Chapter 11
Big Data Analytics in Cloud Computing:
Effective Deployment of Data Analytics Tools

Rajganesh Nagarajan
A. V. C. College of Engineering, India

Ramkumar Thirunavukarasu
VIT University, India

ABSTRACT

In this chapter, the authors consider different categories of data, which are processed by the big data analytics tools. The challenges with respect to the big data processing are identified and a solution with the help of cloud computing is highlighted. Since the emergence of cloud computing is highly advocated because of its pay-per-use concept, the data processing tools can be effectively deployed within cloud computing and certainly reduce the investment cost. In addition, this chapter talks about the big data platforms, tools, and applications with data visualization concept. Finally, the applications of data analytics are discussed for future research.

INTRODUCTION

Big data is an evolving term that describes huge amount of structured, semi-structured and unstructured data. In addition, big data (Kolomvatsos et al., 2015) refers to the use of predictive analytics, behavior analytics, or advanced data analytics for extracting the real inside values from different kind of data (Boyd et al., 2012). The basic characteristics of big data includes volume, variety and velocity (Hilbert & Martin, 2016). Volume represents the quantity of generated and stored data. The size of the data is the key factor to determine the value and its potential insight for considering the whether it is big data or not. Variety includes the data type and its nature such as structured, semi-structured and unstructured

DOI: 10.4018/978-1-6684-3662-2.ch011

data. Velocity represents the speed of the data in which generated and processed to meet the demands and challenges.

Big data environments require clusters of computing servers to support the tools that process the structured, semi-structured and unstructured data. Though big data offers various kinds of analysis such as descriptive modeling, predictive modeling, and prescriptive modeling, it is crucial to analyze and synthesize interesting pattern from diversified data sources. This makes cloud based data analytics, a viable research field and open new research avenues in modeling and analyzing complex data. Furthermore, cloud services enable infrastructures to be scaled up and down rapidly, adapting the system to the actual demand. Hence, this chapter addresses the importance of cloud computing in order to support the big data analytics. In such circumstances, the primary features of cloud computing such as on-demand provisioning, pay-per-usage provides significant improvement in the process of data analytics. The important challenges of big data are discussed and further research directions with the aid of cloud computing is presented.

CATEGORIES OF BIGDATA AND CHALLENGES ASSOCIATED

With the advent of internet and smart devices, the manipulation of data increases rapidly. In addition, there is no such common mechanism followed for the representation of data. In such scenario, it is important to process different kinds of data (Agresti & Kateri, 2011) before formulating the information. With respect to the existing data processing mechanisms, it is essential to invent a new data processing methodology with less capital investment. Accordingly, cloud computing has been highly recommended to incorporate the data processing activities by running the new data analytics tools. Hence, this section highlights the various categories of big data and its processing tool. Finally, the incorporation of cloud computing enhances the data analytics process in an effective manner.

Categories of Data

1. **Structured Data**: Refers to any kind of data that has a proper format and resides in a record or file. Structured data (Chang et al., 2008) are easy to input, query, store, and analyse. Examples of structured data include numbers, words, and dates.
2. **Semi-Structured Data:** The data that are not following the conventional or relational data base system are called as semi-structured data (Sagiroglu & Sinanc, 2013). The data are not organized in table format. In order to analyse the semi-structured data, the complex rules must be used.
3. **Unstructured Data:** The text messages, location information, videos, and social media information (Feldman & Sanger, 2007) are data that do not follow any prescribed format. Always the size of this data is increasing because of the use of new technological devices such as smartphones. Therefore, the understanding of such data become a more challenging one.

Challenges in Big Data

The challenges in big data require more attention to avoid the failure of technology with some unpleasant results. Some of the identified challenges (Labrinidis et al., 2012; Chen et al., 2014) are given here to understand the common issues.