGEMENT SYSTEM.

Data is collected in the form of an organized structure called the database which is the food of any information system. Data huge amount is the major component of the cloud infrastructure. Data can be shared among many tenants. As a result, data management in particular is a key aspect of storage in the cloud [9]. Data in the cloud is distributed across multiple sites and may contain certain privileges and authentic information. It is therefore very important to ensure that data consistency, scalability and security are maintained. In order to address these issues and many other important data issues, there is a need for a database management system for cloud data. The database management system shows the mechanism of storage and retrieval of user data with maximum efficiency, taking into consideration the appropriate security policies. The database management system always provides data independence. No change is made to the storage mechanism and shapes without modifying the entire application. There are several types of database organization, relational database, flat database, object oriented database, hierarchical database.

VI. THE RELATIONSHIP BETWEEN THE CLOUD AND BIG DATA

Cloud computing is a trend in the development of technology, as the development of technology has led to the rapid development of electronic information society. This leads to the phenomenon of big data and the rapid increase in big data is a problem that may face the development of electronic information society [10]. 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. The distributed storage technology in environmental computing helps to manage big data.

Cloud computing and big data are complementary to each other. Rapid growth in big data is regarded a problem. Clouds are evolving and providing solutions for the appropriate environment of big data while traditional storage cannot meet the requirements for dealing with big data, in addition to the need for data exchange between various distributed storage locations. Cloud computing provides solutions and addresses problems with big data. The cloud computing environment is expanding to be able to absorb big amounts of data as it follows the policy of data splitting, that is, to store data in more than one location or availability area. Cloud computing

environments are built for general purpose workloads and resource pooling is used to provide flexibility on demand. Therefore, the cloud computing environment seems to be well suited for big data.

Big data processing and storage require expansion as the cloud provides expansion through virtual machines and helps big data evolve and become accessible. This is a consistent relationship between them. Google, IBM, Amazon and Microsoft are examples of the success in using big data in the cloud environment. In order for the cloud environment to fit with big data the cloud computing environment must be modified to suit data and cloud work together. Many changes are needed to be made on the cloud: CPUs to handle big data and others.

VII. Virtual Machine (VM) between the cloud and big data

Virtual Machine (VM) is a software application that simulates a virtual computing environment that can run the operating system (OS) and its associated applications with multiple virtual machines installed on a single machine. Distributed systems, network computing and parallel programming are not new as one of the key enabling factors of the cloud is virtual technology. By using virtualization technology, one virtual machine can often host multiple virtual machines [11]. Virtualization technology provides the ability to reduce workload in virtual metering devices and unify them into one physical server. Consolidation has become particularly effective after the adoption of multi-core CPUs in computing environments, where many virtual machines can be allocated to a single physical node that improves resource utilization and reduces power consumption compared to multi-node setup.

Virtualization technology is the best platform for big data as well as traditional applications. Assuming big data applications simplifies managing your big data infrastructure, providing faster results and is more cost-effective. The role of infrastructure, whether real or virtual, is to support applications. This includes important traditional business applications, modern cloud, and mobile and big data applications. Virtualized big data applications, such as (Hadoop), provide many benefits that cannot be accessed on physical infrastructure but helps simplify big data management. Today's virtual data constitute a wide range of sources including multidimensional stores, web and data services, XML documents, analytical devices, and indoor and outdoor applications. Data stores (NoSQL) are a modern source type where they support virtual data.

Big data and cloud computing point to the convergence of technologies and trends that make IT infrastructure and their applications more dynamic, more modular and more expendable. Currently, the virtual platform building technology is only in the primary stage, which is mainly based on cloud data center integration technology. Cloud computing and big data projects rely heavily on virtualization. Virtual data is the only way to access and improve heterogeneous environments, such as environments used in big data projects. The cloud computing model allows users to have a default data center that can access data sets that were not previously available by using a shared (API) for disparate data sets.

VIII. Big data Security in cloud computing

Big data and cloud are among the most important stages of IT development. Information privacy and security are one of the most important issues for the cloud because of its open environment with very limited user control. Security and privacy affect big data storage and processing because there is a huge use of third party services and the infrastructure used to host important data or to perform operations as growing data and application growth bring challenges.

A solution is provided for the security services and the level of confidence needed through the third party services within the cloud. The data is stored in a central location known as the cloud storage server, where the data is processed somewhere on the servers, so the client has confidence in the service provider as well as data security. The service level agreement must be standardized to gain trust between service providers and customer. The security of cloud client data varies in protection requirements. Customers require protection of their data only through basic logical access controls, while intellectual property, structured or classified data are confidential and require advanced security controls including encryption, data hiding, login, logging, etc..

The Service Level Agreement (SLA) reflects a service level contract between the user and the service provider. It is one of the ways to enhance the security level, where different levels and complexities of security are determined depending on services to better understand security policies for a cloud consumer, and to protect data. There are rules with service level agreements to protect the data, capacity, scalability, security, privacy, and availability of issues such as data storage and data growth. The technologies available to secure big data, such as registry entry, encryption, and trap detection are essential. In many organizations, big data analytics can be used to detect and prevent malicious hackers and advanced threats. The security of big data in cloud computing is necessary because of the following issues:

1. Protection of big data from malicious intruders and advanced threats.
2. Knowledge about how cloud service providers securely maintain huge disk space and erase existing big data.
3. Lack of standards for checking and reporting big data in the public cloud.

IX. Challenges in big data and cloud computing

The security challenges in cloud computing environments fall under several levels: the network level which includes dealing with network protocols and network security such as distributed nodes, distributed data, and communications between the nodes; authentication level where the user handles encryption / decryption techniques, authentication methods such as contract administrative rights, authentication of applications and nodes, and logging entry; the data level which is concerned with data integrity and availability as well as data protection and data distribution. Cloud computing follows the policy of shared resources, where the privacy of data is very important because it faces some challenges like integrity, authorized access, and availability of (backup / replication). Data integrity ensures that data is not corrupted or tampered with during communication. Authorized access prevents data from infiltration attacks while backups and replicas allow access to data efficiently even in case of technical error or disaster in some cloud location.

Big data face some challenges as they can be classified into groups: data sets, processing and management challenges. When dealing with big amounts of data we face challenges such as volume, variety, velocity and verification which are also known as 5V of big data.

Also, in the field of computer networks the cost of communications is a major concern compared to the cost of processing the same data, as the challenge is to reduce the cost of communications to the minimum while meeting the requirements of storage and additional data from the general cloud to handle big data [12]. Among the factors and challenges that affect the processing of big data in a timely manner is the bandwidth and latency. Where several challenges can be summarized in the relationship between big data and cloud computing.

Data Storage: The storage of big data through traditional storage is problematic because hard drives often fail, data protection mechanisms are not effective, and the speed of big data requires storage systems in order to expand rapidly, which is difficult to achieve with conventional storage systems. Cloud storage services offer almost unlimited storage with a great deal of error tolerance, which offers potential solutions to address the challenges of big data storage.

Variety of data: Big data naturally grow, increase and vary, which is the result of the growth of almost unlimited sources of data. This growth leads to the heterogeneous nature of big data. Generally speaking, data from multiple sources of different types and representations are highly interrelated. They have incompatible shapes and are inconsistent. A user can store data in structured, semi-structured or unstructured formats. Structured data format is suitable for today's database systems, while semi-structured data formats are only fairly suitable. Unstructured data is inappropriate because it contains a complex format that is difficult to represent in rows and columns.

Data transfer: The data goes through several stages: data collection, input, processing, and output. Big data transfer is a challenge, so data compression techniques need to be reduced to reduce the volume, where data volume is a hindrance to transfer speed. It also affects the cost, while cloud computing provides distributed storage resources and data transfer on high-speed lines, reducing costs through virtual resources and resource use at user's request.

Privacy and data ownership: The cloud environment is an open environment and the user's role in monitoring is limited. Privacy and security are an important challenge for big data. Big data and cloud computing come together in practice. According to (IDC) estimates, by 2020, around 40% of global data will be accessed by cloud computing. Cloud computing provides strong storage, calculation and distribution capability to support big data processing. As such, there is a strong demand to investigate the privacy of information and security challenges in both cloud computing and big data.

X. Big Data's Relationship To The Cloud

How does the cloud computing environment correspond to big data? The answer to this question reflects the relationship between them. This is done through the cloud computing features to handle big data, the resources provided by cloud computing, the resource service to provide service to many users where the various physical and virtual resources are automatically set and reset upon request. Cloud computing has access from anywhere to data resources that are spread all over the world by using a (public) cloud to allow those sources faster access to storage. The nature of big data is generated by technologies and locations worldwide, so the cloud resource service provides and helps in the collection and storage of big amounts of data resulting from the use of technologies.

Cloud computing offers features and benefits to big data through ease of use, access to resources, low cost in resource utilization on supply and demand, and reduces the use of solid equipment used to handle big data. Both big data and the cloud aim to increase the value of a company while reducing investment costs. The cloud reduces the cost of managing local software, while big data reduces investment costs by encouraging more prudent business decisions. It seems only natural that these two concepts together provide greater value to companies.

Any system in technology must pass through several main stages. The computer system follows the input, processing and output model. Input is done through devices and then processed through the CPU. Thus, the results of the information are produced. In the relationship between the data and cloud computing, the data is stored on external and remote storage units. On the other hand, in the computer system, the data is stored internally or locally. Therefore, the relationship between the data and cloud computing represents the input, processing and output model as in Figure 3. The big data is entered through devices such as the mouse, cellular devices and other smart devices. Processing is carried out through the tools and techniques used by the cloud computing in providing service, and the outputs are the results, it represents the value of data after processing.

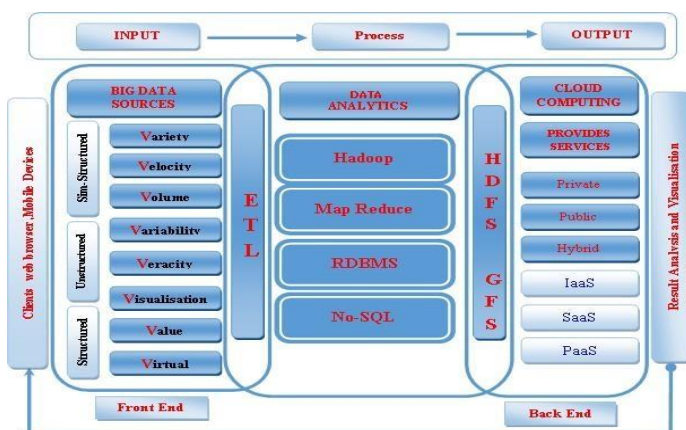


Figure 3 . A Model Showing The Relationship Between Big Data And Cloud Computing

The input and output model defines input, output and processing tasks required to convert input to output. Inputs represent the flow of data and raw materials. The processing step includes all tasks required to transform inputs. The output is data flowing from the transformation process.

XI. Common factor between cloud computing and big data.

The internet of things represents the new concept of the Internet network, which enables communication between several parties to communicate together, and these parties include smart devices, mobile devices, sensors and other [13] where it is considered effective communication between all elements of architecture so that it can Rapidly deploy applications, process and analyze data quickly to make decisions as quickly as possible.

The architecture represents several systems: objects, gates, network infrastructure, cloud infrastructure. [14] Internet objects can benefit from the scalability and performance of cloud computing infrastructure. In fact, Internet applications produce large amounts of data and consist of multiple computer components upon request.

The Internet of Things (IOT) is going to generate a massive amount of data and this in turn puts a huge strain on Internet Infrastructure. As a result, this forces companies to find solutions to minimize the pressure and solve their problem of transferring large amounts of data. . But cloud computing has played a major role in IT, by migrating its data operations to the cloud. Many cloud providers can allow your data to either be transmitted over your traditional Internet connection or via a dedicated direct link. That the real purpose of cloud computing and Internet of things increase efficiency in daily tasks and both have a complementary relationship. The Internet of things generates huge amounts of data, and cloud computing provides a pathway for these data to navigate. By storing data in the cloud, most companies find that it is possible to access large amounts of big data through the cloud. And internet of things is all parts of a continuum. Difficult to think of Internet things without thinking about the cloud, it is difficult to think of the cloud without thinking about the Big data analyzes. Which generates a lot of data, this data is stored in the cloud computing, cloud computing is the only technology suitable for filtering, analysis, storage and access to IoT and other data in ways that are useful, as these data constitute large quantities must be analyzed, Objects is a common factor between the erased cloud and big data.

XII. Common points between big data and the cloud

The cloud computing environment consists of several user terminals and service provider. The big data comes on both sides, as the user collects the data and, in dealing with the technology tools, the big data is produced. The role of the service provider is to save, store and process the big data at the user's request, so cloud computing represents the big data infrastructure. The service provider must ensure that users have on-demand resources or otherwise access their data, systems and applications on a regular basis and is available throughout the service without any interruption.

Data, whether small or big, require storage, processing and security, but the volume and capacity of data requirements differ in accordance with the volume of the data, so cloud computing must provide storage, processing and security requirements for big data in its environment. The cloud environment is scalable and uses sophisticated high-end data management techniques and security policies as the service provider protects and manages data.

Cloud computing provides security, depending not on data volume but the availability of security and protection for small and big data. The service provider guarantees complete confidentiality of user data of all kinds and only allows access to authorized users. Therefore, identity management and access control must be provided for information resources and service resources, according to user needs. The user can connect to the network in these resources through a simple software interface that simplifies and ignores many internal details and processes.

Cloud computing saves the cost of storing and processing data to the user through the availability of geographically dispersed servers and the availability of virtual server technology. The service provider must ensure that the devices and equipment are sufficiently available, and restricted by an integrated and documented entry system for reference when needed. Cloud computing offers the use of high-level applications and software, regardless of the efficiency of the devices the user uses, because it depends on the strength of the network servers and not on the personal resources of your device, regardless of the efficiency of the user's device he can benefit from the cloud service.

CONCLUSION

Big data and cloud computing have been studied from several important aspects, and we have concluded that the relationship between them is complementary. Big data and cloud computing constitute an integrated model in the world of distributed network technology. The development of big data and their requirements is a factor that motivates service providers in the cloud for continuous development, because the relationship between them is based on the product, the storage and cloud for continuous development, because the relationship between them is based on the product, the storage and processing as a common factor. Big data represents the product and the cloud represents the container. The big data is concerned with the capacities of cloud computing. On the other hand, cloud computing is interested in the type and source of big data. Compatibility between them is summarized in Table 2. Cloud computing represents an environment of flexible cloud for continuous development, because the relationship between them is based on the product, the storage and concerned with the capacities of cloud computing. On the other hand, cloud computing is interested in the type and source of concerned with the capacities of cloud computing. On the other hand, cloud computing is interested in the type and source of big data. Depending on the relationship between them, a model was prepared to show the relationship between them as in Figure 3. Compatibility between them is summarized in Table 2. Cloud computing represents an environment of flexible distributed resources that uses high techniques in the processing and management of data and yet reduces the cost. All these cloud for continuous development, because the relationship between them is based on the product, the storage and processing as a common factor. Big data represents the product

and the cloud represents the container. The big data is concerned with the capacities of cloud computing. Depending on the relationship between them, a model was prepared to show the relationship between them.

Characteristics show that cloud computing has an integrated relationship with big data. Both are moving towards rapid progress to keep pace with progress in technology requirements and users.

REFERENCES:

1. Kshetri, Nir. "Cloud computing in developing economies." *Computer* 43, no. 10 (2010): 47-55.
2. https://en.wikipedia.org/wiki/Cloud_computing
3. Klous, Sander, and Nart Wielaard. *We are Big Data: The Future of the Information Society*. Springer, 2016.
4. <https://www.internetworldstats.com/stats.htm>
5. <https://www.ibm.com/big-data/us/en/> Bello-Orgaz G, Jung JJ, Camacho D. Social big data: Recent achievements and new challenges. *Information Fusion*.2016 Mar 31;28:45-59.
6. Boyd, D., & Crawford, K. (2011, September). Six provocations for big data. In *A decade in internet time: Symposium on the dynamics of the internet and society* (Vol. 21). Oxford: Oxford Internet Institute.
7. SHAN, Y. C., Chao, L. V., ZHANG, Q. Y., & TIAN, X. Y. (2017). Research on Mechanism of Early Warning of Health Management Based on Cloud Computing and Big Data. In *Proceedings of the 23rd International Conference on Industrial Engineering and Engineering Management 2016* (pp. 291-294). Atlantis Press, Paris.
8. Parvin Ahmadi Doval Amiri and Mina Rahbari Gavvani, 2016. A Review on Relationship and Challenges of Cloud Computing And Big Data: Methods of Analysis and Data Transfer. *Asian Journal of Information Technology*, 15: 2516-2525
9. Chen, Min, et al. *Big data: related technologies, challenges and future prospects*. Heidelberg: Springer, 2014.
10. Demchenko, Yuri, et al. "Big security for big data: Addressing security challenges for the big data infrastructure." *Workshop on Secure Data Management*. Springer, Cham, 2013. *Environments and evaluation of resource provisioning algorithms.* Software: Practice and experience 41.1 (2011): 23-50.
11. R. Subhulakshmi, S. Suryagandhi, R. Mathubala, P. Sumathi, An evaluation on Cloud Computing Research Challenges and Its Novel Tools, *International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)* Volume 2, Special Issue 19, October 2016.
12. Fonseca, N., & Boutaba, R. (2015). *Cloud services, networking, and management*. John Wiley & Sons.

