

Chapter 1

Big Data: Challenges, Opportunities, and Realities

Abhay Kumar Bhadani

Indian Institute of Technology Delhi, India

Dhanya Jothimani

Indian Institute of Technology Delhi, India

ABSTRACT

With the advent of Internet of Things (IoT) and Web 2.0 technologies, there has been a tremendous growth in the amount of data generated. This chapter emphasizes on the need for big data, technological advancements, tools and techniques being used to process big data. Technological improvements and limitations of existing storage techniques are also presented. Since the traditional technologies like Relational Database Management System (RDBMS) have their own limitations to handle big data, new technologies have been developed to handle them and to derive useful insights. This chapter presents an overview of big data analytics, its application, advantages, and limitations. Few research issues and future directions are presented in this chapter.

INTRODUCTION

With digitization of most of the processes, emergence of different social network platforms, blogs, deployment of different kind of sensors, adoption of hand-held digital devices, wearable devices and explosion in the usage of Internet, huge amount of data are being generated on continuous basis. No one can deny that Internet has changed the way businesses operate, functioning of the government, education and lifestyle of people around the world. Today, this trend is in a transformative stage, where the rate of data generation is very high and the types of data being generated surpass the capability of existing data storage techniques. It cannot be denied that these data carry a lot more information than ever before due to the emergence and adoption of Internet.

Over the past two decades, there is a tremendous growth in data. This trend can be observed in almost every field. According to a report by International Data Corporation (IDC), a research company claims that between 2012 and 2020, the amount of information in the digital universe will grow by 35 trillion

DOI: 10.4018/978-1-5225-0182-4.ch001

gigabytes (1 gigabyte equivalent to 40 (four-drawer) file cabinets of text, or two music CDs). That's on par with the number of stars in the physical universe! (Forsyth, 2012).

In the mid-2000s, the emergence of social media, cloud computing, and processing power (through multi-core processors and graphics processing unit (GPUs)) contributed to the rise of big data (Manovich, 2011; Agneeswaran, 2012). As of December 2015, Facebook has an average of 1.04 billion daily active users, 934 million mobile daily active users, available in 70 languages, 125 billion friend connections, 205 billion photos uploaded every day, 30 billion pieces of content, 2.7 billion likes, and comments are being posted and 130 average number of friends per Facebook user (Facebook, 2015). This has created new pathways to study social and cultural dynamics.

Making sense out of the vast data can help the organizations in informed decision making and provide competitive advantage. Earlier, organizations used transaction-processing systems that inherently used Relational Data Base Management Systems (RDBMS) and simple data analysis techniques like Structured Query Language (SQL) for their day-to-day operations that helped them in their decision making and planning. However, due to the increase in the size of data especially the unstructured form of data (for example, customer reviews of their Facebook pages or tweets), it has become almost impossible to process these data with the existing storage techniques and plain queries.

In this chapter, an overview of big data including its sources to dimensions is given. The limitations of existing data processing approaches; need for big data analytics and development of new approaches for storing and processing big data are briefed. The set of activities ranging from data generation to data analysis, generally termed as Big Data Value Chain, is discussed followed by various applications of big data analytics. The chapter concludes by discussing the limitations of big data analytics and provides direction to open issues for further research.

BACKGROUND AND NEED FOR BIG DATA ANALYTICS

Storage and retrieval of vast amount of data within a desirable time lag is a challenge. Some of these limitations to handle and process vast amount of data with the traditional storage techniques has led to the emergence of the term "Big Data". Though big data has gained attention due to the emergence of the Internet, but it cannot be compared with it. It is beyond the Internet, though Web makes it easier to collect and share knowledge as well data in raw form. Big Data Analytics (BDA) is all about how these data can be stored, processed, and comprehended such that it can be used for predicting the future course of action with a great precision and acceptable time delay.

Marketers focus on target marketing, insurance providers focus on providing personalized insurance policies to their customers, and healthcare providers focus on providing quality and low-cost treatment to patients. Despite the advancements in data storage, collection, analysis and algorithms related to predicting human behavior; it is important to understand the underlying driving as well as the regulating factors (market, law, social norms and architecture), which can help in developing robust models that can handle big data and yet yield high prediction accuracy (Boyd & Crawford, 2011).

The current and emerging focus of big data analytics is to explore traditional techniques such as rule-based systems, pattern mining, decision trees and other data mining techniques to develop business rules even on the large data sets efficiently. It can be achieved by either developing algorithms that use distributed data storage, in-memory computation or by using cluster computing for parallel computation.

22 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/157681

Related Content

Probabilistic Control and Swarm Dynamics in Mobile Robots and Ants

Eugene Kagan, Alexander Rybalov, Alon Sela, Hava Siegelmann and Jennie Steshenko (2014). *Biologically-Inspired Techniques for Knowledge Discovery and Data Mining* (pp. 11-47).

www.irma-international.org/chapter/probabilistic-control-and-swarm-dynamics-in-mobile-robots-and-ants/110453

Mining Partners in Trajectories

Diego Vilela Monteiro, Rafael Duarte Coelho dos Santos and Karine Reis Ferreira (2020). *International Journal of Data Warehousing and Mining* (pp. 22-38).

www.irma-international.org/article/mining-partners-in-trajectories/243412

A New Approach for Fairness Increment of Consensus-Driven Group Recommender Systems Based on Choquet Integral

Cu Nguyen Giap, Nguyen Nhu Son, Nguyen Long Giang, Hoang Thi Minh Chau, Tran Manh Tuan and Le Hoang Son (2022). *International Journal of Data Warehousing and Mining* (pp. 1-22).

www.irma-international.org/article/a-new-approach-for-fairness-increment-of-consensus-driven-group-recommender-systems-based-on-choquet-integral/290891

Measuring the Financial Crisis in Local Governments through Data Mining

José Luis Zafra-Gómez and Antonio Manuel Cortés-Romero (2010). *Data Mining in Public and Private Sectors: Organizational and Government Applications* (pp. 21-46).

www.irma-international.org/chapter/measuring-financial-crisis-local-governments/44281

Literature Review in Computational Linguistics Issues in the Developing Field of Consumer Informatics: Finding the Right Information for Consumer's Health Information Need

Ki Jung Lee (2009). *Handbook of Research on Text and Web Mining Technologies* (pp. 758-765).

www.irma-international.org/chapter/literature-review-computational-linguistics-issues/21756