International Journal of Advanced Trends in Computer Applications
www.ijatca.com

Issues and Benefits in Big Data and Cloud Computing

[1]Prof. Sapna Shukla, [2]Prof. Prabhdeep Kaur
[1,2]University Institute of Computing,
Chandigarh University, Gharuan, Punjab
[1]uic.sapnashukla@gmail.com, [2]uic.prabhdeeplobana@gmail.com

Abstract: *Cloud computing is powerful technology by eliminating the need of infrastructure to the organizations. It is expected that by end of 2016 half of the data will be on cloud. The amount of this data is so huge that large and expensive infrastructure is needed for its successful analysis. This paper is an attempt that gives description of use of cloud computing in Big Data analysis and also defines the issues and benefits of big data in cloud computing .Cloud computing can provide infrastructure to small scale organizations on lease basis, so that they can have its services without high investments. However there are issues of cloud computing such as security, ownership and privacy etc. In this paper the authors discussed about Big Data, Hadoop, Map Reduce, relationship between big data and cloud computing. Benefits and issues of using cloud computing in big data analysis.*

Keywords: *Big data, Map Reduce, Hadoop, Cloud Computing, Big Data analysis.*

I. INTRODUCTION

Data is generated everywhere around us all the time. The amount of this data is so huge that it is termed as Big Data. This data can be is structured, unstructured or semi structured format. Every digital processing, social media exchange produces it where as mobile phones, sensors are the devices that transmits it. This data is arriving from different resources at very high velocity, verity and volume and to extract the meaningful information from this data we need optimal processing, analysis capacity and skills.[5]

The importance of big data does not revolve around how much data we have, but what we do with this huge data. We can take the data from any source and analyze it to find some answers that enable us i) reduced cost (ii) reduced time (iii) new product development (iv) decision making. When we combine big data with high power analysis we can accomplish business-related tasks such as:-

1. Determining route cause of failure, issues and defects in real time.
2. Detecting fraudulent behavior before it effects the organization.
3. Generating coupon on the time of sale based of customers buying habits.[7]

Big data refers to the techniques and initiatives that revolve around diverse, fast changing and massive data for which conventional technologies, skills and infrastructures fails. In saying differently velocity, verity and volume of data is too great to handle conventionally. But today, new technologies make it possible to extract the information from Big Data

Cloud computing refers to the broad set of computing and software that can be sold as a service and managed by third party provider and delivered on the network. Infrastructure as a service, in which on demand processing, network resources and storage is provides to the customer. Customer can pay only for the capacity he is using in the organization rather than in house hosting where he has to pay for the infrastructure where he is using it or not. Benefits of cloud computing as IAAS are flexibility, reliability scalability and in expensiveness. [10]

II. LITERATURE REVIEW

Ibrahim Abaker Targio Hashem [1] depicts the rise of big data in cloud computing .He discussed about Big Data, its classification and introduce cloud computing. Paper highlights about Big Data storage system and technologies like Hadoop, the Challenges of Big Data processing like heteroginity, scalability, volume, data protection, Data transformation, privacy and legal issue etc. further the key issue of big data in cloud were discussed like security and data ownership. It also discussed the background of Hadoop and its components like map reduce and HDFS.

Samiya Khan [2] discussed as the amount of data in increases day by day, the storage of big data is real

issue for the organizations. The cloud computing being an solution to storage of huge amount of data, still have challenges like security , privacy ,lack of ownership and control on the data. There is considerable decrease in the trust as there is a chance of security breach, ownership control and privacy violation, when big data is implemented on the cloud. The paper also discussed about the traditional data mining techniques like OLAP, BPM, mining and data base system like RDBMS.

Caesar Wu,[3] discussed about 9 'V's of Big data instead of 3'V's of big data, also incorporated business intelligence (BI) so that big data analytics ac be used to predict the crucial decision or research results.

Venkata Narasimha Inukollu [4] Focuses on the security issues of cloud computing, Big Data, Hadoop and Map reduce environment. It includes the various solutions for the issue in cloud computing security and Hadoop. The cloud computing security includes data privacy, network security and security of information. With the help of Big Data tools cloud computing plays a vital role in protecting data, applications and the infrastructure.

Marcos D. Assunção [6] Highlights the cloud analytics on the applications of big data. It discussed the four areas (i) Data Managing architecture (ii) Development model (iii) User interaction (iv)Business model. It identifies the possible gaps in the technologies and recommendations to the researchers for the further directions on cloud based big data analytics and computing

Venkatech H.[8] defines how cloud computing is powerful and complex, which eliminates the need to maintain expensive computing hardware, space and software. It is observed that addressing huge amount of data is time demanding and challenging task and also requires big infrastructure to ensure successful data analysis. The paper has discussed about Hadoop, applications of big Data, challenges of big data, Map Reduce and advantages of big data.

Olukunle A Iyanda[9] introduced detailed analysis between Big Data and Cloud Computing. Big Data as a data analysis methodology requires huge amount of hardware and processing resources, making its cost very high. Cloud computing is s set of services, which can provide infrastructure on network for lease basis, which we can scale up the service requirement if needed. Its advantages are scalability, flexibility, efficiency and outsourcing non-core activities. It offers business model of the origination to have services without high investments. The use of big data cloud provides benefits to small sized companies to increase the business without implementing in-house technologies.

III. BIG DATA and HADOOP

Big Data simply means massive data which is being generated around us every time. The amount the data is generated through social media, digital transactions and the data is transmitted using electronic devices from one point to other such as mobile phones. The data can be categorized in the Types:

1. **Structured:** The data which can be stored in relational Data base systems is structured data. Examples of the structured data such as ERP data, backup storage for large volume of data or tabular data.
2. **Semi structured:** The data stored in XML files is semi-structured data. for example call center logs
3. **Unstructured:** The data stored in media files such as documents, images , pdf files all the unstructured data

We can have data from any source like black box data, transportation data, social media data, power grid data, or search engine data.

Some 'V's are used to categorized the big data on the basis of volume, variety and velocity. The different type of the data in datasets can be termed as variety, the size of data in dataset can be termed as volume of data and the speed with which the data is being generated is termed as velocity of data.[2]

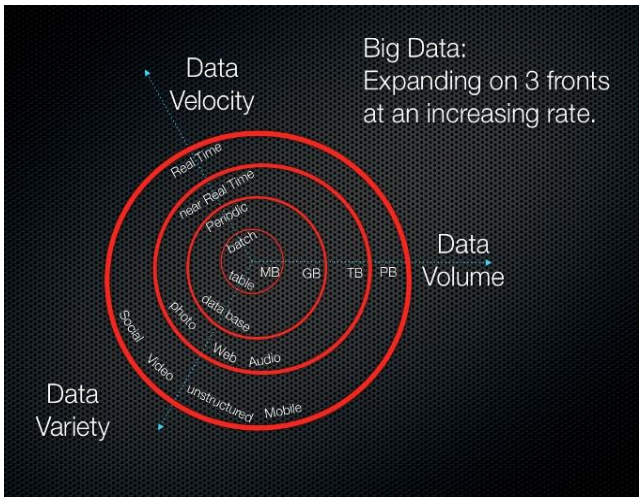

Figure 1: Big Data Characteristics [13]

Big data requires platforms like Hadoop to store the dataset on the distributed clusters and map reduce to coordinate and process and combine the data from multiple resources.[13]

Apache Hadoop is framework which is used to develop data processing applications which are executed on distributed environment. Same manner as data reside in the local file system, the data is Hadoop resides in the distributed file system called HDFS (Hadoop Distributed file system).

Apache Hadoop is consisting of two subprojects:

Hadoop Map Reduce: Hadoop MapReduce is computational model and software framework which is used to write the applications that runs on Hadoop

framework. There MapReduce programs are capable of processing the parallel on the large cluster of computation nodes.

Hadoop Distributed file system: MapReduce uses the data from HDFS. HDFS takes care of the storage part of Hadoop Framework. HDFS create the replicas of data blocks and distribute them on the compute nodes. HDFS clusters are further divided into data node and Name nodes. Name Node manages the meta data where as data node the used to store the actual data in it.[14]

IV. CLOUD COMPUTING

Cloud computing is virtualization based technology which reduces the cost of IT infrastructure. It provides the solution to IT infrastructure at low cost. It enables small scale organizations to take the it infrastructure on lease basis on network where third party is taking care of infrastructure on relatively low cost. The small size organization has to pay only for the capacity they are using rather that in-house infrastructure where they have to invest money whether they use it or not.

Cloud computing service models

SAAS (Software as a service): This is software distribution model, in which application is made available to the customer through cloud over internet. This model is also known as "**on demand software**". Saas is accessed by users using thin client or using web browsers.

IAAS (Infrastructure as a service): This is one of the layers of cloud computing where organizations outsource IT infrastructures like servers, network, processing, storage, virtual machines and other resources. Customers access the resources on the network and pay according to their uses.[15]

PAAS (Platform as a service): It proves the runtime platform for application to run, develop and deployment tool etc.

Benefits of cloud Computing

1. Can configure the application online at any time
2. Can access the application utilities, over the internet
3. It offers online development and deployment tools.

It offers on demand self services.[14]

V. RELATIONSHIP BETWEEN CLOUD COMPUTING AND BIG DATA

Big data and cloud computing is related to each other. Big Data provides user the ability to compute and process the distributed queries on the multiple datasets on distributed clusters where as cloud computing provides the infrastructure, through the use of Hadoop, a class of distributed data processing platform. Large data sources from web or browser are stored in distributed datasets clusters and processed through algorithms in clusters [1].

Cloud computing dignify "As a service "Model by hiding the complexities and challenges involved in building the elastic self service applications. The same is the requirement for Big Data. Hadoop in similar way hide the complexities of the large scale distributed processing from the end users. The user write the MapReduce codes an able to do seamless crunching without warring about the complexities of the node failures, replication, linear scalabilities or fault tolerance etc. thus the simplification provided by cloud computing and big data is a reason of adoption of big data and cloud.[10]

In view of cloud based big data analysis, additional challenges like adoption, implementation of effective big data solution using cloud computing are security and privacy risks. One of the biggest concerns is security while cloud based big data analysis. This is the reason why cloud based big data analytics has attracted immense attention [2]

Not only security but data privacy challenges exists in industries and organizations.

With the increase in the use of big data in business, many companies are fighting with privacy issues. But privacy should be considered as an asset; therefore it becomes a selling point for both stakeholders and companies. [4]

CONCLUSION AND FUTURE WORK

The objective of this paper is survey and overview on big data, its challenges, characteristics and issues. This paper initiates the review of big data analytics using cloud computing its benefits and issues. The author reviewed Hadoop, Map reduce, big data analytics and use of cloud computing in big data. The future research work will be focus on the cloud based big data analysis more challenges and issues of big data analytics.

References

[1] Ibrahim Abaker Targio Hashem, IbrarYaqoob , NorBadrulAnuar , Salimah Mokhtar , AbdullahGani , Samee Ullah Khan - The riseof "big data" on cloud computing:Review and open research issues

[2] Samiya Khan, Kashish Ara Shakil ,Mansaf Alam-CLOUD-BASED BIG DATA ANALYTICS – A SURVEY OF CURRENT RESEARCH AND FUTURE DIRECTIONS

[3] Caesar Wu, Rajkumar Buyya, and Kotagiri Ramamohanarao-Big Data Analytics = Machine Learning + Cloud Computing

[4] Venkata Narasimha Inukollu , Sailaja, 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
[5] http://www.ibm.com/big-data/us/en/
[6] Marcos D. Assunção, Rodrigo N. Calheiros , Silvia Bianchi , Marco A.S. Netto ,
Rajkumar Buyya-Big Data computing and clouds: Trends and future directions.
[7] http://www.sas.com/en_us/insights/big-data/what-is-big-data.html
[8] Venkatesh H, Shrivatsa D Perur, Nivedita Jalihal-A Study on Use of Big Data in Cloud
Computing Environment, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2076-2078
[9] OlukunleA Iyanda- Big Data and Current Cloud Computing Issues and Challenges,ISSN(Print) 2277 128X, Volume 4, Issue 6, June 2014
[10] Cloud and Big Data: Is there a Relation between the two: http://www.infosysblogs.com/cloud/2013/05/cloud_ and_big_ data_is_there_a_.html
[11] https://www.mongodb.com/big-data-explained
[12]http://www.tutorialspoint.com/hadoop/hadoop_big_data _overview.htm)
[13]http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data
[14]http://www.guru99.com/create-your-first-hadoop-program.html
[15] http://www.javatpoint.com/cloud-computing-tutorial