International Journal of Wireless Communications and Mobile Computing
2018; 6(3): 43-48
http://www.sciencepublishinggroup.com/j/wcmc
doi: 10.11648/j.wcmc.20180603.11
ISSN: 2330-1007 (Print); ISSN: 2330-1015 (Online)

SciencePG
Science Publishing Group

Research/Technical Note

Futuristic Correlation of Big Data & Cloud Computing in the New Millennium

Shallu, Sanjay Chaudhary

Department of Computer Science & Engineering at Madhav University, Rajasthan, India

Email address:
dr.bathla@gmail.com (Shallu), schaudhary00gmail.com (S. Chaudhary)

To cite this article:
Shallu, Sanjay Chaudhary. Futuristic Correlation of Big Data & Cloud Computing in the New Millennium. *International Journal of Wireless Communications and Mobile Computing*. Vol. 6, No. 3, 2018, pp. 43-48. doi: 10.11648/j.wcmc.20180603.11

Received: January 19, 2019; **Accepted**: February 21, 2019; **Published**: March 15, 2019

Abstract: Exchange a few words by using information technology in a variety of ways manufacture big amounts of data. Such data requires dealing out and storage. The cloud is an online storage space model where data is stored on multiple virtual servers. Big data processing represents a new face up to in computing, particularly in cloud computing. Data processing involves data acquirement, storage and analysis. In this respect, there are many questions including, what is the connection between big data and cloud computing? And how is big data processed in cloud computing? The answer to these difficulties will be talk about in this paper, where the big data and cloud computing will be studied, in adding up to receiving acquainted with the relationship between them in terms of safety and challenges. It recommends a period for big data, and a model that illustrates the relationship between big data and cloud computing.

Keywords: Big Data, Hadoop, Map Reduce, Resources, Five (Vs), Cloud Computing

1. Introduction

Data is the underdone objects for information prior to sorting, assemble and processing. It cannot be used in its chief form prior to processing. Information represents data after processing and analysis. The technology has been residential and used in all aspects of life, increasing the demand for storing and processing more data. As a result, several systems have been developed including cloud computing that support big data. While big data is answerable for data storeroom and processing, the cloud provides a reliable, easy to get and scalable atmosphere for big data systems of function. Big data is defined as the capacity of digital data created from different sources of technology for example, antennas, digitizers, scanners, mathematical modeling, mobile phones, Internet, videos, e- mails and social networks. The data types embrace texts, geometries, images, videos, sounds and combinations of each. Such data can be in a straight line or indirectly related to geospatial information. Cloud computing refers to on-demand computer supplies and systems obtainable across the network that can provide a number of incorporated computing services without local resources to make easy user access. These resources contain data storage capacity, backup and self-bringing together. Most IT Infrastructure computing consists of services that are making available and delivered through public centers and servers based on them. Here, clouds appear as individual access points for the calculate needs of the consumer. It is normally expected for profitable offers to meet the QoS requirements of customers or consumers, and naturally include service level agreements (SLAs). They are an online storage space model where data are stored on multiple virtual servers, rather than being hosted on a specific server, and are usually provided by a third party. The hosting companies, which have highly developed data centers, rent spaces that are stored in a cloud to their purchasers in line with their needs. [1-4].

2. Big Data

Big data approaches and is created through electronics operations from multiple sources. It necessitates proper processing power and high capabilities for analysis. The significance of big data lies in the systematic use which can help produce an informed judgment to provide better and more rapidly services. The term big data is called on the huge amount of high-speed big data of different types; this data cannot be processed and store in accepted computers. The

main characteristics of big data, called V's 5 As in Figure 1, can be summed up in the fact that the issue is not only about the volume of data, other dimensions of big data, known as 'five Vs', are as follows:[9-10]

2.1. Volume

It symbolizes the quantity of data produced from numerous sources which illustrate the huge data in numbers by zeta bytes. The volume is most evident dimension in what concerns to big data.

2.2. Variety

It symbolizes data types, with, increasing the number of Internet users everywhere, smart phones and group of people networks users, the recognizable form of data has changed from structured data in databases to unstructured data that includes a large number of formats such as images, audio and video clips, SMS, and GPS data.

2.3. Velocity

It characterizes the speed of data frequency from dissimilar sources, that is, the speed of data production such as Twitter and Face book. The huge increase in data volume and their regularity read aloud the need for a system that ensures super-speed data analysis.

2.4. Veracity

It represents the quality of the data. It demonstrates the accurateness of the data and the self-assurance in the data content. The quality of the data incarcerate can vary greatly, which affects the accuracy of analysis. Although there is wide conformity on the forthcoming value of big data, the data is almost worthless if it is not accurate [11-12].

2.5. Value

It represents the value of big data i.e. it shows the importance of data after analysis. This is unpaid to the fact that the data on its own is approximately insignificant. The value lies in cautious analysis of the exact data, the information and ideas it provides. The value is the final stage that comes after processing volume, velocity, variety, contrast, validity and visualization. There have been numerous revisions to the big data until they reached. In this paper, based on the connection between cloud computing and big data, will recommend a new term, virtualization, which practically represents. The data structure is by default. The virtualization of big data is a progression that focuses on creating virtual structures for big data systems. Virtualization technology is the key equipment used to help cloud computing grip large amounts of data flexibly and facilitate the process of managing big data. [13-14].

3. The Type and Nature of the Data

Data in general is a set of values that are in the form of numbers, letters, and symbols and other forms where they are concerned with a particular idea and subject. The data does not make sense without analysis, and is, therefore, compiled for use. It represents input while information is output after processing, i.e. data is entered into the system first, and then processed until it comes out in the form of useful information that has a clear meaning and against which decisions are made. Big data come from multiple sources including sensors and free texts such as social media, unstructured data, metadata and other geospatial data collected from web logs, GPS, medical devices, etc. The big data is gathered from different sources, so it is in several forms, including.

3.1. Structured Data

It is the organized data in the form of tables or databases to be processed.

3.2. Unstructured Data

It characterizes the leading proportion of data; it is the data that people produce daily as texts, images, videos, messages, log records, click-streams etc.

3.3. Semi-structured Data

It is regarded a kind of structured data but not deliberate in tables or databases, for example XML documents or JSON [15-17].

4. Differentiation Between Conventional Data and Big Data

In broad-spectrum, the data in the globe of technology is a set of letters, words, numbers, symbols or images, but with the improvement of multitasking technology tools the data has become different in content and foundation. In light of this, big data become visible which differs from conventional data. Differences between traditional data and big data are shown in Table 1:

Table 1. *Comparison between traditional and big data.*

	Traditional Data	**Big Data**
Volume	MB and GB	PBs and ZBs
Data production Rate	Extended Period of Time	More speedy
Data Type	Structure	Semi Structure, Unstructure
Data Sources	Centralized	Multiple sources, and distributed
Data Store	RDBMS	HDFS, No SQL

5. Cloud Computing

It is a period that refers to on-demand computer supplies and systems that can provide a number of incorporated computer services without being bound by local resources to facilitate user access. These resources include data storage, backup and self-bringing together, as well as software processing and scheduling tasks. Cloud computing is a shared resource system that can offer a variety of online services such as virtual server storage, and applications and licensing for desktop applications. By leveraging common resources, cloud computing is able to achieve expansion and provide volume [19-20].

5.1. Characteristics of Cloud Computing

That cloud computing is one of the distributed systems that correspond to a complicated model. NIST has identified important facet of the cloud, as it shortened the concept of cloud computing in five characteristics as follows:

5.2. On-demand Self-service

Cloud services provide computer resources such as storage and processing as needed and without any human interference.

5.3. Broad Network Access

Cloud computing resources are easy to get to over the network, mobile and elegant devices even sensors can access computing resources on the cloud.

5.4. Resource Pooling

Cloud proposal users allocate a vast array of computing resources; users can settle on the nature of resources and the geographic location they prefer but cannot determine the exact physical location of these resources.

5.5. Rapid Elasticity

Resources from storage media, network, dispensation units and applications are always available and can be increased or decreased in an almost immediate fashion, allowing for high scalability to ensure most favorable use of resources.

5.6. Measured Service

Cloud systems can appraise the processes and expenditure of resources as well as observation, have power over and coverage in a completely transparent manner [21-22]

6. Cloud Computing Service Models

Cloud computing types are confidential on the basis of two models: cloud computing service models and cloud computing consumption models as in Figure 2:

6.1. Software as a Service (SAAS)

Cloud examination providers provide a variety of software applications to users who can use them without install them on their computer. The user is not answerable for whatever thing other than adjusting the settings and customizing the service as correct to his needs. SAAS helps big-data clients to perform data.

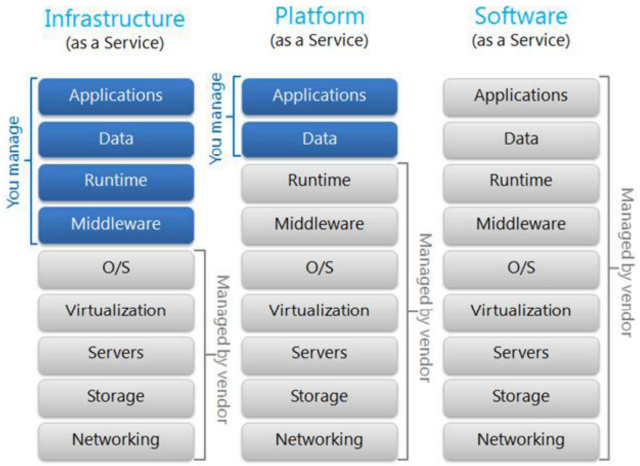

Figure 1. *Cloud computing model.*

6.2. Platform as a Service (PAAS)

Cloud service providers provide platforms, tools and other services to users, where the cloud examination provider manages everything else, including the operating system and middleware with resources that enable you to distribute the whole thing from simple cloud-based appsto sophisticated.

6.3. Infrastructure as a Service (IAAS)

Cloud service providers provide infrastructure such as storage, computing ability, etc. is a form of cloud computing that provides virtualized computing resources greater than the Internet, In an IaaS model, a third-party contributor hosts hardware, software, servers, storage and other infrastructure components on behalf of its users

6.4. DaaS

It is the substitute cloud computing model, as it be at variance from conventional models like (SAAS, IAAS, PAAS) in make available that data to users through the network, as data is considered the value of this model in juxtaposition with cloud computing based on solving some of the challenges in managing a huge amount of data. For these reasons, DaaS is personally related to big data whose technologies must be utilized. DaaS provides highly efficient methods of data sharing and processing. DaaS is closely related to SaaS (storage as a service) and SaaS (software as a service) which can be mutual with one of these models or both of them.

7. The Correlation Between the Cloud and Big Data

Cloud computing is a development in the growth of

technology, as the development of technology has show the way 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. Cloud computing and big data go together, as big data is concerned with storage capacity in the cloud system, cloud computing uses massive computing and storage resources. Thus, by providing big data application with computing capability, big data encourage and pick up the pace the development of cloud computing. The scattered storage technology in environmental computing helps to manage big data. Cloud computing and big data are corresponding to each other. Rapid growth in big data is regarded a problem. Clouds are embryonic and as long as solutions for the suitable environment of big data while conventional 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 explanations and addresses problems with big data. The cloud computing environment is getting higher to be able to attract big amounts of data as it follows the guiding principle of data splitting, that is, to store data in excess of one position 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 atmosphere seems to be well suited for big data. Big data dispensation and storage necessitate spreading out as the cloud provides expansion through virtual machines and helps big data evolve and become easy to get to. This is a regular association between them. Google, IBM, Amazon and Microsoft are examples of the accomplishment 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 required to be made on the cloud: CPUs to handle big data and others. [17-19].

8. Big Data Security in Cloud Computing

Big data and cloud are surrounded by 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 protection 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. The protection of big data in cloud computing is compulsory because of the following issues:

1. Protection of big data from malevolent 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 examination and reporting big data in the public cloud.

9. What Is Big Data's Association to the Cloud Computing

How does the cloud computing situation correspond to big data? The answer to this inquiry reflects the connection sandwiched between them. This is done through the cloud computing features to handle big data, the resources provided by cloud computing, the resource check to provide service to many users where the variety of physical and virtual possessions are robotically set and reset upon request. Cloud computing has right of entry 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 consequential from the use of technologies.

The cloud computing structure can spread out the solid equipment to accommodate small and big data volumes. The cloud can expand to handle big amounts of data by dividing the data into parts, automatically done in IAAS. Expanding the environment is a big data requirement. Cloud computing has the advantage of helping to reduce costs by paying for the value of the resources used, which helps to develop big data. Flexibility is also regarded a requirement for big data. When it needs more storage for data the cloud platform can enthusiastically expand to meet proper storage needs when would like to handle a large number of virtual machines in a single time period. For error tolerance, the cloud helps to handle big data in the extraction and storage process. Error tolerance helps SLAs, as well as QOS levels. Service level agreements specify different rules for adaptable availability of cloud service. Big companies such as Yahoo, Google, Face book, and others offer web-based services, and the amount of data they routinely collect through online user exchanges has overwhelmed traditional IT capabilities. Therefore, the development of basic infrastructure components has to be developed. Apache Hadoop has been introduced as a realistic point of reference for managing big amounts of unstructured data. Apache Hadoop is open platform distributed software for storing and processing data. By using Hadoop, you can dependably store big amounts (pet bytes) on tens of thousands of servers while effectively scaling performance in terms of cost. Map Reduce is based on the distribution of a data set between multiple servers, partial results are then reassembled. Big data are characterized by diversity, i.e. they are of different types and therefore require big data. ETL technology, therefore, deals with data diversity, as ETL represents several

functions such as extraction, conversion, and loading. These three functions are combined into one tool to pull data from one database and place it in another database. It helps to convert databases from one form to another. Big data relies on data truthfulness to be effective. If you store big data at the local level, it will take a huge amount of work to physically merge all data to manage it. The cloud can do this work for the user, offering one site to store and manage all commercial data. In this way, you can get one source of the truth, without fatiguing your time and resources to manually merge the data.

Cloud computing recommends attributes and benefits to big data through ease of use, access to resources, low cost in resource consumption on supply and demand, and reduces the use of concrete 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 direction local software, while big data reduces investment costs by heartening 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, handing out 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 accumulating internally or locally. Therefore, the relationship between the data and cloud computing represents the input, dispensation and output model as in Figure 2. 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.

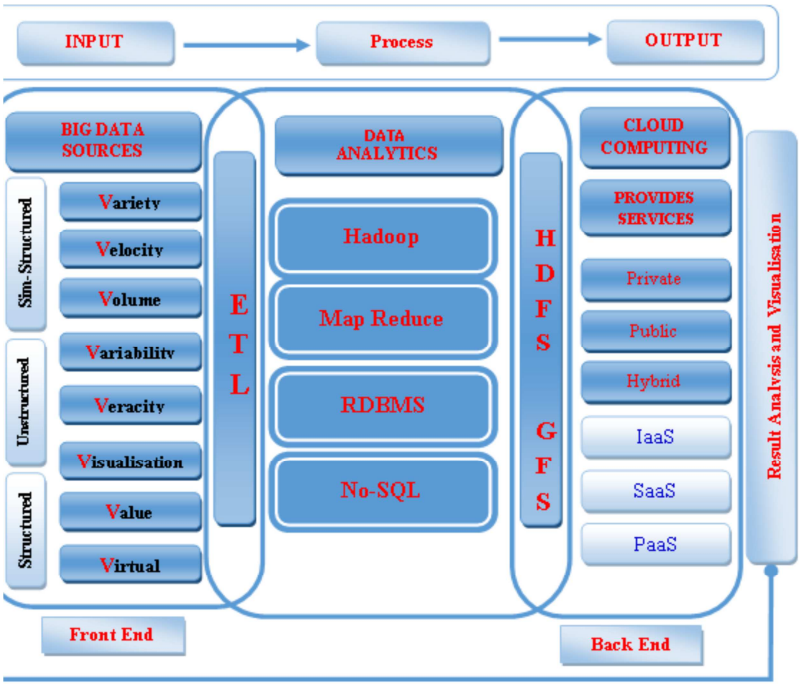

Figure 2. *A Model Showing the Relationship between Big Data and Cloud Computing.*

10. Conclusion

Big data and cloud computing have been universally from more than a few important aspects, and have concluded that the relationship between them is complementary. Big data and cloud computing comprise an integrated model in the world of distributed network technology. The development of big data and their commitments is a factor that motivate service providers in the cloud for continuous development, because the affiliation 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 frightened 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 geared up to show the relationship between them as in Cloud computing represents an environment of flexible distributed resources that use high techniques in the giving out and management of data and yet squeeze the cost. All these characteristics show that cloud computing has an integrated relationship with big data. Both are affecting towards rapid progress to keep pace with progress in technology requirements and users.

References

[1] Charmaz, K., and A. Bryant. "The SAGE Handbook of Grounded Theory: Paperback Edition." (2010).

[2] Neves, Pedro Caldeira, Bradley Schmerl, Jorge Bernardino, and Javier Cámara. "Big Data in Cloud Computing: features and issues."

[3] Klous, Sander, and Nart Wielaard. are Big Data: The Future of the Information Society. Springer, 2016.

[4] 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.

[5] 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.

[6] Parvin Ahmadi Doval Amiri and Mina Rahbari Gavgani, 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.*

[7] Chen, Min, et al. Big data: related technologies, challenges and future prospects. Heidelberg: Springer, 2014.

[8] 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.

[9] McAfee, Andrew, and Erik Brynjolfsson. "Big data: the management revolution." Harvard business review 90.10 (2012): 60-68.

[10] Liebowitz, J. (Ed.). (2014). *Bursting the big data bubble: The case for intuition-based decision making*. CRC Press.

[11] Sremack, Joe. Big Data Forensics–Learning Hadoop Investigations. Packt Publishing Ltd, 2015.

[12] Franks, Bill. Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics. Vol. 49. John Wiley & Sons, 2012.

[13] Furht, Borko, and Flavio Villanustre. Big Data Technologies and Applications, Chapter 1, Springer, 2016.

[14] Calheiros, Rodrigo N., et al. "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms." Software: Practice and experience 41. 1 (2011): 23-50.

[15] 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.

[16] Fonseca, N., & Boutaba, R. (2015). Cloud services, networking, and management. John Wiley & Sons.

[17] https://www.ibm.com/blogs/cloud-computing/2014/01/cloud-computing-defined-characteristics-service-levels/

[18] Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: state-of-the-art and research challenges. Journal of internet services and applications, 1 (1), 7-18.

[19] Ahmed, F. F. (2015). Comparative Analysis for Cloud Based e-learning. Procedia Computer Science, 65, 368-376.

[20] Vacca, J. R. (Ed.). (2016). Cloud Computing Security: Foundations and Challenges. CRC Press. ch-15.

[21] https://support.rackspace.com/how-to/understanding-the-cloud-computing-stack-saas-paas-iaas/

[22] Terzo, O., Ruiu, P., Bucci, E., & Xhafa, F. (2013, July). Data as a service (DaaS) for sharing and processing of large data collections in the cloud. In Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on (pp. 475-480). IEEE.