
ABSTRACT

Big data is a collection of huge quantities of data. Big data is the process of examining large amounts of data. Big data and Cloud computing are the hot issues in Information Technology. Big data is the one of the main problem now a day's. Researchers focusing how to handle huge amount of data with cloud computing and how to gain a perfect security for big data in cloud computing. To handle the Big Data problem Hadoop framework is used in which data is fragmented and executed parallel. Hadoop use two technologies HDFS and Map reduce.

KEYWORDS: Cloud computing, Big data, Hadoop.

INTRODUCTION

Today, technology is increasing very vastly. Organizations are producing and storing vast amount of data. The most popular applications are internet services with millions of users. Websites like Facebook, Google receive millions of clicks daily. These applications generate megabytes to terabytes of invaluable data [5]. Real time capturing, storage, and analysis of this data are common needs of all high-end online applications. To address these problems, a number of cloud computing technologies have emerged.

Cloud computing has been transmute the IT organizations by adding flexibility to the way. IT is devour, enabling organisations to pay only for the resources and services they use [3]. Cloud computing provides virtualized resources over the Internet. Cloud provided infrastructure, platform, and software as services. Software-as-a-service is a service where user can access software according to need without bought any permanent software. In which user can pay according to access service on the bases of time same as in infrastructure as a service and platform as a service. In platform as a service user run applications on different platform according to need on cloud and according to this they used it. Same like that in infrastructure as a service user access machines and components. It's a very cheaper and easier way to access machines through internet on cloud. [6] The service providers are used to provide these services. This helps in delivering the storage and computing services using the internet. It makes ubiquitous data access possible to store data in cloud computing.

Clouds can prevent organisations from spending money for maintaining peak-provisioned IT infrastructure that they are unlikely to use most of the time. Cloud computing is an advance technology to perform complex computing. That's way organizations prefer to access cloud services. Organizations produced mega bytes of data per-second. To handle this data is very complicated it is called big data. Addressing big data is a challenging and time-demanding task that requires a large computational infrastructure to ensure successful data processing and analysis. To handle this problem hadoop technology is introduced.

Google has introduced MapReduce framework for processing large amounts of data on commodity hardware. Apache's Hadoop distributed file system (HDFS) is evolving as a superior software component for cloud computing

[7] combined along with integrated parts such as MapReduce. Hadoop which is an open-source implementation of Google MapReduce, including a distributed file system, provides to the application programmer the abstraction of the map and the reduce. With the help of Hadoop it is easier for organizations to get a grip on the large volumes of data being generated each day, but at the same time can also create problem related to security, data access monitoring, high availability and business continuity [7]. In this paper we discussed about big data problem with cloud computing and hadoop frame work to solve big data problem.

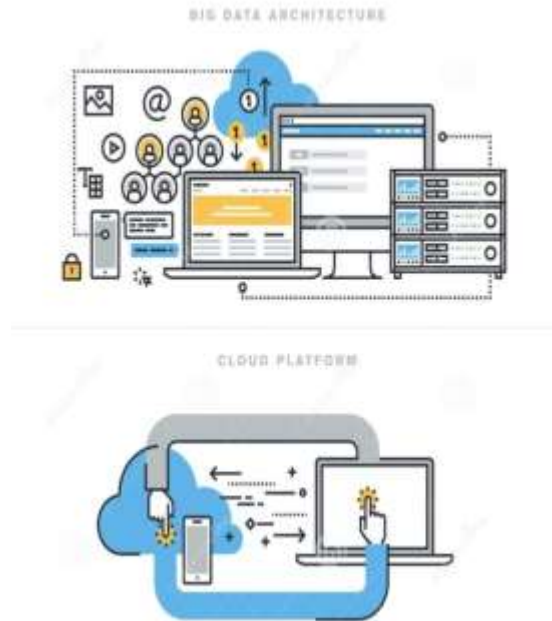


Fig 1: Cloud Architecture

Big data: Big data is a combination of data sets whose size, complexity and velocity make them difficult to be captured, managed, processed or analyzed by conventional technologies and tools such as relational database and desktop statics or visualization packages, within the time necessary to make them useful [2].

Challenges in big data management

Data integration: new protocols and interfaces for integration of data that are able to manage data of different nature and sources.

Data storage:[3] How to store large volumes of information in a way it can be timely retrieved? How to efficiently recognise and store important information extracted from unstructured data? How to store information in a way that it can be easily migrated/ ported between data centres/ cloud providers?

Data Processing and Resource Management: How to optimise resource usage and energy consumption when executing the analytics application?

Data Variety: how to handle an always increasing volume of data ? Especially when the data is unstructured, how to quickly extract meaningful content out of it ? How to aggregate and correlate streaming data from multiple sources?

Big data applications:

Parallelization techniques are used to achieve better scalability and performance for processing big data. Hadoop and Map reduce is the main applications used in big data. Map reduce is a very popularly used tool. It is very simple for use by the programmers to code for the map reduce task.

Hadoop:

Hadoop is a software framework which is used to processing the large data sets in a distributed computing environment [2]. Hadoop is a software where applications are fragmented and these fragmented files are called

blocks. These blocks are executed parallel using cluster machines. Hadoop consists of main two components HDFS and Map Reduce.

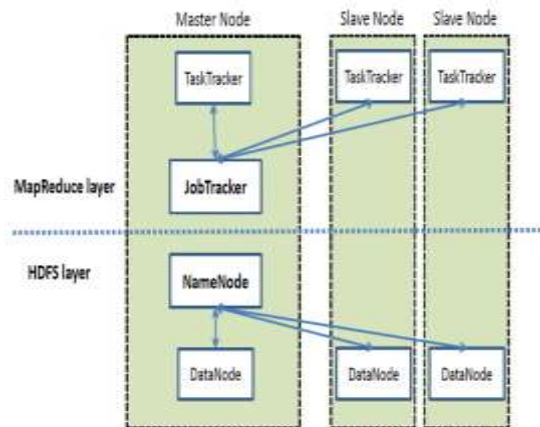


Fig 2: Hadoop Architecture

Hadoop includes a fault-tolerant storage called the Hadoop Distributed File System (HDFS). HDFS is a capable to handle large amount of data without losing it [2]. Hadoop creates clusters of machines and coordinates work among them. Clusters can be built with inexpensive computers. If any cluster is fail Hadoop can-not stop work they continue their work without losing data or interrupting work. It handles work by shifting to other machines. HDFS manages storage on the cluster by fragmented incoming files into pieces. And these fragmented files are called Blocks.HDFS stores three complete copies of each file by copying each piece to three different

CONCLUSION

This paper gave a description about big data problem in the environment of cloud computing. Here we discuss about problem of big data, applications and services which are used to handle big data problem like Hadoop and Map reduce . these problems helps to handle big data problem by fragmented data and executed parallel. Through this system reliability, execution is improved and complexity is decreased. 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. Elmustafa Sayed Ali Ahmed and Rashid A.Saeed, "A Survey of Big Data Cloud Computing Security" International journal of computing science and software engineering (IJCSSE) volume 3, 2014, ISSN: 2409-4285.
2. Harshawardhan S. Bhosale, Prof. Devendra P. Gadekar, " A Review paper on Big Data and Hadoop" International journal of scientific and research publications, Vol 4, Issue 10, ISSN 2250-3153.
3. Marcos D. Assuncao, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S. Netto, Rej Kumar Buyya, " Big Data computing and cloud: Trends and future directions" J.Parallel Distrib. Comput. 79-80 (2015) ELSEVIER.
4. Bo Li, boli "Survey of Recent Research Progress and Issues in Big Data", CSE. Wustlr.edu
5. Amit goyal, Sara dadizadesh, University of British Columbia, Vancouver CS-508, "A Survey on Cloud Computing", The Apache Hadoop Project. Hhttp://hadoop.apache.org, 2009.
6. Pardeep Bhosale, Priyanka Deshmukh, "Enhancing Data Security in Cloud Computing using #D Framework and Digital Signature with Encryption," International Journal of Engineering Research and Technology 2012 ISSN: 2278-0181.
7. Venkata Narasimha Inukollu, Sailaja Arsi and Srinivasa Rao Ravuri, "Security issues associated with Big

Data in Cloud computing” International journal of network security & its applications (IJNSA), Vol. 6, No. 3, May 2014.

8. Shilpa, Manjit Kaur “Big Data and Methodology- A review” International journal of advanced research in computer science and software engineering” (ijarcsse) volume 3, Issue 10, October 2013, ISSN: 2277 128X
9. Venkatesh H, Shrivatsa D Perur, Nivediita Jalihal “A Study on Use of Big Data in Cloud Computing Environment” International Journal of Computer Science and Information Technology, Vol 6, 2015, 2076-2078, ISSN: 0975-9646