Chapter 27
Big Data and Digital Analytics

Sumathi Doraikannan
Malla Reddy Engineering College (Autonomous), India

Prabha Selvaraj
Malla Reddy Institute of Engineering and Technology, India

ABSTRACT

Data becomes big data when then the size of data exceeds the ability of our IT systems in terms of 3Vs (volume, velocity, and variety). When the data sets are large and complex, it becomes a great difficult task for handling such voluminous data. This chapter will provide a detailed knowledge of the major concepts and components of big data and also the transformation of big data in to business operations. Collection and storage of big data will not help out in creation of business values. Values and importance are created once when the action starts on data by performing an analysis. Hence, this chapter provides a view on various kinds of analysis that can be done with big data and also the differences between traditional analytics and big data analytics. The transformation of digital data into business values could be in terms of reports, research analyses, recommendations, predictions, and optimizations. In addition to the concept of big data, this chapter discuss about the basic concepts of digital analytics, methods, and techniques for digital analysis.

BIG DATA: INTRODUCTION

Big data owns 3 V's (Volume, Velocity and Variety). It is defined as the data that appears to be in huge volume of different data that travels with right speed and reaches the destination within the stipulated time frame so that the data could be used for real-time analysis and further reaction. The structure of big data could be structured or unstructured data that could be in any form such as text files, multimedia files, financial data etc. Data comes from several data sources in order to form data inundation. Data sources could be categorized as electronic gadgets like mobile sensors, medical imaging, smart grids, video surveillance, social media video rendering and genetic engineering etc. so as to form data surge. The fastest- growing sources of big data are social media and genetic engineering. Big data might possess data that contains data structures such as structured, unstructured, semi structured and quasi structured.

The configuration of structured data includes reports, contracts, email and other communications. Structured data displays the information in the form of text files. The information is displayed in columns and rows. Furthermore, it could be easily organized and processed with the help of data mining tools. This could be envisaged as a perfectly organized filing cabinet in which identification, labeling is a simple task and it is easy to access.

Most of the organizations might like to be familiar with this type of data and it is used in an effective and efficient manner. Data does not have any intrinsic structure. It is available in the form of complex data sources such as email, multimedia content, web logs, sales automation and social media data as denoted by Marcos et al (2015). As per Holzinger et al (2013), unstructured information may account for more than 70% to 80% of all data in organizations. Semi structured data does not match with the prescribed structure of data models in the perspective of relationships. Examples of semi structured data could be XML, JSON, MongoDB that stores data in JSON. Quasi structured data is defined as the textual data with inconsistent data formats that could be formatted with effort and time. For example, quasi structured data refer to the click stream data that might consist of inconsistent data values and formats. Few organizations, definite methods must be extended so that it has to manage the huge volume of data. The below Table 1 illustrates the generation of data that occurred in several organizations.

Table 1. Growth of unstructured data

Organizations	Generation
Wordpress (Nawsher Khan et al., 2014)	It is a free open source content management system in which nearly 350 new blogs per minute are published by bloggers
LinkedIn (Nawsher Khan et al., 2014)	A business and employment-oriented social networking service. Through this 2.1 million groups have been created
Apple (Nawsher Khan et al., 2014)	Around 47000 applications are downloaded per minute
Facebook	34,722 likes are registered in each and every minute 47% of Facebook users only access the platform through mobile Users generate 4 million likes every minute More than 250 billion photos have been uploaded to Facebook This equates to 350 million photos per day
YouTube	300 hours of video are uploaded to YouTube every minute! Almost 5 billion videos are watched on Youtube every single day. YouTube gets over 30 million visitors per day In an average month, 8 out of 10 18-49 year-olds watch YouTube.
Twitter	Barack Obama's victory tweet was the most retweeted tweet ever with over 800K retweets Top 3 countries on Twitter are the USA at 107 million, Brazil 33 million and Japan at nearly 30 million The average Twitter user has tweeted 307 times

Components of Big Data

As the name implies, big data contributes much for business applications and it has produced a noteworthy impact in various fields that includes the social media, satellite imaging, banking transactions and healthcare machines. S. Hendrickson (2010) noted that information increase takes place in a rapid manner at the rate of 10x every five years. It was found that the capacity for storage in general-purpose

Related Content

Combining Clustering and Factor Analysis as Complementary Techniques

Lakshmi Prayaga, Krishna Devulapalliand Chandra Prayaga (2020). *International Journal of Data Analytics (pp. 48-57).*

www.irma-international.org/article/combining-clustering-and-factor-analysis-as-complementary-techniques/258920

Four Eras of CRM Selling: Why Big Data and Predictive Analytics Will Unlock the Science of Selling

Cindy Marie Gordon (2015). *Strategic Data-Based Wisdom in the Big Data Era (pp. 83-100).*

www.irma-international.org/chapter/four-eras-of-crm-selling/125046

Future Trends in Community of Inquiry Research and Big Data

 (2018). *The Community of Inquiry Framework in Contemporary Education: Emerging Research and Opportunities (pp. 97-112).*

www.irma-international.org/chapter/future-trends-in-community-of-inquiry-research-and-big-data/198530

Data Mining Problems Classification and Techniques

Nayem Rahman (2018). *International Journal of Big Data and Analytics in Healthcare (pp. 38-57).*

www.irma-international.org/article/data-mining-problems-classification-and-techniques/209740

A Brief Survey on Big Data in Healthcare

Ebru Aydindag Bayrakand Pinar Kirci (2020). *International Journal of Big Data and Analytics in Healthcare (pp. 1-18).*

www.irma-international.org/article/a-brief-survey-on-big-data-in-healthcare/253842