

## GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

### AN ANALYSIS ON USE OF BIG DATA IN CLOUD COMPUTING ENVIRONMENT

Dhawal Shah<sup>\*1</sup> & Makhan Kumbhkar<sup>2</sup>

<sup>\*1</sup>IET, DAVV, Indore, MP, India

<sup>2</sup>Christinian Eminent College indore,MP,India

---

#### ABSTRACT

Cloud computing is a powerful technology to perform large-scale and complex computing. Cloud computing removes expensive hardware configuration, lot of space needed and huge amount of software. Cloud computing observed that a huge amount of data or big data generated. Now a day's big data is a challenging and time-demanding task that requires a huge computational infrastructure to ensure successful data processing and analysis.

In this study big data in cloud computing is reviewed. The definition, characteristics and classification of big data along with some discussions on cloud computing are introduced. And also discussed relation between big data and cloud computing.

*Keywords-* Big Data, Cloud Computing.

---

#### I. INTRODUCTION

The fruitful paradigm for the service-oriented programming is the cloud computing. Cloud computing is a big shift from the traditional businesses to new modern era. Cloud computing eliminates the capital expense of buying hardware and software and setting up and running on-site datacentres—the racks of servers, the round-the-clock electricity for power and cooling, the IT experts for managing the infrastructure. It adds up fast. The biggest cloud computing services run on a worldwide network of secure datacentres, which are regularly upgraded to the latest generation of fast and efficient computing hardware. Big Data Analytics platforms like Apache Hadoop, structured and unstructured data can be processed. Cloud computing makes the whole process easier and accessible to small, medium and larger enterprises. The elasticity, pay per use, low upfront investment, transfer of risks is few of the major enabling characteristics that makes the cloud computing the universal platform for deploying economically reasonable enterprise infrastructure.

#### II. BIG DATA

Big data is a word used for explanation of huge amounts of data which are either structured, semi structured or unstructured. The data if it is not able to be handled by the old era databases and software technologies then we categorize such data as big data. The big data is defined using three v's.

##### Volume

We currently see the exponential growth in the data storage as the data is now more than text data. We can find data in the format of videos, music and large images on our social media channels. It is very common to have Terabytes and Petabytes of the storage system for enterprises. As the database grows the applications and architecture built to support the data needs to be reevaluated quite often. Sometimes the same data is re-evaluated with multiple angles and even though the original data is the same the new-found intelligence creates explosion of the data. The big volume indeed represents Big Data.

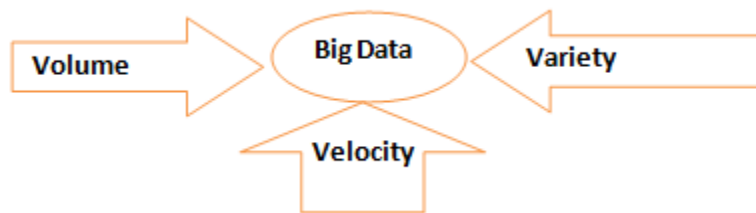
##### Velocity

The data growth and social media explosion have changed how we look at the data. There was a time when we used to believe that data of yesterday is recent. The matter of the fact newspapers is still following that logic. However, news channels and radios have changed how fast we receive the news. Today, people reply on social media to

update them with the latest happening. On social media sometimes a few seconds old messages (a tweet, status updates etc.) is not something interests users. They often discard old messages and pay attention to recent updates. The data movement is now almost real time and the update window has reduced to fractions of the seconds. This high velocity data represents Big Data.

### **Variety**

Data can be stored in multiple formats. For example, database, excel, csv, access or for the matter of the fact, it can be stored in a simple text file. Sometimes the data is not even in the traditional format as we assume, it may be in the form of video, SMS, pdf or something we might have not thought about it. It is the need of the organization to arrange it and make it meaningful. It will be easy to do so if we have data in the same format, however it is not the case most of the time. The real world has data in many different formats and that is the challenge we need to overcome with the Big Data. This variety of the data represents Big Data.



There are few examples for big data usage

### **Hadoop**

Hadoop is a framework that allows us to first store Big Data in a distributed environment, so that, we can process it simultaneously.

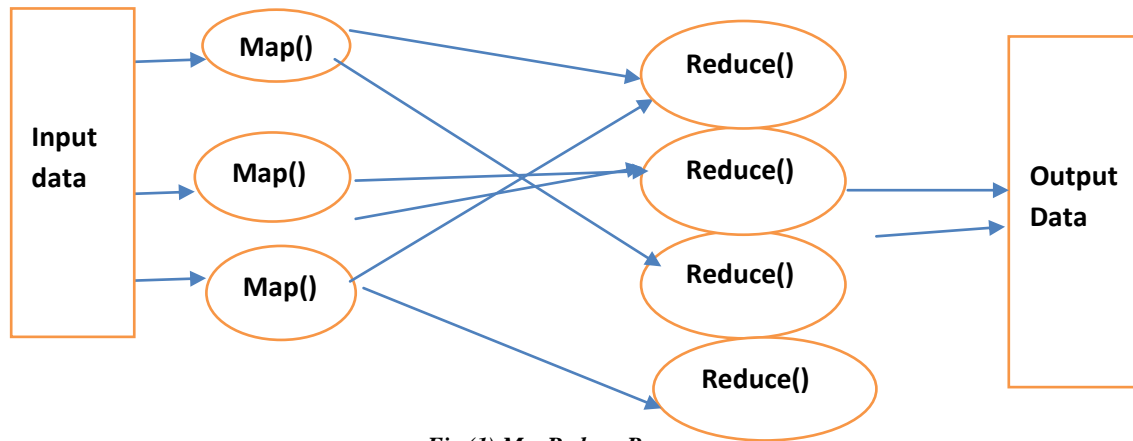
There are basically two components in Hadoop:

The first one is HDFS for storage (Hadoop distributed File System), that allows you to store data of various formats across a cluster. The second one is YARN, for resource management in Hadoop. It allows parallel processing over the data, i.e. stored across HDFS.

### **MapReduce**

MapReduce is a programming model for processing huge data sets with a parallel, distributed algorithm on a cluster (source: Wikipedia). Map Reduce when coupled with HDFS can be used to handle big data. The basic unit of information, used in MapReduce is a (Key, value) pair. All types of structured and unstructured data need to be translated to this basic unit, before feeding the data to MapReduce model. As the name suggests, MapReduce model consist of two separate routines, namely Map-function and Reduce-function. This paper will help us to understand the step by step functionality of Map-Reduce model. The computation on an input (i.e. on a set of pairs) in MapReduce model occurs in three stages:

- Step 1: The map stage
- Step 2: The shuffle stage
- Step 3: The reduce stage



*Fig (1) MapReduce Process*

### III. BIG DATA APPLICATIONS

In the present age of data explosion, parallel processing is very much important for performing a huge volume of data in timely manner. Parallelization techniques and algorithms are used to achieve better scalability and performance for processing big data. Basically, Big Data uses in the following sector

#### *Healthcare*

The level of data generated within healthcare systems is not trivial. Traditionally, the health care industry lagged in using Big Data, because of limited ability to standardize and consolidate data.

#### *Manufacturing*

Predictive manufacturing provides near-zero downtime and transparency. It requires an enormous amount of data and advanced prediction tools for a systematic process of data into useful information.

#### *Media & Entertainment*

Various companies in the media and entertainment industry are facing new business models, for the way they – create, market and distribute their content. This is happening because of current consumer’s search and the requirement of accessing content anywhere, any time, on any device using big data and Cloud Computing

#### *Government*

The use and adoption of Big Data within governmental processes allows efficiencies in terms of cost, productivity, and innovation. In government use cases, the same data sets are often applied across multiple applications & it requires multiple departments to work in collaboration. Big data and Cloud Computing easily can handle these types of task.

### IV. CLOUD BIG DATA CHALLENGES

The main serious challenges of Big Data are capturing, cleaning, curation, integration, storage, processing, indexing, search, sharing, transfer, mining, analysis, and visualization of large volumes of fast-moving highly complex data. Cloud services are configured and priced differently compared to the regular servers deployed in the data centres. To choose and optimize cloud infrastructure for managing the big data workloads, it’s a challenge to integrate the existing systems with cloud services.

This paper given a precise description of Big data and Technologies in Cloud environment. We discussed about the applications, benefits and challenges faced by big data when used cloud computing environment. In future, the challenges are need to be overcome and make way for the even more efficient use of the big data by the user on a cloud computing environment

#### **REFERENCES**

- [1] Venkatesh, "A Study on Use of Big Data in Cloud Computing Environment", *International Journal of Computer Science and Information Technologies*, Vol. 6 (3), 2015, 2076-2078.
- [2] <https://www.qubole.com/resources/big-data-cloud-database-and-computing>.
- [3] Makhan Kumbhkar, "Security Analysing in UNIX for Cloud Computing Environment", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 3, Issue 10, October 2015.
- [4] Vishnu Patidar, "Analysis of Cloud Computing Security Issues in Software as a Service", *International Journal of Scientific Research in Computer Science and Engineering*, Volume-2, Issue-3 ISSN: 2320-7639..