

Chapter 84

Big Data Technologies and Management

Jayashree K.

Rajalakshmi Engineering College, India

Abirami R.

Rajalakshmi Engineering College, India

ABSTRACT

Developments in information technology and its prevalent growth in several areas of business, engineering, medical, and scientific studies are resulting in information as well as data explosion. Knowledge discovery and decision making from such rapidly growing voluminous data are a challenging task in terms of data organization and processing, which is an emerging trend known as big data computing. Big data has gained much attention from the academia and the IT industry. A new paradigm that combines large-scale compute, new data-intensive techniques, and mathematical models to build data analytics. Thus, this chapter discusses the background of big data. It also discusses the various application of big data in detail. The various related work and the future direction would be addressed in this chapter.

1. INTRODUCTION

Big data is the new approach that is used for analyzing large amount of data-set. Data-set can be from different sources which includes both structured and unstructured data, be it a sensor data or emails or social media (Hiba et al., 2014). The standard method used for analyzing and processing the structured data using Relational Databases is no longer satisfactory due to several challenges. Big data deals with high Volume, Velocity, Variety, Veracity and Value of data.

As the amount of data grow day by day, the cost of storage and management became a vital aspect to be taken care. It was then the evolution of cloud computing that came into existence which provides the best methods to store data in a cost-effective way. The rapid progress of IT industries in different fields like medicine, engineering, business and research requires new technique for processing huge volume of data. Each of the specified field contains voluminous data which has to be managed and controlled

DOI: 10.4018/978-1-6684-3662-2.ch084

for processing. To process and manage the data in an effective way certain mathematical models, data-intensive techniques are required together. A new emerging approach that is used for processing multi-dimensional data which combines mathematical models, data-intensive techniques is termed as big data computing (Kune et al., 2016).

Big data keeps the data safe and secure by utilizing different software tools (Venkatesh et al., 2015). Since data are stored on cloud, the three major criteria like speed, scalability and capacity to process the data are achieved. New business opportunities emerge by using big data as it enables the users to visualize the data. Data analytics is another advantage where the datasets are examined, and information's are obtained.

Section 2 introduces the background of big data. Section 3 discusses big data challenges. Section 4 describes the related work. Section 5 explains the future research directions. The conclusion of the chapter is described in Section 6.

2. BACKGROUND

In day-to-day life millions and billions of data are created everyday by people be it on any social networking sites, finance, medicine, sensors, mobile applications and so on. To manage, analyze and process all these data new processing methodologies, techniques and tools must be used. The information's are being flooded and are beyond a range, in which it cannot to measure. A best example is Facebook, where 100 Terabytes of data are uploaded every day. The term "Big data" came into emergence to store and process huge amount of data in a simple way. There are various definitions of big data stated by researches. According to McKinsey, Big data is a huge amount of data in which the size of the data is beyond the size of regular databases. O'Reilly states that big data is nothing but the data that exceeds the limit of normal database where the data or information is large and processed quickly. Hence new approaches have to be followed to process the information. According to Gartner, Enormous data means high volume, rapid speed and variety of data resources which needs new procedures and techniques to handle the data. New approaches and techniques is difficult to apply on traditional databases by using simple software tools used in accessing the relational database. Data can be both structured as well as unstructured data collected from different resources. All the definitions lead to 5 V's which is Volume, Velocity, Variety, Veracity, Value.

Hence new approaches have to be followed to process the information. According to Gartner, Enormous data means high volume, rapid speed and variety of data resources which needs new procedures and techniques to handle the data. New approaches and techniques is difficult to apply on traditional databases by using simple software tools used in accessing the relational databases database. Data can be both structured as well as unstructured data collected from different resources (Sunny Sharma et al., 2016). All the definitions lead to 5 V's which is Volume, Velocity, Variety, Veracity, Value.

1. **Volume:** Large volume of data are collected from enterprises or organizations. These data are usually in form of terabytes ranges from 30-50 or even more TBs. The available storage cannot handle structure and unstructured data; this is an important challenge for organizations.
2. **Velocity:** Velocity describes the speed of data that an organization can handle and analyze to increase the profit of business before the value of information is lost.

9 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-technologies-and-management/291062

Related Content

Focused Error Analysis: Examples from the Use of the SHEEP Model

Deborah J. Rosenorn-Lanngand Vaughan A. Michell (2016). *International Journal of Big Data and Analytics in Healthcare* (pp. 30-48).

www.irma-international.org/article/focused-error-analysis/171403

A Machine Learning-Based Intelligent System for Predicting Diabetes

Nabila Shahnaz Khan, Mehedi Hasan Muaz, Anusha Kabirand Muhammad Nazrul Islam (2019). *International Journal of Big Data and Analytics in Healthcare* (pp. 1-20).

www.irma-international.org/article/a-machine-learning-based-intelligent-system-for-predicting-diabetes/247455

GPS Travel Diaries in Rural Transportation Research: A Focus on Older Drivers

Trevor Hansonand Eric Hildebrand (2014). *Mobile Technologies for Activity-Travel Data Collection and Analysis* (pp. 293-308).

www.irma-international.org/chapter/gps-travel-diaries-in-rural-transportation-research/113217

Big Data-Driven Socioeconomic Development: An Interdisciplinary Approach

Zhaohao Sun, Zhiyou Wuand Kenneth David Strang (2023). *Handbook of Research on Driving Socioeconomic Development With Big Data* (pp. 1-21).

www.irma-international.org/chapter/big-data-driven-socioeconomic-development/319508

Multi-Criteria Decision Making in Marketing by Using Fuzzy Rough Set

Tapan Kumar Das (2017). *Handbook of Research on Intelligent Techniques and Modeling Applications in Marketing Analytics* (pp. 100-118).

www.irma-international.org/chapter/multi-criteria-decision-making-in-marketing-by-using-fuzzy-rough-set/170341