CHALLENGES AND OPPORTUNITIES WITH BIG DATA

Ms. Pushpanjali Saini

Assistant Professor
St. Xavier's College, Jaipur

Abstract

In recent years Internet of Things, Cloud Computing and Big Data have become the hot topics in the Information Technology and Corporate world. Big data describes large volumes of data sets both structured and unstructured, as it acts as superset for data sets that are complex to compute the challenges of Big Data such as Data Analysis, Data Capture, Data Search, Data Sharing, Data Storage and Data Transfer have influenced many aspects of modern society such as retail, manufacturing, financial services, mobile services, life sciences, physical science and many more. All these aspects including Data Visualization, Information Privacy and opportunity of Big Data will be discussed in the paper. The Big Data refers to the huge data sets that may be examined computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions. Big data is basically used for predictive analytics to extract value from the data which leads to more accurate and effective decision making.

Keywords: Big Data, data information technology industry: Analytics: Data Sets.

Introduction

Big data is characterized by the three Vs: volume, velocity, and variety. The data in these days are in the volume of terabytes to petabytes. The rate of data is increase by multiple by petabytes every year thus Velocity is concern with speed of data by which data is collect. Changing data usually lags accumulating data by several orders of magnitude. The variety of data is repetitively growing; it may be both structured and unstructured. There are several organizations in with various rivals but companies are facing problems with uncertainty in managing the data. Resource management, Business, technology, health, government data are the fields where big data samples are accessible.

The "3V's", how Doug Laney calls them in his article 3-D Data Management: Controlling Data Volume, Velocity and Variety, published in 2001, shows main features that are considered to be significant.

Characteristic	Definition	Solution
Volume	The bulk of the data to be processed can be infinite and the speed of processing is constant. Moreover, if data is compressed then firstly data is to be decompress after this huge volume of data analyse which is a tiresome job.	Data Mining is method for managing large data i.e. OLAP(Online Analytical Processing)
Velocity	It is deal with the speed by which data is transfer from one socket to another, therefore streaming of data in real time or close to real time is big task	The solution to this is shrink the data but it leads to problem in the operation volume handling, and then the only solution to this is to invest in infrastructure.
Variety	Big data comes from various sources in different forms; data is both structured and unstructured. Data can be in the form of text, log file, audio and video.	Numerous assertion approaches be present for this resolution, and the main methods are to fill maximum frequently detected values or to build learning models to forecast possible values for each data field, based on the observed values of a given case.

Further Optimization of big data is global search methods is difficult because of large-scale systems. Optimizations of complex systems are essential to cope with a variety of challenges. Meta-heuristic global search methods such as evolutionary algorithms have been successfully applied to optimization [2]. It is also become difficult for firms to search the correct ability to handle new technologies and inferring the data to get significant business acumens. Moreover, data accessibility and connectivity can be a hindrance due to lack of points that are not yet connected and companies often do not have the proper platforms to aggregate and to cope with the data through the enterprise. To solve this enormous mounting bulk of data shaped as a portion of

power grid operation and to focus on clarifications and facilities for system integration and data management, Siemens and Accenture ACN -0.97% recently formed a joint venture in the smart grid field. These aids will allow utilities to integrate functioning technologies, such as real-time grid management, with information technologies like smart metering [3]. So there are various challenges in big data for instance, data, process, management and security issues

Big Data: Challenging Issues

Big data is deal challenges related to infrastructure security for secure computations in distributed programming frameworks and security put into practise for Non-Relational data stores, Privacy of data by follow cryptographic security, Granular access and by preserving privacy for analysis. In addition, data management, integrity and reactive security are also challenges that are to be handling in the process to get valuable data.

1. Data Challenges

1.1 Volume

The volume of data, especially machine-generated data, is exploding. Social media plays a key role: Twitter generates more than 7 terabytes (TB) and Facebook Generate 10 TB of data every day. The challenge is how to deal with the size of Big Data.

1.2 Variety, Combining Multiple Data Sets

Approximately 80% of information now is formless and it is characteristically too immense to manage efficiently. David Gorbet explains [5]:It used to be the case that all the data an organization needed to run its operations effectively was structured data that was generated within the organization. Things like customer transaction data, ERP data, etc. In these days, firms are considering to influence a lot more data from a broader diversity of sources. Complex data of different type is tough to represent in a relational model, so data from types, sources, and format is a challenging task to make structured.

1.3 Velocity

"As businesses get more value out of analytics, it creates a success problem— they want the data available faster, or in other words, want real-time analytics explain by Shilpa Lawande of Vertica[6]

1.4 Veracity, Data Quality, Data Availability

It is also true that several times inconsistent data is existed there due to which quality of data is lower. For instance, when data become unavailable then to get useful information from this enormous data becomes very difficult. This is a huge challenge: how to find high-quality data from the vast collections of data that are out

there on the Web.

1.5 Privacy challenges

There are some key issues are encountered when there is need to make system, that can compute and store data in secure manner and to confirm that private data should remain private as it pass through various phases of input, output and analysis.

Security and privacy issues are magnified by the velocity, volume, and variety of Big Data, such as large-scale cloud infrastructures, diversity of data sources and formats, streaming nature of data acquisition and high volume inter-cloud migration. Therefore, traditional security mechanisms, which are tailored to securing small-scale, static (as opposed to streaming) data, are inadequate. In this talk we highlight the top ten Big Data security and privacy challenges. Highlighting the challenges will motivate increased focus on fortifying Big Data infrastructures.

2. Process Challenges

Processing of data is done in various steps which includes capturing data, arranging data from diverse sources, converting the data into a type that is suitable for analysis, modelling it, Understanding the outcome, visualizing and distribution the results and integration of data represent in fig(1).
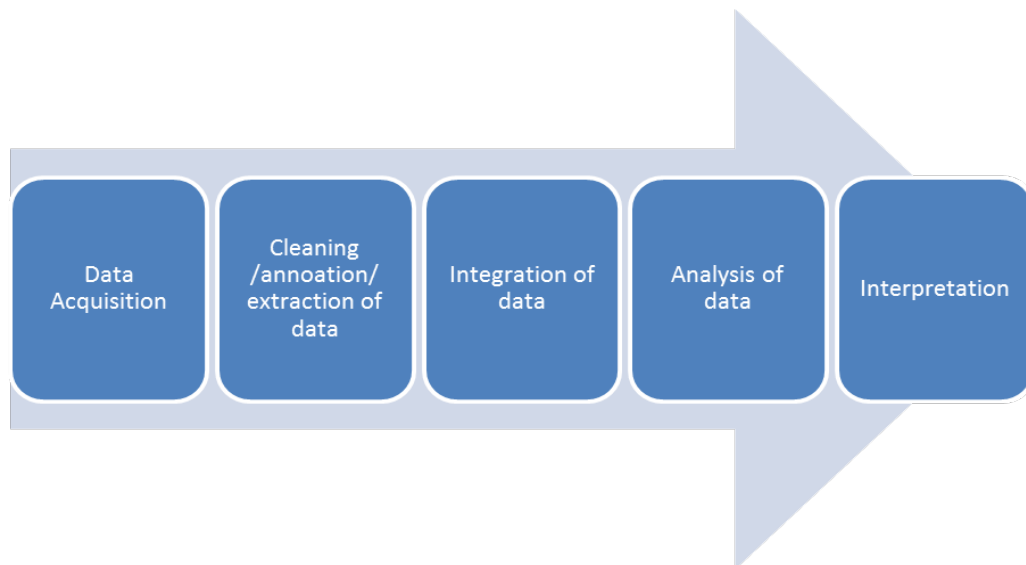
Fig(1)

2.1 Data acquisition challenges

Data that is not useful can be filtered and compressed by orders of degree. First challenge is to describe these filters like this that they do not remove valuable information. Secondly it is a challenge to repeatedly generate

the right metadata to define the type of data is recorded the way it is recorded as well as measured

2.2 Challenges in data cleaning

Reality of big data is away from this believe that Big Data always telling us truth. For example, an educational institute results of students are in different form, some students leave institution in the mid of session, name of student can be entered wrongly, so data cleaning can done only on valid data.

2.3 Data Integration, Aggregation, and Representation:

Heterogeneity of the overflow of data, it is not plenty simply to record it and throw it into a repository. Data analysis is significantly more perplexing than simply discovering, categorizing, understanding, and citing data.

2.4 Data Modeling and Analysis Challenges:

Procedures for mining Big Data are basically changed from out-dated statistical analysis on minor samples. Big Data is often noisy, dynamic, heterogeneous, inter-related and unreliable. However, even noisy Big Data could be additional valued than small samples due to reason that overall statistics gained from regular patterns and association analysis generally conquer individual oscillations and frequently reveal added consistent concealed patterns and knowledge. A difficulty with present Big Data analysis is the absence of coordination between database systems. This is a hindrance to booming over the cooperating elegance of the first generation of SQL-driven OLAP systems.

2.5 Interpretation

It involves usually, the norms, assumptions made and reviewing the analysis. Additionally, there are varied possible causes of error: computer systems can have virus, results can be based on erroneous data. There are often vital assumptions behind the data recorded. In short, it is hardly enough to provide just the results.

Instead of provide added information that clarifies procedure by which outcome was derived, and depend upon what inputs one get is basic information. Moreover, users not only want to see result but also want to understand why they get those results.

Management Challenges

"Many data warehouses contain sensitive data such as personal data. There are legal and ethical concerns with accessing such data. So the data must be protected and access controlled as well as logged foraudits." (Michael Blaha)

The main management challenges are

3.1 Correct usage of data by tracking and managing life cycle of data.

3.2 Platforms Technology for Big Data

Currently industry is still in an undeveloped state and facing eruption of diverse high-tech solutions. Many of the technologies require important technical skills.

Big Data Opportunities

Big data is related for all parts of our society. Industry for make their business intelligence using Big Data in the form of reporting and decision support to forecast. This usage of big data give emphasis to that big data is acute for obtaining actionable information. Also, Regimes are also attracted towards the use of big data and analysis of future to rectify decision making and clearness, to engross citizens in communal affairs, to mend domestic security of country. Healthcare represents another major area to which big data may offer novel opportunities [7]. Knowledge well being systems are currently focusing on turning health care data into knowledge, translating that knowledge into practice, and creating new data by means of advanced information technology as pointed out in [8], the use of big data technologies can lessen the cost of healthcare whereas refining its quality by creating care more protective and personalized and creating it on more home-based uninterrupted observing. Big data is also important for research, various fields of science and engineering are presently encountering from many fold increase in the bulk of data generated as contrast to only ten years before. The obtainability of big data is transforming the way by which research is accompanied and is foremost to the appearance of a innovative standard of science based on data-intensive computing, it poses a important encounter for scientists. To influence these large volumes of data, innovative methods and skills are needed. A new type of organisation, the Research Data Organization, must be planned, applied and enhanced to upkeep the life cycle of scientific data, its movement across scientific disciplines.

Conclusion:

For the practical clarification for huge and multifaceted data is a challenge that firms in these areas are constantly learning and applying new ways to handle it. The leading difficulties concerning Big Data is the infrastructure's more costs. Hardware equipment is very costly for most of the enterprises, even if Cloud solutions are available. Big data system involves enormous processing power and steady and composite network arrangements that are prepared by specialists. In addition to this, software solutions incline to have great prices if the receiver doesn't choose for open source software and if they selected open source, to configure there is requirement specialists with skills to work on it. The negative of open source is that maintenance is not providing as it is in paid software. So, all that is essential to keep a Big Data solution working properly needs. In the case of hardware can only be as fast as present-day technologies can give. Sorting data with human capabilities and set of tools speed up the process. There is need to choose proper plan, tools to enhance opportunity of data and to face challenges.

WORKS CITED

1.   Priya P. Sharma, Chandrakant P. Navdeti, (2014), "Securing Big Data Hadoop: A Review of Security, Issues, Threats and Solution", IJCSIT, 5(2), pp2126-2131

2.   T. Chai, Y. Jin, and S. Bernhard, "Evolutionary complex engineering optimization: Opportunities and challenges," IEEE Computational Intelligence Magazine, vol. 8, no. 3, pp. 12–15, 2013.

3.   Eric Spiegel (@ericspiegel) is the president and CEO of Siemens U.S.A. and the author of the 2009 book "Energy Shift: Game-changing Options for Fueling the Future."

4.   Managing Big Data. An interview with David Gorbet ODBMS Industry Watch, July 2, 2012. http://www.odbms.org/blog/2012/07/managing-big-data-aninterview-with-david-gorbet/

5.   On Big Data: Interview with Shilpa Lawande, VP of Engineering at Vertica. ODBMs Industry Watch, November 16, 2011.

6.   T. Murdoch, A. Detsky, "The Inevitable Application of Big Data to Health Care", JAMA, 2013, 309(13):1351-1352.

7.   H.V. Jagadish et al., "Challenges and Opportunities with Big Data", 2012, available at http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf (Downladed on April 30, 2012).

8.   Big Data Opportunities and Challenges: Discussions from Data Analytics Perspectives Zhi-Hua Zhou, Nitesh V. Chawla, Yaochu Jin, and Graham J. Williams

9.   Alexandru Adrian TOLE Romanian – American University, Bucharest, Romania Database Systems Journal vol. IV, no. 3/2013 31: Big Data Challenges

10. Monica Bulger, Greg Taylor, Ralph Schroeder Oxford Internet Institute, September 2014

11. Data-Driven Business Models: Challenges and Opportunities of Big Data

12. Alexandros Labrinidis University of Pittsburgh Pittsburgh, PA, USA, H. V. Jagadish University of Michigan, Ann Arbor, MI, USA, Challenges and Opportunities with Big Data, August 27th 31st 2012, Istanbul, Turkey, Proceedings of the VLDB Endowment, Vol. 5, No. 12

13. Musa Jafar, Jeffry S. Babb, Kareem Dana, Computer Information and Decision Management West Texas A&M University Canyon, TX 79106 USA2013 Proceedings of the Conference for Information Systems Applied Research ISSN: 2167-1508 San Antonio, Texas, USA v6 n2821, Decision-Making via Visual Analysis using the Natural Language Toolkit and R

14. Hacène Cherfi * Amedeo Napoli * Yannick Toussaint, Towards a Text Mining Methodology Using Frequent Itemsets and Association Rule Extraction