Managing Cloud Server with Big Data for Small, Medium Enterprises: Issues and Challenges

Prerita Gupta
Research Scholar, DAV College, Chandigarh

Dr. Harmunish Taneja
Department of Computer Science and Information Technology, DAV College, Chandigarh, India

Dr. Gagandeep Singh Brar
Department of Computer Science and Information Technology, DAV College, Chandigarh, India

Abstract - The increasing demand of cloud applications and cloud data store leads to the emergence of big data cloud. It has various advantages over conventional physical deployment technologies. Big data cloud can work properly with large projects. Internet of Things (IoT) consists of connectivity of billion of devices that has the ability to sense, communicate and operate other devices. This paper covers the study about big data cloud reshaping the Information and Communication Technology (ICT) for economy in New Zealand for future coming years. The relationship between the cloud computing and big data, technologies and trends of the big data is also discussed.

Keywords: Internet of Things (IoT), Hadoop, Cassandra, Pentaho, Couchbase.

I. INTRODUCTION

In today's era, Big data cloud is an emerging trend in the area of cloud computing. Big data shows its existence in various public and private sectors where right amount of data is produced that result in making on-time decisions, and full utilization of the operations thereby saving money, enhancing outcome and optimizing the profits for the organization [2]. It helps to manage the huge volume of data that is difficult to store, analyze and exploit the traditional database technologies also. The "big data" is a new technique in the area of IT and businesses but merely a new term for researchers or investigators [1]. Traditionally it was very hard to process the large amount of data. Big data stores the various forms of it which includes structured, unstructured and semi-structured. Fig.1 shows the architecture of big data which consists of three layers namely data storage, data processing and data analysis.

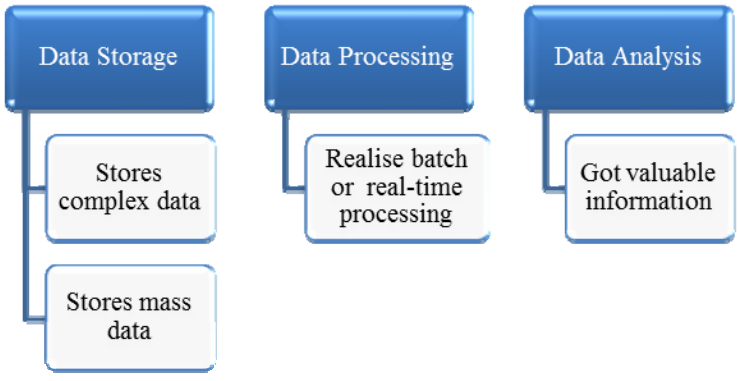

Fig. 1: Architecture of big data

Big data is specified by 6Vs (velocity, variety, volume, veracity, validity).

1. *Volume:* The huge amount of data which is unique in nature. Big data needs to process low-density data such as network traffic on web page and many more. Big data converts low-density data to high-density data. Data might be in petabytes or exabytes that has some value [3]. By data analysis the valuable information and hidden patterns can be extracted from the large amount of data.
2. *Variety:* It refers to the different forms of data collection via social networks. The different types of data could be videos, audio, text, images, etc either structured or unstructured data. Some relevant examples of unstructured data include blogs, internet games and many more whereas online users can also create a variety of structures as well as non-structured type of data. Around 80% of the data is in unstructured form [4].
3. *Velocity:* Data transfer speed is known as velocity i.e. speeds with which the new data generated and travels. In today's era various social media websites are growing vigorously on internet the technology allows us to analyze and generate the data without storing it in the database [5]. The transfer speed matters because of the constant change of the data due to the frequent streaming of the data from the various sources.
4. *Veracity:* It refers to the irregularity, disturbances and deformities in the data. This is the toughest problem to solve and also difficult to understand. There should be some strategy to figure out to keep data clean and processes to restrict dirty data to enter the system [4].
5. *Validity:* Validity means to what extent data is correct and accurate for further use. The valid data or accurate is the key for making right decisions.
6. *Volatility:* Volatility refers to for how long the data is stored and is valid.

Fig. 2: 6V's of Big Data

These 6V's are highly recognized because it shows the exact explanation and the necessity of big data.

1.1. Technologies for Big Data

NoSQL (not-only SQL)*:* A non-relational database which is designed for distributed data stores where there is a need of very large scale of data storage. Various social media websites i.e. facebook, LinkedIn and many more uses NoSQL technologies, as they consist of storing large amount of data. NoSQL database is programmed in various languages and normally associated with the companies which deal with the field of Artificial Intelligence, Computer Simulation, Computer-Aided Design and many more. It is basically a substitute for the old Relational Database Management System (RDBMS) [11].

Hadoop: It is carried out to solve the problem of handling the exploded data onto the internet. This technology is used to handle, store, process and analyze hundreds of petabytes or exabytes of data. No data is too big for Hadoop. This computing data model processes the big data rapidly moreover it is flexible, scalable and free to use. This java enable software from Apache has two main components: Hadoop Distributed file system (HDFS) and MapReduce framework [1][2].

1. The HDFS is highly fault- tolerant and can be deployed easily on client machine which consists of low cost hardware.
2. MapReduce is basically a programming model by Google and is concerned for processing and results in producing of large data sets associated with parallel, distributed algorithms on cluster.

Latest Databases: It consists of various databases which are emerged to handle various kinds of data with latest techniques and technologies. Table 1 shows various databases that are followed by many business companies now days.

Details / Database	Logo	Company	Features
Cassandra2.2		Apache	• Use of insert and select JavaScript Object Notation (JSON) data. • Able to use User Defined Functions (UDF). • Can customize the aggregate functions • Eliminates Cassandra Query language 2(CQL2) and uses Cassandra Query language (CQL3).
Couchbase3.0.1		Couchbase, Inc.	• Clients can export the relevant data from Couchbase to Hadoop and vice versa. • Robust capability of managing large datasets. • Database Change Protocol (DCP) can be able to create changes in the duplicate nodes and data centers through streams. • This version can able to recover the node at faster rate.
Pentaho5.0		Pentaho corporation	• Better data integration with enhanced NoSQL support. • It enhances the navigation from previous versions • Easily analyze the various types of data with accurate results.

Table 1: Databases and its features

Cassandra2.2: This database is created by Apache that sets a perfect example of the case for handling big amount of data. The queries runs on Cassandra Query Language (CQL) interface [8]. The basic structure of Cassandra 2.2 consists of a node, data center, cluster, commit log, Sorted String Table (SST) and CQL table. The advantage of this version is that its client can connect to any node with any data center and use it any time [8].The concept used in reading and writing in Cassandra is based on ACID(atomicity, consistency, Isolation ,Durability) properties and thus avoid reading before writing which ends to the performance problems and many more. Table 1 shows that cassandra2.2 will be helpful to overcome the management of the big data in various cloud applications.
Couchbase3.0.1: It is considered as a most performing NoSQL distributed database. This server satisfies the need of the enterprises which instated with the distributed database performances and its expandability. Its major benefit is that it is flexible enough to deploy its services in a cloud infrastructure and also makes a secure connection between various data centers [9].
Pentaho5.0: Pentaho 5.0 can directly transform the data using SQL transformations. It is the easy way of analytics with updated interface. It combines the big data at the source to reach the accurate results on the destination. The enterprises can easily embed and integrate the data in this version into any cloud environment [10] [13].

This paper focuses on the fundamentals of big data. Various technologies that are used in big data are discussed in Section 1. In Section 2, the trends in big data are discussed. Section 3 includes the survey of New Zealand big data cloud in ICT sector and paper is concluded and the future scope is discussed in Section 4.

II. TRENDS AND ISSUES IN BIG DATA

Big Data in year 2015: Big data exert into the enterprise where most of the cases are on real-time uses [6]. Many enterprises migrates big data deployments into real-time systems from traditional batch systems and the further

development includes data agility, movements of large data for processing on data platforms, acceptance of big data self-service, sellers new business model (Hadoop).

1. *Data Agility* is the key for development of big data technologies as the transactions around legacy databases and data warehouses are not too much flexible and are too slow in the businesses. To structure the unstructured data, Database Administrator (DBA) requires these kinds of legacy databases and data warehouses. The cost factor delays the access of these new data sources and becomes very hard to modify the data over the time. This results in legacy database hard to move quickly in the data organization. Initially, most of the big data projects focused on the storage of target data sources instead focusing on its management. Enterprises should concentrate on measuring data agility. They should also focus on the ability of processing and examine the outcomes from processing.
2. *Movements of large data to data platforms* consider that the data hub consists of various types of data in different forms. A data hub constitutes of a scalable infrastructure in minimum cost and rapidly moves the data from various data platforms [7].
3. *Acceptance of big data self-service* will ease the life of IT users.

The data usage is phenomenally rising in 2015 as compared to last 4 years [6]. The IT companies should look forward to the cloud computing structure which supports the big data projects. It is recorded that 80% unstructured data and 20% of structured data is transmitting on the network [6]. One of the main focus of IT mangers is to move the big data from batch systems to real-time systems. The trend of Big data cloud computing includes Big Data-as-a-Service (BDaaS), with the combination of cloud hosted services which offers services like analysis of the huge amount of complex data . In the Cloud computing architecture, big data needs to be specified so that businesses can took advantage of it [14]. IT managers focus on the new ways to store the clean data and the security protocols which will be the major point in big data cloud. [7] Focus on engines running to process the data in a single space which will create a point of security and to access the data efficiently. Now the point of view is on the integration of the various technologies like Hadoop into large scale processing platforms as well as the continuous accessing and the events on the real time data systems.

III. BIG DATA CLOUD IN NEW ZEALAND FOR SMALL ,MEDIUM ENTERPRISES(SMES) AND ITS CHALLENGES

The big data is reshaping the New Zealand government policy and is expected to be beneficial for future of New Zealand. New Zealand needs the answer for the privacy and the power of the shared data. The ICT sector of New Zealand is growing rapidly and the challenge is to manage this speed to provide proper services to the clients. For most of the New Zealand based business companies analysis of the data is very crucial to their current business strategy and almost 41% of companies have already incorporated the big data. The problem which is arising is that business hubs want to access the data from the data lakes which will be further beneficial for the outcomes of the businesses [15].

There are umpteen challenges for researchers in big data. The foremost challenge is related to the storage issue. As the different types of data transmit from source to its destination and this data can have various forms. The problem arises where to store the data. Another problem is related to the management issue i.e. managing the big data needs to be solved. Traditionally there are various protocols that runs in sequence which ensures the validity and accuracy. It is almost impossible to get the validate data every time by seeing the amount of the data using now days. Their do not exists the perfect way of managing the data till date. The new tools and analytics should be evolved to solve the problem related to the performance. The effective processing of the petabytes or zettabytes of data requires data analytics which provides the time to time information to its clients. Still there are unknown challenges that could arise if there is an expansion and new innovations in analytics [16].

IV. CONCLUSION AND FUTURE SCOPE

In this paper the various technologies and trends are studied related to the big data. The latest technologies using the big data concept is adapted in the business quickly. The ICT has faced some issues related to big data which is related to the storage, management and performance problems. The issues mentioned in the paper should be monitored for the proper and accurate management of the large amount of data. The latest big data technologies and different problems mentioned in this paper are the pinpoint for the researchers and practitioners. They should know

the techniques for better management of the big data Cloud**.** The issues mentioned in this paper will be the further challenges for the researchers to get the optimal solutions**.**

REFERENCES

[1] Ibrahim Abaker Targio Hashern , Ibrar Yaqoob, Nor Badrul Anuar, Salimah mokhtar, Abdullah Gani, Samee Ullah Khan, " The rise of "Big data" on cloud computing: Review and open research issues", PP 98-115, 2015
[2] Arkady Zaslavsky, charith Perera, Dimitrios georgakopoulos, " Sensing as a Service and Big Data"
[3] http://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf
[4] http://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/
[5] http://www.slideshare.net/BernardMarr/140228-big-data-volume-velocity-variety-varacity-value
[6] http://www.intel.com/content/dam/www/public/us/en/documents/product-briefs/big-data-cloud-technologies-brief.pdf
[7] http://www.cio.com/article/2862014/big-data/5-big-data-technology-predictions-for-2015.html
[8] http://docs.datastax.com/en/cassandra/2.2/cassandra/features.html
[9] http://www.couchbase.com/nosql-databases/about-couchbase-server#RealTimeBigData
[10] http://www.graymatter.co.in/BlogPost/6-reasons-to-migrate-to-pentaho5-0
[11] http://www.hindawi.com/journals/ijg/aa/502795/
[12] https://en.wikipedia.org/wiki/MapReduce
[13] http://www.pentaho.com/5.0
[14] https://www.bbvaopenmind.com/en/cloud-computing-big-data-and-mobility-2015-tech-trends/
[15] http://www.scoop.co.nz/stories/BU1505/S00598/big-data-comes-of-age-for-new-zealand-businesses.htm
[16] Stephen Kaisler, frank Armour, J.Alberto Espinosa, William Money, "Big data: Issues and Challenges Moving forward", Hawaii International Conference on System Sciences, PP 995-1004, January 2013.