

## Characteristics and Issues of Big Data in Cloud Computing

Poonam Rani<sup>1</sup>, Neha<sup>2</sup>

<sup>1</sup>Assistant Professor, Department of Computer Science,  
Sanatan Dharma College, Ambala Cantt

<sup>2</sup>Assistant Professor, Department of Electronics & IT,  
Sanatan Dharma College, Ambala Cantt

### ABSTRACT:

The term "big data" was coined in response to the rapid growth of global data as a technology capable of storing and processing big and heterogeneous volumes of data, offering both businesses and science with deep insights into their clients/experiments. Cloud computing offers a secure, fault-tolerant, scalable, and available framework for big data distributed management systems. In this article, we provide an overview of both technology and successful integration scenarios for big data and cloud frameworks. Despite the fact that big data solves many of our current problems, it still has some weaknesses and issues that need to be addressed. Other obstacles, such as security, privacy, scalability, data governance policies, data heterogeneity, and disaster recovery mechanisms, have yet to be addressed. There are other concerns as well. This paper presents an overview of both cloud and big data technologies describing the current issues with these technologies.

**KEYWORDS:** Big Data, Cloud Computing, Big Data Issues

**1. INTRODUCTION:** In current era, there has recently been an increasing demand to store and even process more and more data, inside domains such as finance, science, and even government. Systems which support big data, and host these people using cloud computer, are already developed and even used successfully (Hashem et al., 2014)[1].Whereas big data is in responsible for storing and processing data, Cloud offers a reliable, fault-

tolerant, available and scalable environment so that big database systems can perform (Hashem et al., 2014). Big data analytics viewed both scientific and business data as a way to correlate data, find patterns and predict new trends. To fulfill the current requirements, big data systems must be scalable, flexible and fault tolerant [2]. In this paper, we describe both features and issues of big data in cloud computing.

## **2. BIG DATA & CLOUD COMPUTING :**

In digital world, information are created from different sources and the quick change from computerized advancements has prompted development of big data. In general, it refers to the assortment of enormous and complex datasets which are challenging to deal with traditional database management tools or data processing applications [3]. These are accessible in organized, semi-organized, and unstructured arrangement in petabytes and beyond.

Cloud computing delivers all these through hardware virtualization. Thus, big data and cloud computing are two compatible concepts as cloud enables big data to be available, scalable and fault tolerant.

Business regards big data as a valuable business opportunity. As such, several new companies such as Cloudera, Hortonworks, Teradata and many others, have started to focus on delivering Big Data as a Service (BDaaS) or DataBase as a Service (DBaaS). Companies such as Google, IBM, Amazon and Microsoft also provide ways for consumers to consume big data on demand.

## **3. CHARACTERISTICS OF BIG DATA IN CLOUD COMPUTING**

Big Data contains a large amount of data that is not being processed by traditional data storage or the processing unit.

It is used by many multinational companies to process the data and business of many organizations. The data flow would exceed 150 hexabytes per day before replication. There are five v's of Big Data that explains the characteristics as shown below in Figure 1. The name Big Data itself is connected with a enormous size. Big Data is a tremendous 'volumes' of information created from many sources every day, for example, business processes, machines, web-based social media platforms, organizations, human collaborations, and others many more. Facebook can produce approximately a billion messages, 4.5 billion times that the "Like" button is recorded, and in excess of 350 million new posts are uploaded every day. Variety means Big Data can be structured; unstructured, quasi-structured

and semi-structured that is being collected from different sources [4]. But these days the data will comes in array forms, that are PDFs, Emails, audios, SM posts, photos, videos, etc. Veracity means how much the data is reliable. It has many ways to filter or translate the data. Veracity is the process of being able to handle and manage data efficiently. Big Data is also essential in business development. Veracity means how much the data is reliable. It has many ways to filter or translate the data. Veracity is the process of being able to handle and manage data efficiently. Big Data is also essential in business development. Value is an essential characteristic of big data. It is not the data that we process or store. It is valuable and reliable data that we store, process, and also analyze. Velocity plays an important role compared to others. Velocity creates the speed by which the data is created in real-time [5]. It contains the linking of incoming data sets speeds, rate of change, and activity bursts. The primary aspect of Big Data is to provide demanding data rapidly.

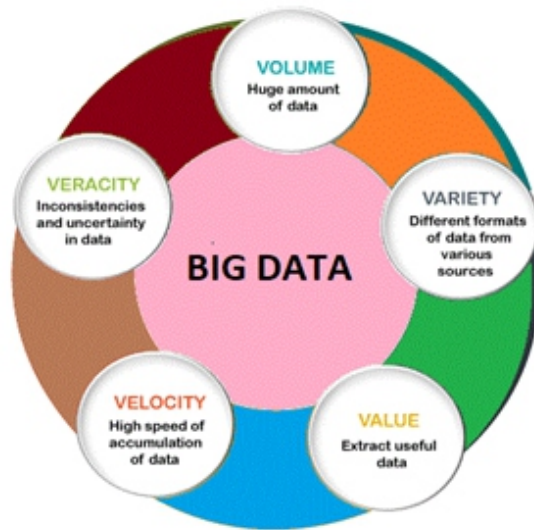


Figure 1

#### 4. ISSUES OF BIG DATA IN CLOUD COMPUTING

Although big data solves many current problems regarding high volumes of data, it is a constantly changing area that is always in development and that still poses some issues. In this section we present some of the issues not yet addressed by big data and cloud computing [6].

As the amount of data grows at a rapid rate, keeping all data is physically cost-ineffective. Therefore, corporations must be able to create policies to define the life cycle and the expiration date of data. Moreover, they should define who accesses and with what purpose clients' data is accessed. As data moves to the cloud, security and privacy become a concern that is the subject of broad research. Big data DBMSs typically deal with lots of data from several sources and as such heterogeneity is also a problem that is currently under study. In this paper we provide an overview over these issues [8].

- **Security:** Cloud computing and big data security is a current and critical research topic (Popović & Hocenski, 2015)[7]. This problem becomes an issue to corporations when considering uploading data onto the cloud. Questions such as who is the real owner of the data, where is the data, who has access to it and what kind of permissions they have are hard to describe. Corporations that are planning to do business with a cloud provider should be aware. Securing this large amount of information is one of the overwhelming difficulties of Big Data. Often organizations are so busy in understanding, storing and examining their informational collections that they push data security for later stages.
- **Privacy:**  
Ensuring data security and protecting privacy has become extremely difficult as information is spread and replicated around the globe.
- **Lack of proper understanding of Big Data**  
Organizations fail in their Big Data drives because of insufficient understanding. Representatives may not know what information is, its capacity, processing, significance, and sources. Data experts might realize what is happening, however others might not have a clear picture.
- **Data growth**  
The amount of data being stored in data centers and databases of companies is increasing rapidly. As these data sets grow exponentially with time, it gets extremely difficult to handle. Most of the data is unstructured and comes from documents, videos, audios, text files and other sources. This means that you cannot find them in databases.
- **Confusion while Big Data tool selection**  
Companies regularly become confused while choosing the best tool for Big Data investigation and capacity. Is HBase or Cassandra the best innovation for data storage? Is Hadoop Map Reduce adequate or will

Spark be a superior choice for data investigation and storage? These queries trouble organizations and sometimes they can't track down the responses. They wind up settling on unfortunate choices and choosing improper tool. Subsequently cash, time, efforts and work hours are wasted.

## 5. CONCLUSION

The importance of big data does not resolve around how much data you have, but what you do with it. We can take data from any source and analyze it to find answer that enables cost and time reduction, new product development, optimized offerings and also helps in taking smart decisions. Big Data includes Black Box Data, Social Media Data, Stock Exchange Data, Power Grid Data, structured, semi structured, unstructured, quasi-structured data, search engine data etc [9].

Two IT initiatives are top of mind for organizations across the world i.e. Big Data Analytics and Cloud Computing. As a delivery mode for IT services, it has the potential to enhance business agility and productivity while enabling greater efficiencies and reducing cost. In the current era, big data is a big challenge for the organizations. To store and process such large volume of data, velocity and variety of data Hadoop or Map Reduce came into existence.

## 6. REFERENCES:

- [1] Hashem, I.A.T. et al., 2014. The rise of “big data” on cloud computing: Review and open research issues. *Information Systems*, 47, pp.98–115.
- [2] M. K.Kakhani, S. Kakhani and S. R.Biradar, Research issues in big data analytics, *International Journal of Application or Innovation in Engineering & Management*, 2(8) (2015), pp.228-232.
- [3] A. Gandomi and M. Haider, Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*, 35(2) (2015), pp.137-144.
- [4] C. Lynch, Big data: How do your data grow?, *Nature*, 455 (2008), pp.28-29.
- [5] X. Jin, B. W.Wah, X. Cheng and Y. Wang, Significance and challenges of big data research, *Big Data Research*, 2(2) (2015), pp.59-64.
- [6] R. Kitchin, Big Data, new epistemologies and paradigm shifts, *Big Data Society*, 1(1) (2014), pp.1-12.
- [7] C. L. Philip, Q. Chen and C. Y. Zhang, Data-intensive applications,

challenges, techniques and technologies: A survey on big data, *Information Sciences*, 275 (2014), pp.314-347.

- [8] K. Kambatla, G. Kollias, V. Kumar and A. Gram, Trends in big data analytics, *Journal of Parallel and Distributed Computing*, 74(7) (2014), pp.2561-2573.
- [9] S. Del. Rio, V. Lopez, J. M. Bentez and F. Herrera, On the use of mapreduce for imbalanced big data using random forest, *Information Science*