# 5 V'S OF BIG DATA VIA CLOUD COMPUTING: USES AND IMPORTANCE

**Mohammad Tabrez Quasim[1], Prashant Johri[2], Mohammad Meraj[3], SK.Wasim Haider[3]**
[1] Department of Computer Science, College of Computing & Information Technology, University of Bisha, Saudi Arabia, mtabrez@ub.edu.sa
[2] School of Computing Sciences and Engineering, GalgotiasUniversities, Gr.Noida, India
[3]Nims University, Rajasthan, India

***ABSTRACT** : Presently in our daily life use, either on a personal or global basis,the volume of data coming from various sources which the traditional system fails and is no longer satisfy the challenges of handling this huge data (big data). In this perspective,, various big data cloud providers and their working mechanism with this huge data has been discussed. It also presents the utilization and significance of big data usingcloud computing in the present situation.Most of the organization right now are using big data for their strategic decision. This study offers a broad overview of the advantages, uses and importance of big data linked with cloud computing in various fields like Banking System, Healthcare, Transportation, Education, Manufacturing Natural Resources, Insurances, Media, Entertainment and different Government sectors and furthermore talked about how to apply inventive solution utilizing this terminology*

**Keywords**: Big Data, 5 V, Cloud Computing, Application of Big data

## 1. INTRODUCTION

Big Data is a new term, used to describe a huge amount of structured, semi-structured and unstructured data that is so extensive. Sometimes it is exceptionally hard to process like conventional database and software techniques directly. The measure of the size of data looks like in the form of $10^3$(KB),$10^6$(MB),$10^9$(GB),$10^{12}$(TB),$10^{15}$(PB),$10^{18}$(EB),$10^{21}$(ZB),$10^{24}$(YB) which are obviously large data. So to handle this type of data we use the concept of Big Data which could be started from any point. For instance, a 100MB file is needed to send using email, but many times we feel it is unable to send easily that means this is the big data in respective to email because it cannot process using the traditional system. There are sorts of technological devices available in the market from different vendors including IBM, Microsoft, Amazon, etc., to manage big data. The types of data in a big data plateform are of three types,Structured data: Relationship data like files and Tables etc[2,3]

Semi Structured data: XML data. Unstructured data textual: Word, PDF, Text, Media Logs and non-textual audio, video document.

## 2. 5 V's OF BIG DATA

### 2.1Volume

Volume represents the amount of data. The volume of data has increased exponentially in recent times because of vast development in the technological world. The number of emails, social media posts, video clips, or even new text added per day is in excess of several billion entries. Facebook alone can generate over five new profiles a second, 136,000 new pictures uploads a minute, and 510 comments per minute. Other than FB we have a lot of other sources like Twitter, Instagram, LinkedIn, youtube, E-Commerce Website, etc[4, 5]. The traditional storage cannot handle structured and unstructured data. This is a big challenge for enterprises nowadays.

### 2.2Velocity

Velocity represents the speed at which new data is generated and the moves around. Just think of different social media messages going viral in seconds, the speed at which credit card transactions are checked for fake activities, In the current scenario every day even every moment the number of emails, photos, video clips, twitter messages, etc. increases at tremendous speeds around the whole world. now Big data technology allows us to analyze the data during being generated, without putting it into databases.

### 2.3 Variety

It refers to the diversity of data types and data sources. 80 percent of the data in the world today is unstructured and at first, look does not demonstrate any sign of connections. An organization can get Data/information from a wide range of sources: from in-house gadgets to Smartphone GPS innovation or what individuals are stating on social networks. though, this category itself is extremely broad and could incorporate blog entries, social media updates, social media profiles, pictures, videos, audio files, and so forth. With inventive enormous big data technologies, organizations can now create structure within big data to better collect, store, and analyze it to drive value . Because of Big Data such algorithms, data can be arranged in an organized way and examined for relationships. Data does not generally contain just regular datasets, yet in addition pictures, video, animation and speech recordings.

### 2.4 Value

When we talk about value, we're referring to the worth of the data being extracted. This is the main challenge in big data because the value is essential for businesses. It is all about the quality of the data that is stored and further use of it [2].

### 2.5 Veracity

Veracity is all about making sure the data is accurate. For example, If X is sending the email to Y,Y will have the exact content that X sent it, otherwise, the email services will not be reliable and people will not use it[2,3]. A pictorial diagram of 5V's of a big data is shown below.

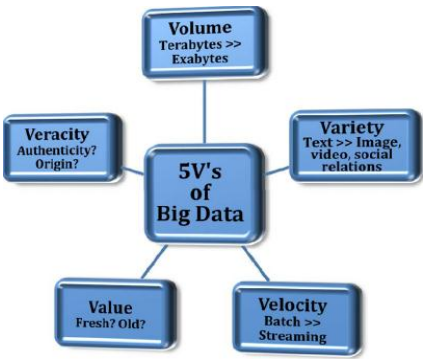


**Fig 1 Characteristics of Big Data**

## 3. CLOUD COMPUTING

Cloud computing is a new business model and service model. Its central idea is that it doesn't depend on the local computer to do computing, but on computing resources operated by third parties that provide computing, storage, and networking resources. Today, the industry says that Cloud computing is the fifth public resource ("the fifth utility") after water, electricity, gas, and oil. The industry generally trusts that Cloud computing can be separated into the following bottom-up categories, depending on the type of service[1, 4, 6]:

**3.1 Infrastructure as a Service (IaaS) in the Cloud**: Provides infrastructure, including physical and virtual servers, storage, and network bandwidth services directly to users[1, 4]. Users design and implement applications based on their practical requirements, like Amazon EC2 (Amazon Elastic Cloud Computing).

**Table-1: Service type classification of cloud computing**

| Classification | Service type | Flexibility/ Generality | Difficulty level | Scale and example |
|---|---|---|---|---|
| IaaS | Basic computing, storage, network resources | High | Difficult | Large, Amazon EC2 |
| PaaS | Application hosting environment | Middle | Middle | Middle, Google App Engine |
| SaaS | Application with specific function | Low | Easy | Small, Salesforce CRM |

**3.2 Platform as a Service (PaaS) in the Cloud**: Provides a hosting Cloud platform in which users can put their applications onto the Cloud platform. Development and deployment of the applications must comply with the specific rules and restrictions of the platform, such as the use of certain programming languages, programming frameworks, and data storage models. For example, Google App Engine provides an operating environment for Web applications; once the applications are deployed, other involved management activities— like dynamic resource management—will be the responsibility of the platform.

**3.3 Application as a Service in the Cloud**: Provides software that can be used directly, most of which is browser-based and specific for a particular function. For example, Salesforce provides the customer relationship management system (CRM). The application is easy to use in the Cloud, but its flexibility is low and it is generally only used for a specific application.

## 4. RELATIONSHIP BETWEEN CLOUD COMPUTING AND BIG DATA:-

Cloud Computing and Big Data are linked together. The use of cloud computing in big data is shown in Fig 2[4].

Big Data is using Distributed storage technology which depends on cloud computing rather than local storage to a computer or electronic devices. Big data summarizes by quickly developing cloud-based innovations which is created using by virtualized technologies. In this manner, cloud computing not just gives facilities for the automation and processing of big data yet, in addition, serves as a service model. The following **Table-2** is showing the comparison of various big data cloudproviders[8, 9, 10].

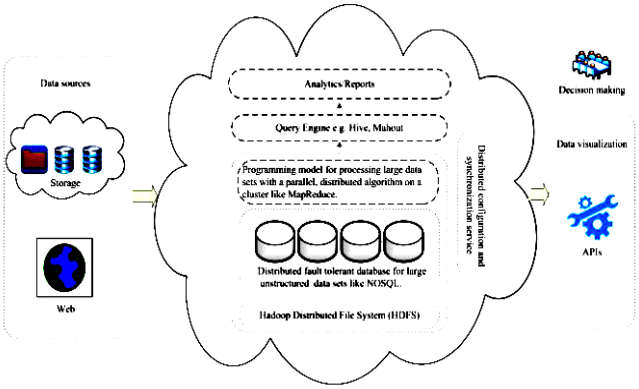


**Fig 2**

**Table-2**

| | Google | Microsoft | Amazon | Cloudera |
|---|---|---|---|---|
| Big Data Storage | Google Cloud Services | Azure | S3 | |
| Map Reduce | AppEngine | Hadoop on Azure | Elastic Map Reduce (Hadoop) | Map Reduce YARN |
| Big Data Analytics | Big Query | Hadoop on Azure | Elastic Map Reduce (Hadoop) | Elastic Map Reduce (Hadoop) |
| Relational Data Base | Cloud SQL | SQL Azure | My SQL or Oracle | My SQL,Oracle, PostgreSQL |
| NoSQL Database | AppengineDataStore | TableStorage | DynamoDB | Apache Accumulo |
| Streaming Processing | Search API | Streaminsight | Nothing Prepackaged | Apache Spark |
| Machine Learning | Prediction API | Hadoop+ Mahout | Hadoop+ Mahout | Hadoop+ Oryx |
| Data import | Network | Network | Network | Network |

## 5. Application of Big Data

Big Data application analyzes and gives the solution of real-world problems. Industries are facing challenges and how to solve these challenges using big data. In this article also mention the big data providers that are offering solutions in the specific industry [8,11, 12].

### 5.1 Banking and Securities

The difficulties in the banking industry incorporate card fraud detection, client information exchange, IT tasks examination, tick investigation.

The Big Data is being utilized by the Securities Exchange Commission (SEC) to look at financial market exercises. They are using natural language processors and network analytics to get unapproved exchanging activities in the financial markets. Retail merchants, Big banks, mutual funds, and other supposition estimation and so on.

### 5.2 Communications, Media and Entertainment

In the present situation, clients can get too rich media on-request even in various format and in an assortment of devices. Some big data challenges in the communications, media and entertainment industry are :

a. Collecting, analyzing, and using client's inevitable
b. Controlling distinctive internet-based media and mobiles content
c. Understanding and examining the patterns of real-time media content utilization
d. Applications of big data in the Communications, media and entertainment industry
e.Construct content for individual target groups of onlookers
f.Recommend content on interest

**5.3Healthcare**

At present healthcare sector approaches expansive volume

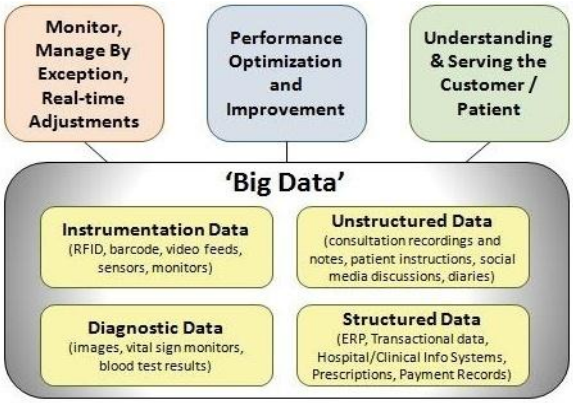


**Fig-3**

of data yet shockingly fails in impeccably utilizing these data to deal with the expense of rising medicinal services and by wasteful frameworks that anticipate better and quicker healthcare benefits over the world. It is only because of electronic data is inaccessible, unusable and inappropriate. The healthcare databases which keep health-related data have made it hard to connect data that can indicate patterns valuable in the medical field.

Some advanced hospitals gathered data from various cell phone app, from a large number of patients, to enable specialists to utilize evidence-based medicine instead of regulating a few medical/lab tests to all patients who go to the hospital. The University of Florida and some others universities used free public health data and Google Maps to make visual data that takes into consideration for faster identification and productive analysis of healthcare services data, used in tracking the spread of chronic infection (Fig-3) [7].

**5.4 Education**

At present Big data is fundamentally utilized in advanced education. A distinctive university like the University of Tasmania, an Australian university with more than 30000 students has conveyed a Learning and Management System that tracks the general advancement of a student over time, it is additionally used to quantify instructor's adequacy to guarantee a decent ordeal for both students and educators. Educator's performance can also be judged against quantities of students, students socioeconomics, subject matter, student desires, behavioral classification, and many other factors

**5.5 Government**

Presently some government organizations are utilizing big data to make our life better and simpler A few examples are:-

a) The Food and Drug Administration (FDA) is utilizing big data to watch and study patterns of food-related illnesses and infections which has prompted quicker treatment and less death.

b) Governments Transportation System are utilizing big data for traffic control, route arranging, congestion management (by foreseeing traffic conditions), intelligent transport frameworks, Private sectors are likewise utilizing big data in transport: revenue management, innovative improvements, logistics and for competitive advantage

Individual utilization of big data incorporates route planning to save money on fuel and time, for travel arrangement in tourism, etc.

c) The insurance agencies are additionally utilizing big data for better client protection with regards to claims management, a prescient investigation from big data has been utilized to offer quicker administration since huge measures of data can be analyzed particularly in the guaranteeing stage. Fraud detection can also enhanced.

## 6. CONCLUSION

The present work describes the concept of Big Data and examines its distinctive qualities. In this article, we have also examined about 5 V's of big data i.e. volume, velocity, variety, veracity and the connection between the cloud computing and big data. We have focused on different big data applications that are utilized in IT organization or other association to store huge amount of data using technologies like Hadoop, HDFS(Hadoop distributed file system), map-reduce framework, NoSql, Pig, Hive ,Sqoop, Zookeeper.

## 7. REFERENCES

[1] Yang K, Jia X. "Data Storage Auditing Service in Cloud Computing: Challenges, Methods and Opportunities" World Wide Web. 2012; 15(4):409−28. Crossref

[2] Alexandru Adrian TOLE "Big Data Challenges "Database Systems Journal vol. IV, no. 3/2013

[3] Srinivasan S. "Basic Cloud Computing Types. Cloud Computing Basics:" Springer; 2014. p. 17−41. Crossref.

[4] C. Lakshmi*, V. V. Nagendra Kumar "Survey paper on Big Data " International Journal of Advance Research in Computer Science and Software Engineering Volume 6, Issue 8, August 2016.

[5] Yuri Demchenko "The Big Data Architecture Framework (BDAF)‖ Outcome of the Brainstorming " Session at the University of Amsterdam 17 July 2013.

[6] R Kalaivani "Security Perspectives on Deployment of Big Data using Cloud: A Survey"International Journal of Advanced Networking & Applications (IJANA) Volume: 08, Issue: 05 Pages: 5-9 (2017) Special Issue.

[7] S.J.Samuel, K.RVP, K.Sashidhar, C.R.Bharathi, "A survey on big data and its research challenges", ARPN Journal of Engineering and Applied Sciences, Vol.10, No.8, Pp.3343, 3347, 2015.

[8] S.Kuchipudi, T.S.Reddy, "Applications of Big data in Various Fields", International Journal of Computer Science and Information Technologies (IJCSIT), Vol.6, N o.5, Pp.4629 - 4632, 2015.

[9] S.Mukherjee, R.Shaw, "Big Data Concepts, Applications, Challenges and Future Scope" International Journal of Advanced Research in Computer and Communication Engineering, Vol.5, No.2, 2016.

[10] A.Misra, A.Sharma, P.Gulia, A.Bana, "Big Data: Challenges and Opportunities", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol.4, No.2, Pp.41 - 42 2014.

[11] L.Venkata, S.Narayana, "A Survey on Challenges and Advantages in Big Data, International Journal of Computer Science and Technology Vol.6, No.2, 2015.

[12] Jifa G. Some Issues on Big Data, Science and Technology for Development. 2014; 1:004