Extracting Market Value of Business and Business Decision from Big Data Analytics

Sarswati Kumar Pandey
Software Developer
Tarantula Labs Pvt. Ltd.
1947, Infocity
Gandhinagar, Gujarat

ABSTRACT

Organizational data are exponentially increasing day by day that causes the data explosion, computation capacity and storage management of data in organizations due to variety, volume and velocity of data exceeded. The impact of Data Explosion on organizations computing technology and architecture are reducing the timely and accurate decision making. Big Data is not only the solution that assists the organizational success but also to increase the market more effectively and business operational improvements. Thus big data technologies help an organization to effectively take decisions and increase the market value.

General Terms

Big Data, Big Data Analytics.

Keywords

Data explosion, Computational capacity, Architecture, Big Data, Decision making.

1. INTRODUCTION

The explosion of data in the organizations is now a big concern. The computing architecture used by the organization now becomes overhead to the currently used tools and technologies. There is a problem to extract the business values from different sources due to explosion of data. More than 2.5 petabytes of data is being handled by millions of customers each hour in Wal-Mart. Facebook is transacting 250 millions of images, photos upload and communication among 800 million users that are actively taking part and apart of this all more than 900 million of different types of objects are being transfer in the storage media. There is a daily cause of data explosion that is nothing but what we texting, calling, browsing and tweeting on mobile devices worldwide [1]. The organizations are flooded with petabytes of data from every transactional and operational system.

According to International Data Corporation (IDC) the amount of information as it is exponentially increasing will be imperative as many bits of information in the digital world as stars in the physical world. The data explosion is not a new problem. It started many years ago as in the 1970s. The main alteration is affecting the growth velocity and the diversity in the data. The main objective of big data is to harness and harvest every relevant data bytes and using it for the effective decision making. Big data technology helps the organization to collect the large amount of relevant data and helps to understand the advantage of full value of the relevant data. Thus the threat to traditional computing architecture by data explosion now has an effective solution and a way to deal with the value of relevant data [5].

2. DATA MANAGEMENT RETHINKING

The organizations require a better way to manage the data because the shifting in amount and potential of data are so much epic now a day. The large organizations must think about managing their data and making the important decisions on the data that are useful and relevant in their computing architecture and business operations. Most organizations are showing their progress slowly to extract the relevant and valued data from the big data. Some companies are attempting to use traditional data management practices on big data only to understand that the old rules will no longer be applicable [1]. Organizations are now using integration of data, metadata management, data quality and governance altogether instead of viewing data integration as a standalone techniques.

The traditional data approach extract-transform-load has been integrated with one that reduces the movement of data and enhances processing power. The most of the organizations treat the data as a core and valuable asset and thus to manage the information the enterprises are using the predictive and proactive approach.

3. A GLIMPSE OF BIG DATA

Big data defines a situation where the velocity, variety and volume of data exceed the storage or computing capacity for timely and accurate decision making of an organization. Transactional data stores hold the various parts of data that are being daily used in the organizations [6]. Intra machine interaction like details of call records, metering and Radio Frequency Infrared detector systems produces a large amount of data waves. Every form of data is exponentially expanding and the data is also coupled with semi- structured and unstructured data from various sources as social media. By using some standard estimation many organizations in different sectors are having more than petabytes of data. It is very big problem that this number is going to be doubled in every six months. Fig.3 shows that Big data can be defined by volume but lesser in context that is moving constantly by target than by its exponentially increasing variability, velocity, volume and complexity [1].

3.1 Variability

The flow of data from its source to destination is highly variable and it is challenging to manage.

3.2 Velocity

The real time data like RFID tags and metering are producing a greater need to deal with tolerance of data. The velocity explosion is pressuring on organizations to build the essential infrastructure to act quickly enough.

3.3 Volume

The non-numeric and unstructured data in the organizations are up to 85 percent. The data is still folded in decision making and qualitative analysis.

3.4 Complexity

The expanding nature of data sources in the universe is associated with the linking, matching and transforming of the data across the business systems and entities. Organizations must understand the relationships such as linkage of data and complex hierarchies, among all the data.
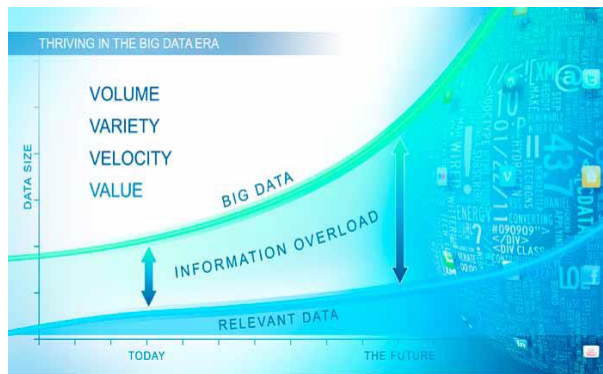

Fig 3: Finding relevant data is key to delivering the value from huge amounts of data

It is very important to know that all the available data in the universe are not useful or relevant. The responsibility of an organization must be separating the vague, unstructured and semi-structured data from the relevant data. It will decrease the information overload in the computing architecture.

4. EXTRACTING BUSINESS VALUE USING THREE KEY TECHNOLOGIES FROM BIG DATA

The new generation architectures and technologies as big data technology are designed to extract the relevant data from a large volume and variety of data. The technology enables the high velocity capture, analysis and discovery. This analysis and discovery is required in real time and it must be secure, achievable and affordable. Benefit of big data analytics is possible for many advance technologies. The storage capacity has become so cheap and abundant that there is no problem to store large volumes of data in technological processes. Many computing models such as clustering, parallel processing, cloud computing, grid computing and virtualization are combined with high speed connectivity with possible redefinition [1].

There are following three technologies that will help the organizations to extract the relevant business information from the big data.

4.1 Information management for big data

Managing the data as a valuable asset in the organization is very important. Organizations are coping with the management of existing data. Big data is adding more complexity. It is very necessary to know what data must be stored and till how long time it must reside in the storage volumes. Big data will increase the need of quality of data and governance that embed the analytics into operational systems.

4.2 High performance analytics for big data

Analyzing all the available data to extract the relevant and precise information has been possible by high performance analytics.

4.3 Flexible deployment options for big data

This deployment model brings the choice of deploying the application on hybrid technologies platforms.

Big Data is new generation architecture and technology that is designed and developed to extract the relevant and valuable data from a large amount and variety of data much more economically. It is possible when high velocity capturing has been enabled to analyze the relevant data in traditional architectures. Big data is used to create high performance and flexible computing infrastructure. Depending upon the technical requirement of the organization big data can be managed in many ways such as complete and targeted data scenarios. Big data help an organization to be successful, enhance effective performance and to increase the market value efficiently. Big data help to take the business decisions more efficiently and effectively [1].

5. CHALLENGES AND OPPORTUNITIES WITH BIG DATA

The decision making which is data driven is being recognized broadly and the idea of big data is growing very fast. Some challenges such as heterogeneity, scale, timeliness, complexity and privacy problems are slowing down the big data analysis pipeline phases that can produce the relevant and valuable data. The first problem is acquisition of data, when any natural disaster takes place it is very mandatory to make decision that what data is relevant and what data is not, what data is to be discarded and what to be kept. Today most of the data are not in well-structured format. The problem is to keep relevant data efficiently with their exact metadata [2].

5.1 Processing Pipeline Phases

There are the various challenging phases of processing pipeline in the big data. The processing pipeline phases are shown in fig 5.1

5.1.1 Data Acquisition and Recording

Big data is generated and recorded from a large volume of data sources. Simulations and scientific experiments are generating petabytes of data now a day. These all generated data contains most of the data that is of no any interest. These data can be easily compressed by their order of magnitude and can be easily filtered. The challenge is to develop these data filters in such a way that they do not throw away the essential and useful data. The second challenge is to exactly generate the metadata that may able to define that what and how data is measured and recorded. The metadata must be recorded along with the observational data [8].

Human burden to record the metadata will be minimized by the metadata acquisition systems. The recording information will be useful only when it can be interpreted and carried along with the data analysis pipeline [2].

5.1.2 Information Extraction and Cleaning

The frequently collected information will not be in the format that is needed for data analysis. Information extraction tools are required for extracting the relevant data and information from the various sources and it must be in structured format

for data analysis. It is a big challenge to do this completely and correctly [9]. These data includes images and will include audio and video contents in the future. Such type of data extraction is very much application dependent. The data cleaning process is to recognize the constraints very well and to well-understood the data [2].
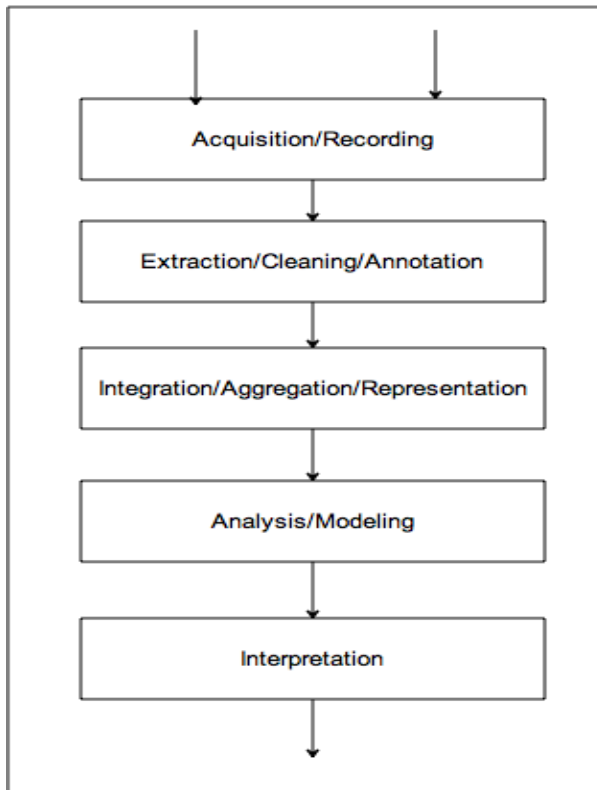

Fig 5.1: Processing pipeline phases

5.1.3 Data Integration, aggregation, and Representation

Analyzing the data is more challenging than identifying, locating, citing and understanding data. Data integration is the process of integrating a large amount of useful and relevant data from different data sources that provide some of the answers while business decision making [11]. The same information can be stored in the database using the different modeling techniques. Designing the database is an art now a day. The database design is effectively created and can be used in the absence of intelligent database design with the help of professionals and domain scientists [2].

5.1.4 Query Processing, Data Modeling and Analysis

Query processing in big data is different from traditional approaches that were done in the small samples of data. Big data includes noise, heterogeneity, and dynamism, inter-relation and untrustworthy. Large heterogeneous information networks are formed by interconnected big data. Information redundancy can be explored to overcome for the missed data and to cross check the conflict, to disclose inherent clusters, to uncover hidden models and relationship and to validate trustworthy [2]. Big data enables the next generation interactive data analysis with real time responses. In future the query related to big data will be automatically generated to create the contents on the websites, to populate the recommendations and hot lists, to provide data set values based on ad hoc analysis whether to it or discard it [3].

5.1.5 Interpretation

If users can not understand the analysis it is of limited value to analyze the big data. A decision maker has an ability to provide analysis result with the interpretation of the result. This involves retracing the analysis and examining all the assumptions made [7]. There may be many error sources like bugs in the computer system, model assumptions and result based on erroneous data. This is a challenge for big data with its complexity. Data records are often having so many assumptions. It is very difficult to provide just only results. There must be some supporting information to support how the result is derived with the precise inputs. This type of supplementary information is known as provenance of the data [2]. User can view each piece of data in a few clicks and can understand the provenance which is the main features to understand the data.

5.2 Big Data Analysis Challenges

There are some of the common challenges that lie sometimes all of the pipeline analysis phases. The fig. 5.2 shows the underlying various challenges

5.2.1 Heterogeneity and Incompleteness

A great deal of heterogeneity is tolerated when humans consume information. The natural language can provide the information in depth. The heterogeneity of data is creating problem in the data integration processing phase. After error correction and data cleaning some errors and incomplete information remain in the processing phase.

5.2.2 Scale

The main issue of Big Data is its size. Managing large, complex and ever growing amount of data is being a continuous challenging issue for many decades. In the past by increasing the processor speed this problem was mitigated till some extent.
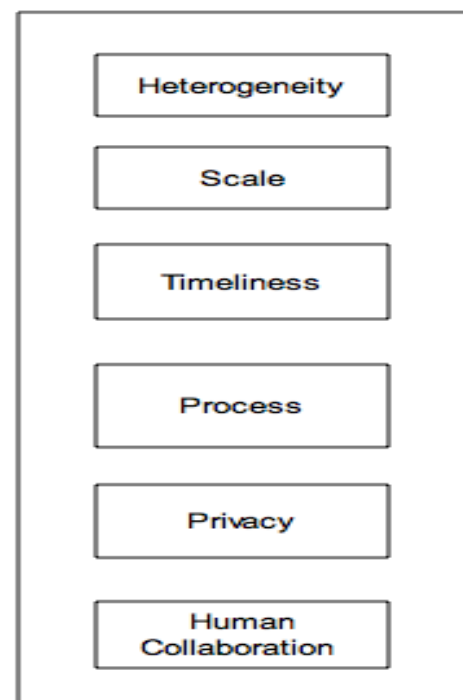

Fig 5.2: Processing pipeline phases

5.2.3 Timeliness

If the data set is too large it will take longer time to be analyzed. We have to design a system in such a way that

processes the large amount of data faster. In many situations the data analysis results are needed immediately. For example a credit card fraudulent, the data analysis must be accomplished before the transaction is completed. Such type of system structure design becomes challenging that response immediately for large volume of data [10].

5.2.4 Privacy

The data privacy is a concern and in case of big data the privacy is even more important concern. Managing the privacy of personal data, health records particularly through linking of data from different sources is a big privacy issues. There is a big public fear about the inappropriate use of personal data. Location of users can be traced very easily through many static connection points. Now a day hiding location of a user is much challenging than the identity of a user. Many online applications share private data, how data sharing is linked and how to provide control over sharing to the users. The privacy of big data analysis is a big concern [4].

5.2.5 Human Collaboration

There are many advance technology that are being used in computational analysis such as patterns like CAPTCHAS that a human can easily identify than any algorithms. Big data analysis will not be completely computational; rather it will be designed explicitly to have a human in loop [2].

6. CONCLUSION AND FUTURE WORKS

Technologies have entered an era of Big Data. Through better data analysis of the large volumes that are becoming available, there is the potential for building faster advancement in many scientific disciplines and improving the profitability and success of many organizations. However, many technical challenges described in this paper must be addressed before this potential can be realized fully [2]. The challenges include not just the obvious issues of scale, but also heterogeneity, lack of structure, error-handling, privacy, timeliness, provenance, and visualization, at every stages of the analysis pipeline from data acquisition to result interpretation. These technical challenges are common across a large variety of application domains, and therefore not cost-effective to address in the context of one domain alone. Furthermore, these challenges will require transformative solutions, and will not be addressed naturally by the next generation of industrial products. We must support and encourage fundamental research towards addressing these

technical challenges if we are to achieve the promised benefits of Big Data. Big Data Analytics are used to extract the relevant information from a large amount of data that improve the market value of business and helps to take effective business decisions.

7. REFERENCES

[1] Badyal, Sargam, and Naveen Kumar. *"Suggesting Advance Methods for Obtaining Instantly Recognizable and Accurate Images."* (2014).

[2] Labrinidis, Alexandros, and H. V. Jagadish. "Challenges and opportunities with big data." *Proceedings of the VLDB Endowment* 5.12 (2012): 2032-2033.

[3] Russom, Philip. "Big data analytics." *TDWI Best Practices Report, Fourth Quarter* (2011).

[4] Zikopoulos, Paul, and Chris Eaton. *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, 2011.

[5] Chen, Hsinchun, Roger HL Chiang, and Veda C. Storey. "Business Intelligence and Analytics: From Big Data to Big Impact." *MIS quarterly* 36.4 (2012): 1165-1188.

[6] Beath, Cynthia, et al. "Finding value in the information explosion." *MIT Sloan Management Review* 53.4 (2012): 18.

[7] Barney, Jay B. "How a firm's capabilities affect boundary decisions." *Sloan Manage* (2012).

[8] Chaudhuri, Surajit, Umeshwar Dayal, and Vivek Narasayya. "An overview of business intelligence technology." *Communications of the ACM* 54.8 (2011): 88-98.

[9] Bharadwaj, Anandhi, et al. "Digital business strategy: toward a next generation of insights." *MIS Quarterly* 37.2 (2013): 471-482.

[10] Demirkan, Haluk, and Dursun Delen. "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud."*Decision Support Systems* 55.1 (2013): 412-421.

[11] Sarswati Kumar Pandey and Madhulika. "Dealing with Heterogeneity of Data and Knowledge using Object Modeling." International Journal of Advances in Engineering & Technology (IJAET), Volume 7 Issue 6, pp. 1812-1817, Jan. 2015.