

# Cloud Based Big Data Analysis: A Review

Arpitha J C<sup>1</sup>, Akhila S J<sup>2</sup>, Shwetha S Kulloli<sup>3</sup>

<sup>1-3</sup>Assistant Professor, BCA, KLE Society's Degree College, Nagarbhavi, Bangalore, Karnataka, India

\*\*\*

**Abstract** - Big data is a data analysis approach that provides in the enormous growth of various field used in daily life like social network analysis, semantic web analysis and bioinformatics network analysis. Today, organizations, business units, government sectors, etc. are adopting the big data technique to store the large data generated by them. Cloud computing is a model that is growing everywhere to deliver big data services. Overview of big data along with its technologies is considered. This paper shows the survey over the cloud based big data, cloud computing architecture, service models of cloud computing. The basic problems and challenges in delivering big data are mentioned in the rest of the paper.

**Key Words:** Cloud computing, cloud service model, big data, big data technologies.

## 1. INTRODUCTION

Cloud is a server assortment that is spread throughout the internet to store, process and handle the data. Cloud computing provides the hardware and software services over internet. It allows big data to manage and give the stored data in proper way [1]. Cloud computing provide the security to big data with hadoop, environment like file and network, encryption, logging, nodes authentication, layered framework for assuring cloud etc. Different types of models are using for mining the big data. Cloud computing approach is important as it has the adopted technologies to manage the large amount of data. It provides the interconnectivity between the devices and data that is further assisted to exchange the data and connected to other devices. Big data term is essentially used to convert the data into information. The major purpose of big data is to store and manage, visualize, and analyze huge amount of data per day.

## 2. BIG DATA

The word big data is nothing but large data, further complex data and relationship analysis of these data sets. The main advantage of big data is that it performs the better analysis of huge data than conventional analysis methods. Due to this reason the big data has increased very much importance in the present generation, which has development in the data collection, data storage and performs the data interpretation. From last few eras, the use of digital media is being increased in many areas which making the enormous amount of data, for example, hospital data, bank data, social networking data, etc. The data

storage cost is reducing day by day by which we can store the entire data rather than removing it. In addition to this, many of the data analyzing techniques are developed and have not accomplished in efficient data analysis.

The big data in the real world is like the collection of huge resources which can be used consistently. The big data provides the vast application advantages, but the conventional data analysis methods fail to provide the proper privacy mechanism. The privacy concern of the big data includes the private data disclosure to the world.

### 2.1 Big Data Types

The Big data can be categorized into various types that help in better understanding of the BD characteristics. This categorization of BD provides the better understanding of large data stored in the cloud system. The types are: data sources type big data (BD), data processing type, format content type, data staging type, and data stores type and are explained as below.

**Type 1- Data Sources:** The data sources type of big data contains social media data, machine data, sensing data, transactions data and IoT data.

**a. Social media data:** This is data created by many social blogs, microblogs, Facebook, Twitter through the virtual communication.

**b. Machine data:** This is data generated by some hardware's or software's based on hospital machines, computers without humans help.

**c. Sensing data:** This is data gathered by sensors and converts them as signals.

**d. Transactional data:** The data's like bank transactions, work records, and other time dimension-based events data are considered as transactional data.

**e. Internet of things (IoT) data:** The data generated by the interface of Smartphone's, tablets and digital cameras with the internet are considered as IOT data.

**Type 2- Data processing data:** This type of data includes real-time and batch data.

**a. Batch data:** Many organizations in recent days are adopted the Map Reduce systems for running the batch

jobs, such systems offers the large data scaling generated from many nodes.

**b. Real-time data:** The data generated with the real-time computation, processing, etc.

**Type 3- content format data:** The content format data types are divided into structured, semi structured and unstructured data.

**a. Structured type data:** These kinds of data are handled with SQL. The data is easy to apply as input, for storage, query and analysis purpose. Numbers, dates, and words are the structured data.

**b. Semi-structure type data:** These are may be of structured and are not organized as rational databases or tables.

**c. Unstructured type data:** This data includes location information, texts, videos and social data.

**Type 4- Data staging:** The data staging includes cleaning data; transform data, and normalization data.

**a. Cleaning data:** This is unreasonable, incomplete data identification or cleaning.

**b. Transform data:** The process data were transforming as the suitable form of data for analysis.

**c. Normalization data:** This is database structuring method to reduce the redundancy.

**Type 5- Data stores:** This is divided into document oriented type of data, Column- oriented data, graph database, key-value type, and all are used to store the data.

### 3. CLOUD COMPUTING

Cloud computing provides enormous computing resources to the user applications through internet in the large-scale distributed computing environment. Cloud computing is characterized by multiagency, fast deployment, low cost, scalability, rapid provisioning, elasticity and ubiquitous network access [1]. Therefore, users can access the resources as per necessity from the cloud on pay-as-use basis. They are not required to buy and install the required resources locally [2]. For example, if we need a resource occasionally then it is the better option to use the resource from cloud instead of providing to own the resource. In the modern world in the aura of cloud, we can store the large data on cloud and may access it whenever and wherever required. Cloud computing is inherently flexible architecture as we can access the resources on our demand and need basis [1,3]. Resources are always

available on the cloud and are made available to different user fields on demand basis. Furthermore, through virtualization the high capacity servers in cloud are provisioned to different users in parallel. Virtual machines which are fabricated using virtualization software are allocated to the users instead of the physical machines [4]. Cloud computing reduces the resource expense by pay-as-per usages and offers high storage facility.

### 3.1 CLOUD SERVICE MODELS

#### A. Cloud Service Models

**a. Software as a Service (SaaS) -** In this service model user can retrieve the software applications and database accessible on the cloud. It provides the environment to run a application. If a user does not have the required software or hardware sources locally, he can retrieve it from the cloud [5].

**b. Platform as a Service (PaaS) -** In this service model user can configure and install the software on the cloud.

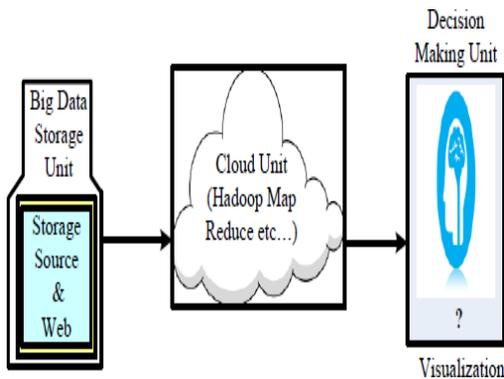
PaaS gives operating system, programming language and web server to design the software. User can save the time and price to buy all these resources [5].

**c. Infrastructure as a Service (IaaS) -** In this service model cloud provides infrastructure like virtual machines, storage, network, IP addresses and other specialized software or hardware resources through high speed networks to the users [5]. In IaaS, an environment is provided to deploy and run the infrastructure (hardware/software) in the distributed cloud environment.

### 3.2 Cloud-based Big Data Analysis Architecture

The combined architecture of cloud system and big data system are shown in figure 3. Huge data sources from the cloud and Web are put away in a circulated shortcoming tolerant database and controlled through a programming model for extensive data sets with a parallel dispersed calculation in a group. The main purpose behind information representation is to see explanatory results displayed outwardly through various diagrams for basic leadership. Enormous information uses disseminated capacity Innovation

Considering distributed computing as opposed to nearby stockpiling appended to a PC or electronic gadget. Big data assessment is driven by quickly developing cloud-based applications created utilizing virtualized advancements. In this manner, cloud computing not just gives computation and processing of enormous data additionally serve as a service mode.



**Fig - 1: Architecture for Cloud-based Big Data Analysis Architecture**

#### 4. CHALLENGES IN BIG DATA

**a) Availability** - Cloud computing permitted to access the data by personally authorized. Due to the distribution of data on various clouds, it reduces the performance of accessing data[1]. Another issue is to tackle the converting of data into suitable forms.

**b) Security** - The security becomes at extreme risk when users accessed personal and other sensitive information through credit/debit cards[2]. Moreover, different organizations have their own rules and regulations about the secure data. Hence, multi-level security is needed for all types of data [2]. It needs a privacy preserved data model to provide security to the sensitive data of any organization or personal information. Sometimes hackers store or lose the data due to the poor security.

**c) Scalability** - There is a mismatching of data speed and CPU speed. Due to the large volume of data, it is not delivered to the processor at proper time. Many applications require parallel computing like navigation, social networks, finance, internet search, timeliness etc[2]. There is a requirement of the cloud service to provide the services of infrastructure, platform and applications at required time to maintain the scalability [1].

**d) Big Storage** - Data can be in different forms it may be text, images, audio/video etc. Such type of data is use by different mediums like mobile devices, aerial sensory technologies, remote sensing, radio frequency identification readers etc. Such type of data require large storage device with huge space and higher input/output speed [1]. It is most difficult to access the information from unstructured data [1]. Large amount of data could not be retrieved at proper time. Hence, it becomes more tuff in file systems.

#### 5. CONCLUSION

The modern technology like cloud computing will suggest a scalable service for the big data with optimized cost. But the concern of privacy and security is still unresolved. In the recent time cloud computing has emerged as a model in computing science. Big data is an emerging platform to manage and distribute the large- scale data. It changes the conventional data base techniques into valuable innovative and machine learning techniques. Although, big data is serving with many facilities still it has many issues and challenges.

#### REFERENCES

- [1] M. Mayilvaganan and M. Sabitha, "A cloud-based architecture for Big-Data analytics in smart grid: A proposal," Computational Intelligence and Computing Research (ICCIC), IEEE International Conference, Enathi, pp. 1-4, 2013
- [2] G. Baciú, C. Li, Y. Wang and X. Zhang, "Clouds: Cloud-based cognition for large streaming data," Cognitive Informatics & Cognitive Computing (ICCI\*CC), 2015 IEEE 14th International Conference, Beijing, 2015, pp. 333-338.
- [3] Hemanth Kumar N. P. and Prabhudeva S. Security and Privacy Preservation on Cloud-based Big Data Analysis (CBDA): A Review "Communications on Applied Electronics (CAE) – ISSN : 2394-4714" Volume 6 – No.1, October 2016
- [4] Dr. Harjinder Kaur Dr. Major Singh Goraya "Role of Big Data in Cloud Computing: A Review", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181 Vol. 8 Issue 07, July-2019
- [5] J. Srinivas, K.Venkata Subba Reddy and Dr. A. Moiz Qyser, "Cloud Computing Basics", International Journal of Advanced Research in Computer and Communication Engineering, Vol.1(5), 2012.
- [6] Bernice M. Purcell, "Big data using cloud computing", Journal of Technology Research, <http://www.aabri.com>
- [7] D.P. Acharjya, Kauser Ahmed P, " A Survey on Big Data Analytics: Challenges, Open Research Issue and Tools", International Journal of Advanced Computer Science and Applications, Vol. 7(2), 2016

## BIOGRAPHIES



Arpitha J C completed BE from STJIT and M Tech from JNNCE Shvaimogga and currently working as Assistant Professor in KLE Society's Degree College, Nagarbhavi, Bangalore. Her research fields include Cloud Computing and Big data.



Akhila S J completed BE from BGSIT, B G Nagar and M Tech from DBIT Bangalore and currently working as Assistant Professor in KLE Society's Degree College, Nagarbhavi, Bangalore. Her research fields include Sensor

Cloud and IoT.



Shwetha S Kulloli completed BE from SGBIT and M Tech from VTU Belgaum and currently working as Assistant Professor in KLE Society's Degree College, Nagarbhavi, Bangalore. Her research fields include Image Processing and machine

Learning.