

Quality Management and Improvement in Big Data Analytics

Dr. Rekha Chouhan

(Assistant Professor, Pratibha Institute of Business Management, Pune, INDIA)

rekha09@gmail.com

Abstract

In the information era, enormous amounts of data have become available on hand to decision makers. Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Due to the rapid growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets. Furthermore, decision makers need to be able to gain valuable insights from such varied and rapidly changing data, ranging from daily transactions to customer interactions and social network data. Such value can be provided using big data analytics, which is the application of advanced analytics techniques on big data. This paper aims to analyze some of the different analytics methods and tools which can be applied to big data, as well as the opportunities provided by the application of big data analytics in various decision domains so to improve the quality.

Keywords: *big data, data mining, analytics, decision making.*

Introduction

Big Data Analytics

The term “Big Data” has recently been applied to datasets that grow so large that they become awkward to work with using traditional database management systems. They are data sets whose size is beyond the ability of commonly used software tools and storage systems to capture, store, manage, as well as process the data within a tolerable elapsed time.

Big data sizes are constantly increasing, currently ranging from a few dozen terabytes (TB) to many petabytes (PB) of data in a single data set. Consequently, some of the difficulties related to big data include capture, storage, search, sharing, analytics, and visualizing. Today, enterprises are exploring large volumes of highly detailed data so as to discover facts they didn't know before [1].

Hence, big data analytics is where advanced analytic techniques are applied on big data sets. Analytics based on large data samples reveals and leverages business change. However, the larger the set of data, the more difficult it becomes to manage. Naturally, business benefit can commonly be derived from analyzing larger and more complex data sets that require real time or near-real time capabilities; however, this leads to a need for new data architectures, analytical methods, and tools. A huge repository of terabytes of data is generated each day

from modern information systems and digital technologies such as Internet of Things and cloud computing. Analysis of these massive data requires a lot of efforts at multiple levels to extract knowledge for decision making. Therefore, big data analysis is a current area of research and development. The basic objective of this paper is to explore the potential impact of big data challenges, open research issues, and various tools associated with it. As a result, this article provides a platform to explore big data at numerous stages. Additionally, it opens a new horizon for researchers to develop the solution, based on the challenges and open research issues [4].

The integration of quality, analytics, and big data guided by the strategic direction of the organization can potentially result in new sources of customer value and a new source of competitive advantage. This research report describes some of the advances related to quality, analytics, and big data and explores the benefits of integrating them in practice. Analytics and big data techniques can help improve product and service quality by generating new customer insights and enhancing decision making. Also, it is argued that analytics and big data initiatives can be improved if quality principles are applied. It is explained that analytics is not just about numbers and big data doesn't necessarily mean we need millions of data points. A new emergent definition of big data is introduced plus ten ideas for better integrated quality, analytics, and big data practices.

Therefore the successive section will elaborate the big data analytics tools and methods, in particular, starting with the big data storage and management, then moving on to the big data analytic processing. It then concludes with some of the various big data analyses which have grown in usage with big data.

Characteristics of Big Data

Big data is data whose scale, distribution, diversity, and/or timeliness require the use of new technical architectures, analytics, and tools in order to enable insights that unlock new sources of business value. Three main features characterize big data: volume, variety, and velocity, or the three V's. The volume of the data is its size, and how enormous it is. Velocity refers to the rate with which data is changing, or how often it is created. Finally, variety includes the different formats and types of data, as well as the different kinds of uses and ways of analyzing the data [2].

Data volume is the primary attribute of big data. Big data can be quantified by size in TBs or PBs, as well as even the number of records, transactions, tables, or files.

Additionally, one of the things that make big data really big is that it's coming from a greater variety of sources than ever before, including logs, clickstreams, and social media. Using these sources for analytics means that common structured data is now joined by unstructured data, such as text and human language, and semi-structured data, such as eXtensible Markup Language (XML) or Rich Site Summary (RSS) feeds. There's also data, which is hard to categorize since it comes from audio, video, and other devices. Furthermore, multi-dimensional data can be drawn from a data warehouse to add historic context to big data. Thus, with big data, variety is just as big as volume[5].

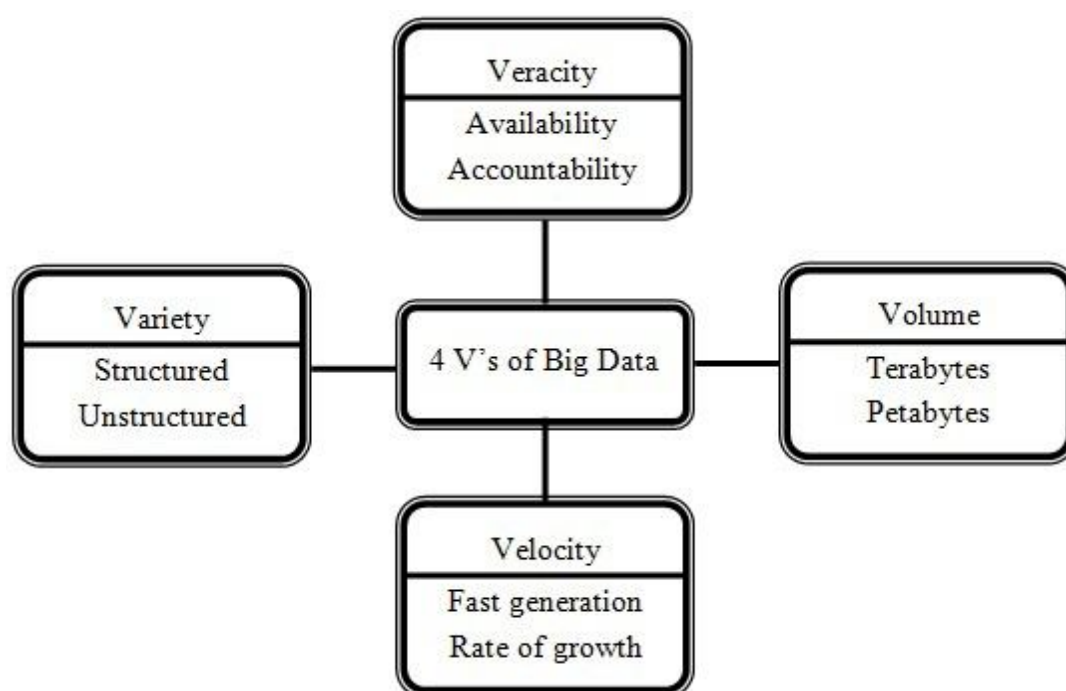


Fig.1: Characteristics of Big Data

Moreover, big data can be described by its velocity or speed (Fig.1.). This is basically the frequency of data generation or the frequency of data delivery. The leading edge of big data is streaming data, which is collected in real-time from the websites.

Some researchers and organizations have discussed the addition of a fourth V, or veracity. Veracity focuses on the quality of the data. This characterizes big data quality as good, bad, or undefined due to data inconsistency, incompleteness, ambiguity, latency, deception, and approximations.

Big Data Analytics Tools and Methods

With the evolution of technology and the increased multitudes of data flowing in and out of organizations daily, there has become a need for faster and more efficient ways of analyzing such data. Having piles of data on hand is no longer enough to make efficient decisions at the right time.

Such data sets can no longer be easily analyzed with traditional data management and analysis techniques and infrastructures. Therefore, there arises a need for new tools and methods specialized for big data analytics, as well as the required architectures for storing and managing such data. Accordingly, the emergence of big data has an effect on everything from the data itself and its collection, to the processing, to the final extracted decisions.

Consequently, proposed the Big – Data, Analytics, and Decisions (B-DAD) framework which incorporates the big data analytics tools and methods into the decision making process. The framework maps the different big data storage, management, and processing tools, analytics tools and methods, and visualization and evaluation tools to the different phases of the decision making process. Hence, the changes associated with big data analytics are reflected in three main areas: big data storage and architecture, data and analytics processing, and, finally, the big data analyses

which can be applied for knowledge discovery and informed decision making [3].

Each area will be further discussed in this section. However, since big data is still evolving as an important field of research, and new findings and tools are constantly developing, this section is not exhaustive of all the possibilities, and focuses on providing a general idea, rather than a list of all potential opportunities and technologies.

Especially for the manufacturing, energy and utilities, and telecommunications industries, big data can be used for quality management, in order to increase profitability and reduce costs by improving the quality of goods and services provided [7]. For example, in the manufacturing process, predictive analytics on big data can be used to minimize the performance variability, as well as prevent quality issues by providing early warning alerts. This can reduce scrap rates, and decrease the time to market, since identifying any disruptions to the production process before they occur can save significant expenditures. Additionally, big data analytics can result in manufacturing lead improvements. Furthermore, real-time data analyses and monitoring of machine logs can enable managers to make swifter decisions for quality management.

Also, big data analytics can allow for the real-time monitoring of network demand, in addition to the forecasting of bandwidth in response to customer behavior. Moreover, healthcare IT systems can improve the efficiency and quality of care, by communicating and integrating patient data across different departments and institutions, while retaining privacy controls [4]. Analyzing electronic health records can improve the continuity of care for individuals, as well as creating a massive dataset through which treatments and outcomes can be predicted and compared. Therefore, with the increasing use of electronic health records, along with the advancements in

analytics tools, there arises an opportunity to mine the available de-identified patient information for assessing the quality of healthcare, as well as managing diseases and health services [2].

Additionally, the quality of citizens' lives can be improved through the utilization of big data. For healthcare, sensors can be used in hospitals and homes to provide the continuous monitoring of patients, and perform real-time analyses on the patient data streaming in. This can be used to alert individuals and their health care providers if any health anomalies are detected in the analysis, requiring the patient to seek medical help [8]. Patients can also be monitored remotely to analyze their adherence to their prescriptions, and improve drug and treatment options.

Moreover, by analyzing information from distributed sensors on handheld devices, roads, and vehicles, which provide real-time traffic information, transportation can be transformed and improved. Traffic jams can be predicted and prevented, and drivers can operate more safely and with less disruption to the traffic flow. Such a new type of traffic ecosystem, with "intelligent" connected cars, can potentially renovate transportation and how roadways are used. Accordingly, big data applications can provide smart routing, according to real-time traffic information based on personal location data. Furthermore, such applications can automatically call for help when trouble is detected by the sensors, and inform users about accidents, scheduled roadwork, and congested areas in real-time [6].

Furthermore, big data can be used for better understanding changes in the location, frequency, and intensity of weather and climate. This can benefit citizens and businesses that rely upon weather, such as farmers, as well as tourism and transportation companies. Also, with new sensors and analysis techniques for developing long term climate models and nearer weather

forecasts, weather related natural disasters can be predicted, and preventive or adaptive measures can be taken beforehand.

Conclusion

In this research, we have examined the innovative topic of big data, which has recently gained lots of interest due to its perceived unprecedented opportunities and benefits. In the information era we are currently living in, voluminous varieties of high velocity data are being produced daily, and within them lay intrinsic details and patterns of hidden knowledge which should be extracted and utilized. Hence, big data analytics can be applied to leverage business change and enhance decision making, by applying advanced analytic techniques on big data, and revealing hidden insights and valuable knowledge.

Accordingly, the literature was reviewed in order to provide an analysis of the big data analytics concepts which are being researched, as well as their importance to decision making. Consequently, big data was discussed, as well as its characteristics and importance. Moreover, some of the big data analytics tools and methods in particular were examined. Thus, big data storage and management, as well as big data analytics processing were detailed. In addition, some of the different advanced data analytics techniques were further discussed. The integration of quality, analytics, and big data guided by the strategic direction of the organization can potentially result in new sources of customer value and a new source of competitive advantage. This research report describes some of the advances related to quality, analytics, and big data and explores the benefits of integrating them in practice. Analytics and big data techniques can help improve product and service quality by generating new customer insights and enhancing decision making. Also, it is argued that analytics and big data initiatives can be improved if quality principles are applied. It is explained that analytics is not just about numbers and big data doesn't necessarily mean we need millions of data points.

References

- [1] Liedtke, C. A. (2015). Quality, Analytics, and the Big Data Revolution. Presentation at the Global Quality Futures Workshop, Winnipeg, Canada, August 26, 2015.
- [2] C. L. Philip, Q. Chen and C. Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on big data, *Information Sciences*, 275 (2014), pp.314-347.
- [3] T. K. Das and P. M. Kumar, Big data analytics: A framework for unstructured data analysis, *International Journal of Engineering and Technology*, 5(1) (2013), pp.153-156.
- [4] D. P. Acharjya, S. Dehni and S. Sanyal *Computational Intelligence for Big Data Analysis*, Springer International Publishing AG, Switzerland, USA, ISBN 978-3-319-16597-4, 2015. (IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 2, 2016.
- [5] M. K. Kakhani, S. Kakhani and S. R. Biradar, Research issues in big data analytics, *International Journal of Application or Innovation in Engineering & Management*, 2(8) (2015), pp 228-232.

- [6] A. Gandomi and M. Haider, Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management, 35(2) (2015), pp.137-144.
- [7] Elgendy, N.: Big Data Analytics in Support of the Decision Making Process. MSc Thesis, German University in Cairo, p. 164 (2013)
- [8] EMC: Data Science and Big Data Analytics. In: EMC Education Services, pp. 1–508 (2012).