International Journal of Research in Advent Technology (IJRAT) Special Issue
E-ISSN: 2321-9637
National Conference "MOMENTUM-17", 14th & 15th February 2017
Available online at www.ijrat.org

Opportunity and Challenge of Big Data Storage, Integration, Security Practices

Ms. Tejashri Maruti Dumbre
HOD of Department Information Technology
tejashridumbre89@gmail.com

Ms. Pooja Somnath Popalghat
Information Technology
poojapopalghat1997@gmail.com

Ms. Rituja Vasant Dhamdhere
Information Technology
dhamdhererituja10@gmail.com

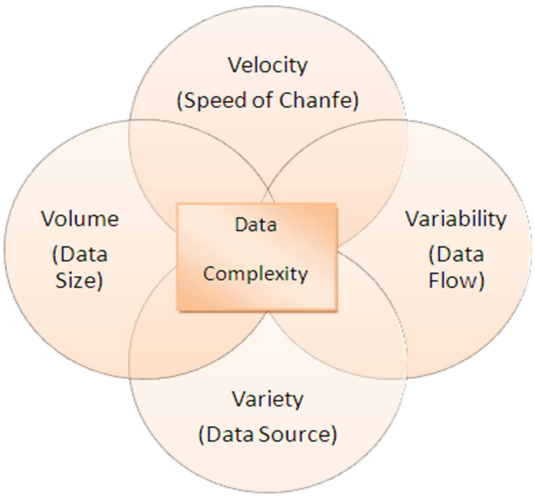
ABSTRACT
Big Data at is victimize the random illustration. It consist of Thousands of companies worldwide, we are impact on company performance. Big Data is the collection of data set and so large. Big Data is the complex we confirm and extends early results of a productivity impact form big data. Big Data can improve the big data and analysis the performance. Data representation means data is represented by the different way and it stored on a computer. Data demonstration refers the internal method and data with much unexpected attribute these can't be controlled and processed by the current traditional software systems, which became absolute problem. The several challenges and bare in adapting and accepting Big data technology, its tools are also discussed in detail along with the problems Hadoop is facing. The paper bringing up rear with the Good Big data practices to be followed. Process group includes all the challenges disagreement while developing the Big Data. Data challenges are the group of the perfectibility relates to the characteristics of the information itself. Big Data is started with compering the data and end with presenting output to the user. Many institutions are taking steps to focus the produce problems of advanced persist in threats, attacks and fraud. What counts as "big data" severals depending on the perfactability of the users and their tools, and expanding capabilities make big data a moving target.

Index Terms-Big Data, Security, Privacy, and Acquisition.

1. INTRODUCTION

Data is the collection of Excellency and variable and some related sense. In recent year the size of data is increases. Data are collected to create the information for suitable making decisions. Hence the data provide to reach resources for the knowledge discovery and it support to the decisions. Data base is organized collection of data so that it can be easily update, access, insert and the access to it. Big data is becoming the new area for scientific data research and for business applications. That will help the research to choose the best data mining tools for their frame work. The volume, variety and velocity of data coming into your organization continue to reach unprecedented levels. The Big Data occurrence has emerged as a result of enormous amounts of records that are becoming available across a wide range of application domains across science, business, and government. In now day's data sets grow rapidly in part because they are progressively more gathered by cheap and numerous information sensing mobile devices, aerial, software logs, cameras, microphones. investigation of data sets can find new correlation to "stain big business trends, avoid diseases, and combat crime and so on."

2. BIG DATA



2.1. *CHARACTERISTICS OF BIG DATA*

2.1.1. *Volume*:

Big Data describe the large volume of data. it has two basis on data there are structured and unstructured. It covers a business on day to day basis. But it does not the amount of data is important. It has stored the big amount of data.

2.1.2.*Variety*:

Verity of data means the different type of data refers. It has number of data is required. The verity of data forms the email, video, textdocument, structure and numeric data. It has financial transaction. It has include the different or number of verities

2.1.3.*Variability:*

It considers the consistency of data flow. It consist the fixed pattern of data, range, mean, variance and standard deviation is four commonly use the measure variabilities. Range means the amount between our smallest and largest in the sets. You can find the range by subscribing smaller to larger numbers. Variance of a data sets. It gives the rough idea and how spread out your data. Small number of variance means the data is tightly clustered together and large number means the large number of data is spread apart. Varience rearly useful. It has expected to calculate the deviation standerds. Standard deviation means the dispersion of set of data values.

2.1.4.*Velocity:*

Velocity refers the speed of processing data. For example Facebook has 300 billion images may same like lot of facebook people upload the more photos a day to day. Velocity is the measure of how fast the data incoming and then process it.

3. FIGURES, TABLES AND PHOTOGRAPHS

Acquisition/Recording: Acquisition means the all methods of simultaneous or sequential time measurement and counting for measurable or quantifiable data and group of contigeneous data.
Data Acquisition: It is the process of sampling signal That measure real world physical condition and converting the resulting samples into digital numeric values that can be manipulated by a computer. Digital Data Acquisition is a process of collecting signal samples to measure the physical phenomena and it is convert the information to a desktop computer program according to National Instruments. Data Acquisition method is used to access the voltage, sound pressure or electronic current by the sensor, temperature etc. Then another method is used to access the application specific hardware and then programmable software. Data Acquisition consisted of measurements collected and recorded by technicians. Is software work in verity of situations?
Data Extraction and cleaning: Most of the information or data collected will not be a format ready for analysis. For example consider the information collected for hospital system, consisting the hand written dictations from several physicians, sensor, and x-ray, image data. It is difficult to analys. Data cleaning is the part of analysis. It cleans the data then performs statically analysis on time. The data is on time clean and data cleaning means the major challenges in Big data analytics. Big Data has the potential to revaluation many businesses driven by the many opportunities. Big Data is Largest Technology. It involving to a point in which the more organization are prepared to model.
Data Integration, Aggregation and Presentation: Data can be many heterogeneous and may have different metadata. There are different data aggregation and representations stratergies. May be needed for different data analysis task. Data analysis is more considerable more challenging than simply identifying locating and understanding the data.
Query processing and Analysis Modeling: It is suitable for big data to deal with the noise, dynamic and unsaturated data, heterogeneous and the data characterized by complex relation. Method for querying and mining Big Data are fundamentally

different .Big data forms the large heterogeneous information network. A problem with current big Data

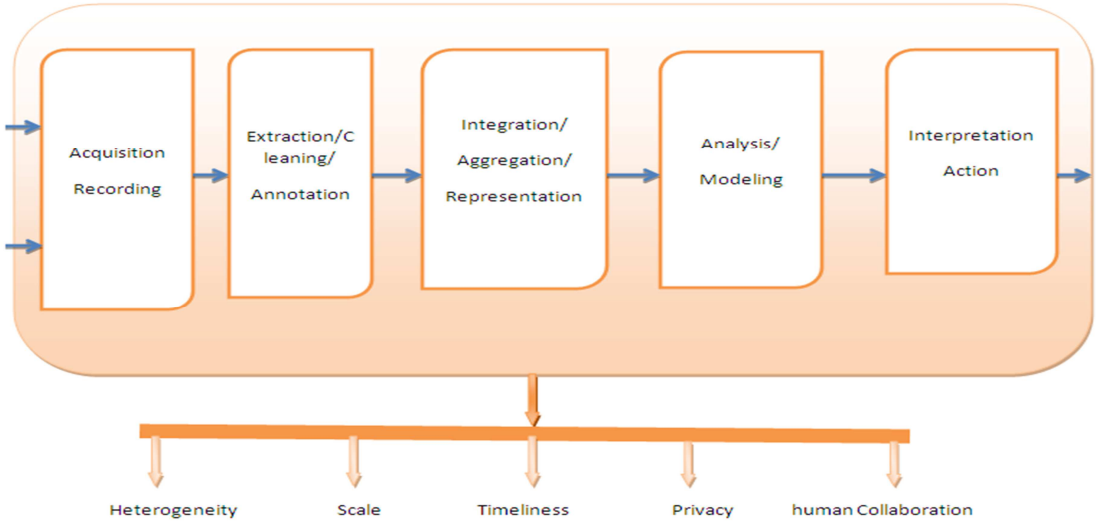


Fig. 1.Big Data Pipelining.

analysis is the lack of co-ordination between database system which hosts the data and provides SQL queries.

Interpretation: Interpretation is a communication process, designed to expose meanings and associations of our intellectual and natural legacy, through involvement with objects, artifacts, landscapes and sites.

4. BIG DATA PRIVACY AND SECURITY CHALLENGES

Security and privacy issues are magnified by velocity, volume and the variety of big data. Itlocation based information being collected by big data applications. Network attacks and Network intrusions usually generate the traffic of specific data patterns. It has used the encryption technic to avoid the traffic. Encryption technic is used to can't be access the data unauthorized users.

CONCLUSION

We have entered an era of Big Data. The paper describes the concept of big data along with three v/s Volume, Velocity, Variety, Variability of Big Data. Big Data is seldom the province of significant human effort and there will rarely be a particular person who could be said to be an author for copyright purposes. Big Data is to highlight: cleaning, capture, integration, virtualization of large volume of fast moving highly complex big data. In most of the people daily life part is social network and because data is also a form of Big Data.

ACKNOWLEDGMENTS
First and foremost, we would like to thank my author Ms.DUMBRET.M., Ms.POPALGHATP.S., Ms.DHAMDHERE R.V**.** for his guidance and support. We will perpetually remain satisfying for the constant support and guidance extended by guide, in making this paper. Through our many discussions and ideas. The invaluable discussions we had with her, the penetrating question, has all led to the development of this paper.

(A.1)

REFERENCES

[1] Jonathan Paul Olmsted "Scaling at Scale: Ideal Point Estimation with 'Big-Data" Princeton Institute for Computational Science and Engineering.

[2] Following the Breadcrumbs to Big Data Gold. Yuki Noguchi. *National Public Radio*, Nov. 29, 2011. http://www.npr.org/2011/11/29/142521910/the-digital-breadcrumbs-that-lead-to-big-data

[3] https://en.wikipedia.org/wiki/Big_data

[4]Big data: The next frontier for innovation, competition, and productivity. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. McKinsey Global Institute. May 2011.