IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

Big Data in Cloud Computing: Benefits and Challenges

Harmanpreet Kaur

Post Graduate Government College for Girls, Sector -11, Chandigarh

Available online at: www.ijcseonline.org

Accepted: 03/Jun/2018, Published: 30/Ju/2018

Abstract:- In these days the most important emerging technologies to enter IT mainstream are big data and cloud computing. These two technologies are coming together to provide powerful results and making intelligent decisions in businesses. Cloud computing is a blessing for an organisation that wants to update technology under limited budget. It provides reliable, scalable and fault-tolerant environment to big data distributed management systems. Big data has ability to store and process different types of data at a high speed which is stored at different locations. In this paper the benefits and challenges involved in deploying big data through cloud computing has been discussed. Although there are couple of challenges and risk that come cross the path of integration between cloud computing and big data but with the invasion of new tools and technologies we are able to cope up with these problems. This paper presents an overview of both the technologies big data and cloud computing and also introduces the characteristics, benefits and challenges of big data in cloud computing.

***Keywords*:** Big data, Cloud Computing, types of data, services of cloud, benefits, challenges.

I. Introduction

Big data has become the need of the today's world. It helps the companies to improve operations and make intelligent decisions. Big data is a broad term which contains all type of data that exist in present day world. Therefore, it is collection of data from traditional and digital sources inside and outside of an organisation like hospital records, paperwork of offices, businesses and all type of institutions.

Examples of Big data are petabytes (1024 terabytes), exabytes (1024 petabytes) of data which contains trillions of records of millions of people. The sources of data are email, mobile devices, database services and applications. The data when captured, formatted, manipulated, stored and then analysed can help an organisation to improve its operations in various fields.

Due to vast Information technology there is an emergence of cloud computing which made it easier to provide the best technology in most cost effective manner and availability of wide range of applications to the small companies. Example 2.5 Quintillion bytes (2.3 trillion gigabyte of data is created everyday).

The high volume of data presents a challenge to a cloud environment. So, cloud computing and big data are ideal combination as they provide a scalable and accommodating solution for big data.

The main contribution of this paper is to provide an overview over the challenges that come across while deploying both technologies.

This paper is organized as follows Section 2 provides types of data and characteristics of big data ; Section 3 discusses benefits of big data in cloud computing; Section 4 discusses risk and challenges that is faced while deploying technologies and Section 5 presents the conclusion.

II. Big Data: Types of data and characteristics

Big data is used as a concept that refers to ability to make use of increasing volumes of data. It is classified into three main types.

Unstructured data: This data is not well organised or easily interpreted by traditional data models.

Example twitter tweets, emails, videos, audio files, satellite images, scientific data, web pages and social media messages.

Semi-Structured data: This data is neither raw nor typed data in conventional database. It is a form of structured data but is not organised in tabular form. **Example** XML, NoSQL databases, Email, other Markup languages.

Multi-Structured data: This data is derived from interaction between people and machines. It is available in many formats and resides in non-transactional systems such as sensors and customer interaction streams such as social networks or web applications that are combination of text and visual images along

© 2018, IJCSE All Rights Reserved

with structured data like transactional information.

The five different characteristics used to describe Big Data also known as Five V's are as follows:

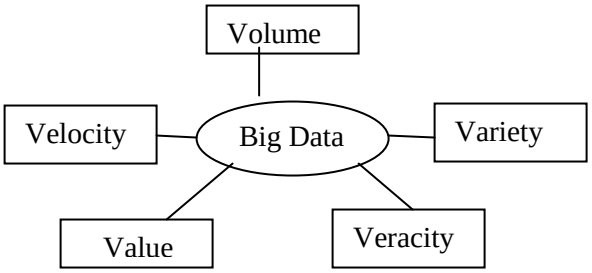

Figure 1. Five V'S of Big Data

Volume: It describes the vast amount of data generated ever from relative sources which include social media, businesses transactions and machine-to- machine data. So, with advancement of new technologies (Hadoop) the issue of storage has been reduced.

Velocity: It concerns the rate at which the data is generated and processed to meet the challenges and demands of today's world. Big data is available in real time.

Variety: It concerns various types of data from different sources the big data framework have to deal with. It draws text, images, video, audio and complete missing data through data function.

Value: It describes the importance of data regarding information they contain. Large amount of data is worthless unless they provide some value.

Veracity: It concerns the trust worthiness of the data. The quality of captured data can vary that affect in having accurate analysis.

The three basic services it offers are:

IaaS (Infrastructure as a Service): It delivers virtual infrastructure such as machine, storage and other hardware assets as a resources that satisfies the needs of the clients.

PaaS (Platform as a Service): It provides operating system, transactions and control structures. It allows users to deploy cloud applications created using run time environment. At this level big data DBMS are implemented.

SaaS (Software as a Service): It is one of the most known cloud model that provide complete operating environment with applications, management and user interface. Example: Oracle on demand, Google Apps.

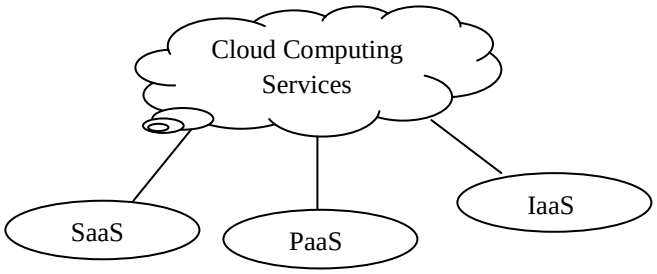

Figure 2. Benefits of Big data in cloud

III. Benefits of Big Data in Cloud Computing

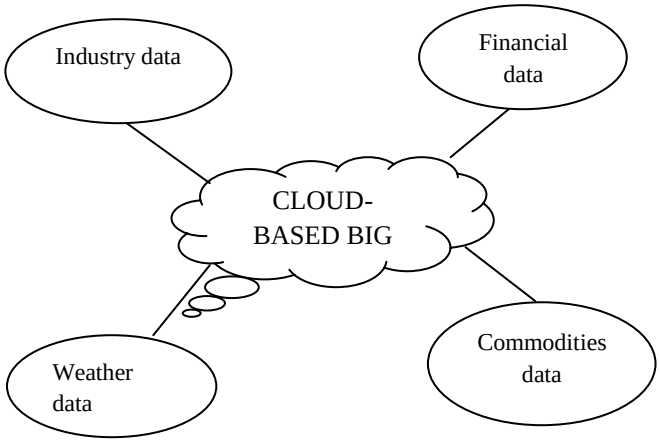

Figure 3. Types of cloud

Affordability: Cloud computing is a blessing for an organisation that wants to have updated technology under their limited budget. Organisation can pick what they want and pay for that. So, the big data are easily available in low cost. Now, the organisation can host their big data on offside servers and can update their data at low cost.

© 2018, IJCSE All Rights Reserved

Instant Infrastructure: The traditional infrastructure of managing a data can takes weeks to just install and run a server. With emergence of cloud computing it provides the company all resources which are required. Database of cloud computing enable an organisation to have thousands of virtual servers and get them working in minutes.

Data Processing: Big data analytics platform like Apache Hadoops, both structured and unstructured data can be processed.

Example social media generates a lot of unstructured data which can't be processed under one category like posts, photos, videos, blogs etc. With the help of cloud computing the whole process is made easier to all type of enterprises.

Feasibility: The virtual nature of the cloud allows unlimited resources on demand as compared to traditional solutions which requires more physical servers to increase processing power an storage space.

With the emergence of cloud computing companies can demand desired level of processing power and storage space easily. So, it is a perfect platform for the entire IT requirements. There is no need for additional infrastructure, as cloud provides most solutions in SaaS model.

IV. Risk and Challenges

There are some challenges and risk that should be considered while deploying big data in cloud environment. The most important issue is the security and privacy of big data cloud environment which are of major concern for businesses and cloud providers' today.

The reality is many attackers keep inventing new ways to find an entry points in the system and many of these attacks are due to server misconfiguration.

Other security issue that arises is because of integration between cloud computing and big data as they are creating new platform heterogeneity. So, new security tools are needed to work with new big data cloud platform such as Fraud detection patterns, encryptions, smart solutions, event logging and monitoring.

Another challenge is the location and type of data which is stored at different places in big data for processing. Such data is either moved to the processing environment or the processing is performed on the location.

The only way to deal with these challenges is to implement next generation technology which can predict an issue before it can cause damage.

V. Conclusion

Now days the most emerging technologies are Big data and Cloud computing. Cloud environment has potential for cost savings to the enterprises and is a perfect vehicle for hosting big data workloads. However, the cloud environments provide scalable, fault-tolerant solutions to the big data systems. The capability of big data to store different types of data at various locations and processing it at high speed result in data that can guide businesses and educational institutors in developing fast. The five V's volume, velocity, variety, value, veracity and the benefits to describe big data and three basic services of cloud computing have been discussed in this paper.

However, there are couple of challenges that come across the path of integration of both the technologies but with the invasion of new tools, technologies and investment of thousands of dollars in creating an environment which is suitable to work with terabytes of data.

References

[1] Rajsingh**, E.B.,** Veerasamy**, J.,** Alavi**, A.H.,** Peter**, J.D. "**Advances in Big Data and Cloud Computing"

[2] N.Turner, "Cloud Computing: A Brief Summary", Lucid Communications limited, 2009.

[3] Dillon,T., Wu,C., Chang, E.: "Cloud Computing: Issues and Challenges", 24th IEEE international Conference on advance Information Networking and Applications

[4] Marcello Trovati, Richard Hill, Ashiq Anjum, Shao Ying Zhu Lu Liu Editors, "Big Data Analytics and Cloud Computing-Theory, Algorithms and Applications"

[5] Arun K.Somani, Ganesh Chandra Deka, "Big Data Analytics-Tools and Technology for effective Planning"

[6] Chaowei Yang, Qunying Huang, Zhenlong Li, Kai Liu & Fei Hu, "Big Data and Cloud Computing: innovation opportunities and challenges".

[7]J.Preethi,,N.Aswathy,"A Survey on Cloud Applications", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 2, Issue 2, pp.10-17

[8] Xiaofang Li, Yanbin Zhuang & Simon X. Yang, "Cloud Computing for Big Data Processing" .

[9] K. Phaneendra, Dr. M. Babu Reddy, "Prototype Survey of Different Resource Provisioning Procedures in Cloud Computing", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 1, Issue 1, pp.108-115.

Author's Profile

Mrs Harmanpreet kaur received the M.C.A degree in 2013 from Panjab University, Chandigarh, India and cleared UGC-NET exam in January 2018. She works as a Resource Person in Post Graduate Government College for Girls Sector 11, Chandigarh, India in the Department of BSc Computer Science.

© 2018, IJCSE All Rights Reserved