

Chapter 2

Big Data Analytics: A Necessary Roadmap for Enterprises

Iman Raeesi Vanani
Allameh Tabataba'i University, Iran

Faezeh Mohammadipour
Allameh Tabataba'i University, Iran

ABSTRACT

The idea that we can get value from data has been discussed, but the main challenge is to use data effectively in order to facilitate smarter and better decision making and surpass our competitors. The change leaders in organization are now dealing with big data from both within and outside the enterprise, including structured and unstructured data, machine data, online and mobile data to supplement their organizational data pool and provide and facilitate the way through which the businesses can compete and operate successfully. Companies that invest in big data can have a distinct advantage over their competitors. Therefore, in this chapter, the concepts of big data analytics along with the relevant description of different categorization, capabilities, challenges are firstly explained, and then big data analytics techniques and methods are introduced and discussed to make the readers familiar with the way big data is applied in the enterprises.

INTRODUCTION

Big Data has been a buzzword in the last decade and it refers to the dynamic, large and disparate volumes of data being created by people, tools and machines. Big Data Analytics is also the process of examining Big Data to uncover hidden

DOI: 10.4018/978-1-5225-5137-9.ch002

patterns, unknown correlations, market trends, customer preferences and other useful information that can help organizations make more informed business decisions. Big Data requires new, innovative and scalable technology to collect, host and analytically process the vast amount of data in order to create real-time business insights that relate to consumers, risk, profit, performance, productivity management and enhanced shareholder value (Erevelles et al., 2016). It includes information obtained from social media, internet-enabled devices (including smartphones and tablets), machine data, video and voice recordings and the continued preservation and logging of structured and unstructured data. Moreover, Big Data nowadays is characterized by the 5 following “V” (Lee, 2017):

- **Volume:** Volume refers to the vast amount of data being created every second in comparison with the traditional data sources. In this regards, we can think of all email, twitter messages, photos, video clips, sensor data that we produce and share every second. On Facebook alone we send 10 billion messages per day, click the “like” button 4.5 billion times and upload and upload 350 million new pictures every day.
- **Variety:** Variety refers to data comes from different sources and is being created by machines as well as people. In the past, we only use the structured data in the rational databases but currently 80% of data is unstructured data such as photos and videos that cannot be stored in the tables.
- **Velocity:** Velocity refers to speed of data which is being generated extremely fast; a process that never stops even while we sleep. For example, it takes millisecond for a trading system to analyze social media networks to pick up the signals that trigger decisions to buy or sell shares.
- **Veracity:** Veracity refers to Big Data that is sourced from many different places; as a result, you need to test the veracity/ quality of the data.
- **Value:** It is well and good to have access to Big Data but if we cannot to drive value from data, it becomes absolutely useless. Therefore, it is the most important “V” in Big Data and we should have the clear understanding of the costs and benefits of Big Data before starting any related initiatives in this area.

Besides from different characteristics of Big Data, the type of data is also important. Therefore, two types of data are found in the area of Big Data: structured data and unstructured data. Structured data refers to information that existed in a relational database and is searchable by simple, straightforward search engine algorithms or other search operations; whereas unstructured data is essentially the opposite and usually refers to information that doesn’t reside in a traditional row-column

27 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/big-data-analytics/218739

Related Content

Mining for Mutually Exclusive Items in Transaction Databases

George Tzanis and Christos Berberidis (2007). *International Journal of Data Warehousing and Mining* (pp. 45-59).

www.irma-international.org/article/mining-mutually-exclusive-items-transaction/1789

A Probabilistic Deep Learning Approach for Twitter Sentiment Analysis

Mostefai Abdelkader (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 367-381).

www.irma-international.org/chapter/a-probabilistic-deep-learning-approach-for-twitter-sentiment-analysis/308497

Clustering of COVID-19 Multi-Time Series-Based K-Means and PCA With Forecasting

Sundus Naji Alaziz, Bakr Albayati, Abd al-Aziz H. El-Bagoury and Wasswa Shafik (2023). *International Journal of Data Warehousing and Mining* (pp. 1-25).

www.irma-international.org/article/clustering-of-covid-19-multi-time-series-based-k-means-and-pca-with-forecasting/317374

Big Data Computing and the Reference Architecture

M. Baby Nirmala and Pethuru Raj (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 56-72).

www.irma-international.org/chapter/big-data-computing-and-the-reference-architecture/150158

An HL7-Aware Decision Support System for E-Health

Pasquale De Meo (2009). *Handbook of Research on Text and Web Mining Technologies* (pp. 670-683).

www.irma-international.org/chapter/hl7-aware-decision-support-system/21750