Chapter 15

Introduction to Big Data and Business Analytics

Dineshkumar Bhagwandas Vaghela
Shantilal Shah Government Engineering College, India

ABSTRACT

The term big data has come due to rapid generation of data in various organizations. In big data, the big is the buzzword. Here the data are so large and complex that the traditional database applications are not able to process (i.e., they are inadequate to deal with such volume of data). Usually the big data are described by 5Vs (volume, velocity, variety, variability, veracity). The big data can be structured, semi-structured, or unstructured. Big data analytics is the process to uncover hidden patterns, unknown correlations, predict the future values from large and complex data sets. In this chapter, the following topics will be covered more in detail. History of big data and business analytics, big data analytics technologies and tools, and big data analytics uses and challenges.

INTRODUCTION

The modern technologies generate very complex and unstructured data in very huge amount such as RFID data, web logs, sensors devices, Internet searches, machinery, social networks like Facebook, Twitter and many more, vehicle sensors, portable computers, cell phones, call center records and GPS devices. All these technologies are used in different types of the applications. In sentiment analysis, the sentiments from different sources on specific topic can be collected which are in terms of large volume. The sentiments about the product, movies or any person can be viewed from different official sites or from social media sites such as Twitter, Facebook, Instagram and many more. Politicians and governments often use sentiment analysis to understand how the people feel about themselves and their policies. The Figure 1 shown below represents the sources of Big Data.

The rapid generation of the large volume of data has 5V's characteristics. Here 5V refers Volume, Velocity, Variety, Variability and Veracity. Here 5 V's of big data has been clearly explained in Figure 2.
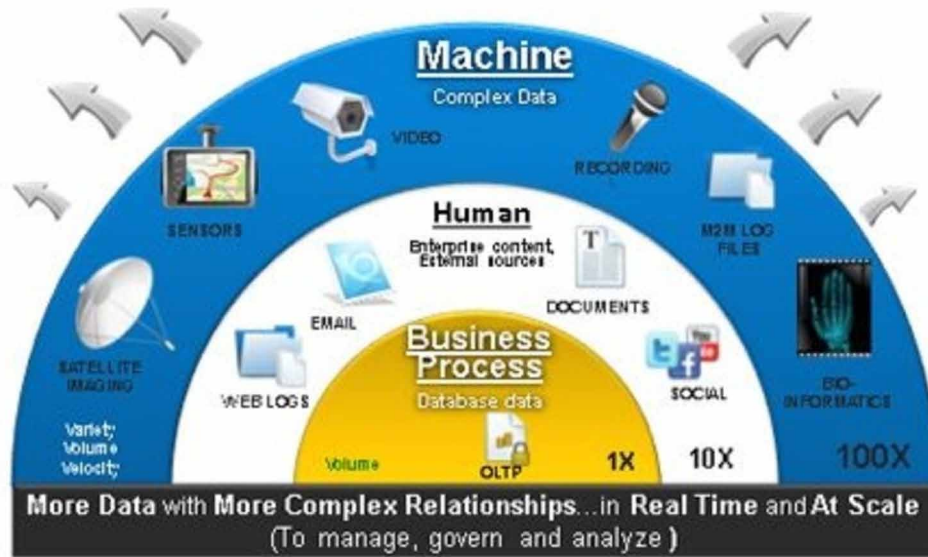
Figure 1. Sources of Big Data

Figure 2. 5 V's of Big Data

5V's (Volume, Velocity, Variety, Variability, Veracity)

Volume

Volume refers an amount of data generated every unit time. Here the data can be of any form like emails, sensor data, video clips, photos, twitter messages etc. which the people generates and share within the unit time period. These data are in the form of hundreds of Zetta Bytes or Bronto Bytes. It has been noticed that on Facebook approximately 10 billion messages have been sent per day with clicking "like" button 4.5 billion times and upload more than 350 million photos every day (He et al., 2013). The amount of data generated per unit time is exponentially increasing; i.e all the data generated in the world between the beginning of time and 2008, the same amount of data will soon be generated every minute (Chen et

Related Content

Qualitative Modeling of Ice Sheet Accumulation for Identification of Mass Balance Variation
Kirill Khvorostovsky, Pavel Lunevand Victoria Shterkhun (2018). *Dynamic Knowledge Representation in Scientific Domains (pp. 99-118).*
www.irma-international.org/chapter/qualitative-modeling-of-ice-sheet-accumulation-for-identification-of-mass-balance-variation/200173

Neural Networks and Bootstrap Methods for Regression Models with Dependent Errors
Francesco Giordano, Michele La Roccaand Cira Perna (2009). *Intelligent Data Analysis: Developing New Methodologies Through Pattern Discovery and Recovery (pp. 272-285).*
www.irma-international.org/chapter/neural-networks-bootstrap-methods-regression/24224

Semantics Enhancing Knowledge Discovery and Ontology Engineering Using Mining Techniques: A Crossover Review
Elena Irina Neaga (2007). *Knowledge Discovery and Data Mining: Challenges and Realities (pp. 163-188).*
www.irma-international.org/chapter/semantics-enhancing-knowledge-discovery-ontology/24906

Boosting Prediction Accuracy of Bad Payments in Financial Credit Applications
Russel Pearsand Raymond Oetama (2010). *Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection (pp. 255-269).*
www.irma-international.org/chapter/boosting-prediction-accuracy-bad-payments/36911

Conceptual Data Warehouse Design Methodology for Business Process Intelligence
Svetlana Mansmann, Thomas Neumuth, Oliver Burgertand Matthias Röger (2010). *Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development: Innovative Methods and Applications (pp. 129-173).*
www.irma-international.org/chapter/conceptual-data-warehouse-design-methodology/39591