



Big Data Analytics of Health Care in Nepal using Cloud Computing Environment & Machine Learning Technique

Dambarbhadur Rawal¹, Dr. Pharindra Sharma²

Department of Computer Science and Engineering, Nirwan University,
Jaipur, Rajasthan, India.

ABSTRACT- Now a day Cloud computing play important role in the market. In this research paper we analysis health care day in Nepal using machine learning and cloud computing. One of the best parts of Cloud is to share data in a common center by using internet with the help of various cloud providers. Cloud computing is very good and power full technology for complex computing using Cloud environment. In this era Cloud Computing using big data tool using hadoop & machine learning technique used to analysis various health care data to find accurate result from these tools.

KEY WORDS: Big Data Analytics, Health Care, Cloud Computing, Virtualization, Hadoop, Machine Learning Technique.

1. INTRODUCTION

One of the big challenge things in the world, we have large amount of data coming from various industry like manufacturing, health care center, IT industry, social media platform, data forecasting etc. For these types of data coming from various industry have problem like data storage, data processing and data computing issue are facing. This type of problems is avoided to using big data analytic tools, cloud computing and Machine learning technique. In normal life we know about cloud means, water molecules. Cloud Computing is like a network connected from different part. The user can use various module of cloud to storage data, to analysis data, but from cloud we have to pay on demand using by user. For this no need to take care of handling network just we need web browser to use cloud. There is different type of cloud available in the market.

TYPES OF CLOUD

1. Public Cloud: The public cloud is a cloud computing provided by third party which can be used by any user. This service is used by user use them and they have to pay whatever service used by user.

Private Cloud: This cloud is private used by organization on premise. Private cloud is used by the specific user which have only access and nor for common user. In this cloud there is higher security provide because this is private cloud.

Hybrid Cloud: This could is combination of public cloud and private cloud. In Hybrid cloud each cloud can manage independent but for applications and data can be shared in the hybrid cloud.

2. Virtualization

Virtualization is a new technology that come in picture their have no physical hardware available in our premises. This is also call as virtual machine. Virtualization can play import role in the cloud computing environment. It provides to produces the capability of pooling computing services from the server dynamically when we want.

3. PURPOSE OF STUDY

The main objective of this paper specifies to analysis health care data in Nepal using machine Learning technique i.e. supervised learning and cloud computing . In this paper we describe machine learning technique, data visualization tool and cloud computing environment. In this Cloud computing we have various storage to store data for health care and various data security in cloud.

Cloud service can provide three type of service based on:-

1. Infrastructure as a Service (IaaS): Infrastructure as a Service (IaaS) refers to providing infrastructure services such as physical computing resources, scaling, security, storage, and networking to its users.

2.Platform as a Service (PaaS): Platform as a Service (PaaS) refers to offering a development environment (typically like operating system, database, Integrated Development Environment, etc.).

3.Software as a service (SaaS): In the Software as a Service (SaaS) model, users can gain access to applications and software from anywhere in the world.

Characteristics of Cloud Computing

Security is one of the characteristic of cloud computing. Data security is main things.

On-demand self-service:- It is based on a self-service model where users manage their services like- allotted storage, functionalities, server uptime, etc., making users their own boss. Cloud service providers do not tell users how to manage their services, they only charge them at the end of the billing cycle based on their usage of the services.

Cost-effective:- This is cost effective because this is pay on use.

Resource pooling: This is providing by the cloud. Computing resources (like networks, servers, storage) that serve individual users can be securely pooled to make it look like a large infrastructure.

Cloud computing provide easy data backup recovery, only pay for whatever you used, low maintenance cost etc.

Cloud Provider

There are various clouds available in the market today are AWS, IBM Cloud, Microsoft Azure, Google Cloud, Sales force, Alibaba Cloud, etc.

4. Big Data Analytics

As mentioned in earlier, the complexity of big data is very huge which is unable to process using commonly used software tools? Also data sizes are continuously increasing and changing from time to time. Hence, the used technique for analyzing data is called big data analytics. The process of collecting, keeping and analyzing the data with the aim of revealing hidden patterns, unknown correlations and other facts is called big data analysis [8][4]. Analyzing big data is needed analytical capabilities as well as optimal processing power [3]. In big data analysis, raw format data are converted into standard format with the support of tools [12] [20]. This is a process of information gathering, data analysis, visualization and scheduling [12].

5. BIG DATA

Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size. Big Data is becoming one of the most talked about technology trends nowadays. The challenge with the big organization is to get huge amount of data is available and for future what data is coming. How to handle existing data and make it meaning full that provide

accurate analysis of the data for future business. Now a day's maximum company move into big data to handle huge amount of data for that organization. As organizations are evaluating and architecting big data solutions they are also learning the ways and opportunities which are related to Big Data. There is not a single solution to big data as well there is not a single vendor which can claim to know all about Big Data. Big Data is too big a concept and there are many players – different architectures, different vendors and different technology. The big data is classified into 3 types

1. **Structured Data**
2. **Unstructured Data**
3. **Semi-structured Data**

Characteristics of big data

There are 4 characteristic given below

1. Volume: - The volume means size of data. Whether a particular data can actually be considered as a Big Data or not, is dependent upon the volume of data.

2. Variety: - The variety of data means table format, file, xml format, text format etc.

3. Velocity: - The Velocity means speed of data. How many second? We have to fetch data from various centers.

4. Variability: - This refers to the inconsistency data which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

6. APACHE HADOOP

Apache Hadoop is an open source framework that is used to efficiently store and process large datasets ranging in size from gigabytes to petabytes of data. Instead of using one large computer to store and process the data, Hadoop allows clustering multiple computers to analyze massive datasets in parallel more quickly. Apache Hadoop is one of the main parts of the Big Data technology. This is a concept How to used big data in hadoop platform. Apache people continuously improve the hadoop. The Hadoop is a collection of library develops by Apache to handle Big Data.

MAPREDUCE

Map Reduce was developing by Google to create large amount of web search indexes. This is framework to write applications that processes a large amount of structured or unstructured data over the web. This takes the query and breaks it into parts to run it on multiple nodes. By distributed query processing it makes it easy to maintain large amount of data by dividing the data into several different machines. Map Reduce is a software framework for easily writing applications to manage large amount of data sets with a highly fault tolerant manner applications. More tutorials and getting started guide can be found at Apache Documentation. Map Reduce provides Automatic parallelization, distribution,

I/O scheduling, Load balancing, Fault tolerance.

HDFS

HDFS is a distributed file system that handles large data sets running on commodity hardware. It is used to scale a single Apache Hadoop cluster to hundreds (and even thousands) of nodes. This is a java based file system that is used to store structured or unstructured data over large clusters of distributed servers. Now HDFS is a part of Apache Hadoop project. HDFS supports a traditional hierarchical file organization

HIVE

Hive was developed by Facebook, now it is made open source for some time. Hive works like a bridge in between SQL and Hadoop, it is basically used to make SQL queries on Hadoop clusters. There are two types of tables available in Hive – managed and external. Apache Hive is basically a data warehouse that provides ad-hoc queries, data summarization and analysis of huge data sets stored in Hadoop compatible file systems. Hive supports client applications based on Java, PHP, Python, C, and Ruby coding languages.

7. RELATIONSHIP BETWEEN CLOUD COMPUTING AND BIG DATA

Cloud computing and big data are combined. Big data provides practitioners the ability to use commodity computing to process distributed queries across large data sets and return result in a timely manner. Big Data and Cloud Data have a mutual relationship, as the Cloud infra effectively enables storage, real-time processing, and Big Data analysis to scale up and quickly. The big benefit of using Cloud storage for your Big Data is this scalability. Cloud storage is available on a pay-as-you-go basis. Analytical results presented visually through different graphs for decision making. Based on cloud computing big data utilizes distributed storage technology rather than local storage attached to electronic device. Cloud computing provide in build hadoop service used in the cloud.

8. BIG DATA MANAGEMENT SYSTEM

Many researchers said that Traditional DBMS is not suitable for huge amount of data to be stored. This is architecture to store limited amount of data in a DBMS tool. Big data management is a broad concept that encompasses the policies, procedures and technology used for the collection, storage, governance, organization, administration and delivery of large repositories of data. It can include data cleansing, migration, integration and preparation for use in reporting and analytics.

9. CONCLUSION

Currently, big data become more popular and powerful with development of the technology. This paper is used to analysis of health care data using big data tools cloud

technologies and machine learning. We have to analysis of data by using machine learning technique and read data from cloud storage service.

KNOWLEDGMENT

My sincere gratitude goes to Dr. Pharindra Sharma, Nirwan University, Department of Computer Science and Engineering who gave advices and guidance to complete this work at this level. Also, I am thankful to the people who have published researches on this area. Not only that, I would like to thank my parents who supported me.

REFERENCES

1. R. L. Villars, C. W. Olofson, M. Eastwood, Big data: what it is and why you should care, White Paper, IDC, 2011, MA, USA.
2. D.E. O’Leary, Artificial intelligence and big data, IEEE Intell.Syst.28 (2013)96–99.
3. M. Chen, S. Mao, Y. Liu, Big data: a survey, Mob. Network. Appl. 19(2)(2014)1–39.
4. IEEE, vol. 13
5. R. Cumbly, P.Church, Is Big Data creepy? Comput.LawSecur.Rev. 29 (2013)601–609.
6. Y. Cao, C. Chen, F. Guo, D. Jiang, Y. Lin, B. Ooi, H. Vo, S. Wu, and Q. Xu, “Es2: A cloud data storage system for supporting both oltp and olap,” in Data Engineering (ICDE),2011 IEEE 27th International Conference on. IEEE, 2011,pp. 291–302.
7. F. Chang, J. Dean, S. Ghemawat, W. Hsieh, D.
8. Wallach M. Burrows, T. Chandra, A. Fikes, and R. Gruber, “Big table: A distributed structured data storage system,” in 7th OSDI,2006, pp. 305–314.
9. Aguilera, M. K., A. Merchant, M. A. Shah, A. C.
10. Veitch, and C. T. Karamanolis, “Sinfonia: A New Paradigm for Building Scalable Distributed Systems,” SOSP 2007.
11. Aguilera, M., W. Golab, and M. Shah, “A Practical
12. Scalable Distributed BTree,” VLDB 2008. [23] Amazon Elastic Compute Cloud: <http://aws.amazon.com/ec2/>, Retrieved date: Sep. 27, 2009.
13. Volume No: 1(2014), Issue No: 9 (September) ISSN No: 23484845
14. Apache Hadoop, <http://hadoop.apache.org/>, Retrieved date: Sep. 27, 2009.
15. Apache HBase, <http://hadoop.apache.org/hbase/>,
16. Retrieved date: Sep. 27, 2009. [28] Sang Woo Han, Jong Won. A multiagentbased management system for pervasive collaborative computing environment[C]. IEEE international conference on computing and communications.USA: Institute of electrical and electronics engineers, 2009:16.
17. Han Xu, CaoYongcun. The application of computersupported collaborative technologies [30] Han J, Kamber M. Data mining concepts and techniques [M]. San Francisco: Morgan kaufmann, 2006. Massachusetts: MIT press, 2004:191211.
18. S.Kaisler,F.Armour,J.A.Espinosa,W.Money,BigDat
19. S. Chakrabarti, “Mining the web: Discovering
20. Knowledge from Hypertext Data”, Morgan Kaufmann