ISSN: 0474-9030

Our Heritage

Vol-68-Issue-15

January-2020

UGC Care Listed Journal

A Study -Challenges of Big data storage and its Techniques

Mr.DigambarSawalkar[1],Dr.Manisha Kumbhar[2]
Assistant Professor,Indira Institute of Management,Pune
Professor,Sinhgad Institute of Management,Pune

Abstract:
Inlatestyears, bigdataisone of the interesting topicfor researcher and industry.Growth of big Data is very fast because need to study and find out a solutions and itwillhelp to manage the data and systematically extract value of data and gain the usefulknowledge from datasets.Inthispaper, we focuses on concept of bigdata and also identifying challenges inbigdata storage and its techniques. The prime objective of this paper is to study the big data challenges and itstechniques associated with it.

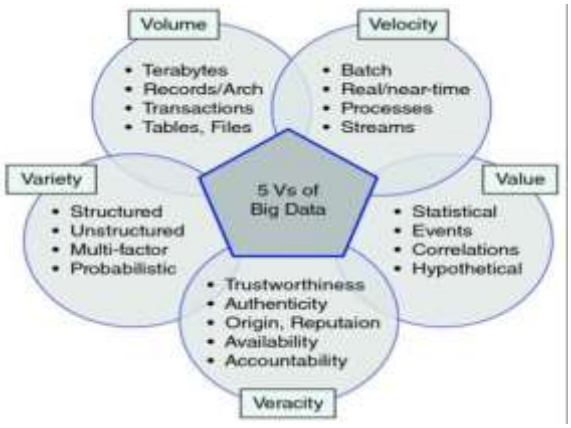
Keywords : *Big data, Hadoop, Map Reduce.*

1. Introduction

Today the big data is a buzz word and still in developing stage. Managing business data, basically the problem of initial value is considered as a case of big data.Inthis paper we are discussing Challenges of Big data storage and its techniques.Big data is a term that describes collection of data and it's too large and very complexand it becomes difficult to process by traditional data processingapplications. Big Data refers to new techniques for analyzing, storing massive or complex data.Thebenefit of big data isto increase theprocessing of database systems. Thedata is too large, it growsvery fast, it is difficult to fit in the structuresof database architectures and it also helps for gaining value from thesebig data, it should be alternative way to process it.

2. Characteristics of Big data

The characteristics of the big data arelarge Volume andspeed of data generation is very fast means velocity and generated data in different format means Variety**.** Figure 1.



Big data originals model is 3V model as follows

1. Volume: Big data is high volumes of data. The volume of big data to be analyzed is massive.
2. Variety: The variety means that the data sets have many different types of data, including structured and unstructured.It is useful for researcher for analyzing the data and willusethedata for their work.
3. Velocity: new data is generated with high speedand data will process to complete the challenges for development.
4. Veracity:It helps to filter through which data important or not and it also generated understanding of data.
5. Value:It is good to access valuable big data.

3.Challenges of Big data

1. **Infrastructure**
   Data needs a place for storage, like objects need a container; data must occupy space. If you are storing large amounts of data, it's necessary to have infrastructure to store it, which means investing in high-tech servers that will occupy space in your office or building. One of the easiest way to use cloud hosting and cloud storage, which take advantage of another company's infrastructure to save you that space and the trouble of setting things up yourself. Figure.2

2. **Cost**
   Running your own data center is an expensive operation. You'll need to spend money on initial setup, ongoing maintenance, and the costs associated with the people responsible for maintaining it. Again, the best solution here is to outsource the work; you'll probably have to pay a monthly fee, but it will save you money in the long run.
3. **Security**
   Security is a major issue to overcome. Hypothetically, if your data is stored somewhere, it's possible for a third party to obtain it. There are some layers of security that can prevent from unauthorized access, including encryption and reliance on third-party providers, but there is also a limit to how well these can protect youeven the FBI has trouble maintaining the security of its data when its own best practices aren't followed. You will need to run a tight operation, choosing the best partners and keeping your own team adhering to best practices at all times.
4. **Scale**
   You might be able to find a storage solution that serves your current needs adequately, but what happens if those needs change suddenly? How will you account for your needs as they stand in 5

years? For data storage needs some capacity to scale. Here, it pays to give yourself as many options as possible, since you won't be sure exactly how your needs will change in the future.

4. Techniques for big data handling

There are many techniques available for data analysis and data handling in big datalike Simple DB, Not NoSQLbut these traditional approaches are only applicable to traditional data and not big data as it cannot be stored on a single machine. These are the Big Data handling techniques -Hadoop, MapReduce. Out of these, Hadoop is one of the most widely used technologies.

4.1 Hadoop

Hadoop is an open source framework and High volumes of data are processed by Hadoop. Hadoop is used for storage and processing for data sets. The following are components of Hadoop.

1.HDFS

Hadoop has a distributed File System called HDFS, which stands for Hadoop Distributed File System. It is used for storing very large files with streaming data access patterns, running on clusters on commodity hardware. There are two types of nodes in HDFS cluster, namely namenode and datanodes.

Name node –It manages the file system and directory .it also handles the metadata of files and directories.

Datanode-it stores and retrieve blocks as per the instructions of clients or the namenode. Whatever data isretrievedsent back to the namenode with lists of blocks. Without the namenode it is not possible to access the file. So it is very important to make name node resilient to failure.

2.Map Reduce-It is a programming paradigm for managing applications on multiple distributed servers. It is divide and conquer methodwhich is used to break the large complex data into small units and process them. It reads the data from HDFS and italso read the data from other sources like mounted local file systems, the web, and databases. It divides the readable data between different computers (servers, or nodes). It is also fault-tolerant. If some of nodes fail, Hadoop knows how to continue with the computation, by re-assigning the incomplete work to another node.

5. Conclusion

In current era Bigdata has impact in all sector and industry. In this paper, wehave briefly reviewed grand challenges that big data brings us.

In latest years,too large and complex data are generated. These data need to analyze and it is difficult using traditional database management tools. In this paper, we have studied challenges and techniqueswhich is usefulfor analyzing big data. This paper present that every big dataplatform has its individual focus. In future researcher need to give more attention tobig data techniques to solve problems of big data effectively.

References

[1]M. K.Kakhani, Research issues in big data analytics, International Journal of Application or Innovation in Engineering & Management, 2(8) (2015).

[2] A. Gandomi Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management, 35(2) (2015).

[3] C. Lynch, Big data: How do your data grow?, Nature, 455 (2008).
[4]Xiao longJinBenjaminW. WahXueqiChengYangzhouWang, Significance and Challenges of Big Data Research
[5] M.H.Padgavankar1, Dr.S.R.Gupta, Big Data Storage and Challenges, International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014.
[6] D. P. Acharya, A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, International Journal of Advanced Computer Science and Applications,Vol. 7, No. 2, 2016
[7] Neha Begam, Big Data Challenges and Techniques, International Journal of Engineering Science and Computing, April 2017