Journal of Soft Computing and Decision Support Systems

Big Data Tools: Advantages and Disadvantages

Maria Ijaz Baig [a,*], Liyana Shuib [a,*], Elaheh Yadegaridehkordi [a,*]

[a] Department of Information Systems, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

* Corresponding authors email address: mariaijazbaig@hotmail.com, liyanashuib@um.edu.my, yellahe@gmail.com

Abstract

Big data tools have increasingly become crucial requirements of managing the complex and voluminous data. The selection of right tool requires an in-depth knowledge of existing big data tools and their potentiality. This paper provides a review of big data tools by selecting 34 articles from 2011 to 2018. This study provides pertinent information about most popular big data tools. The various big data tools related advantages and disadvantages are also discussed in detail. The findings of this study categorized the big data tools according to their potentiality. This study is beneficial for researchers to explore the big data sets according to its potentiality. It also provides deep insight of big data tools applicability to apply in real-time environment. This research is also helpful for practitioners to select the right big data tool according to requirement.

Keywords: Big data tools, Big data advantages and disadvantages, Big data potentiality

1. Introduction

The modern world enormously generated big amount of data. This big data is producing from variety of sources. The size of data is increasing day by day and recently reached to several xenottabyte. Big data is known by three characteristics, referred as the 3V's (Volume, Velocity and Variety). Volume refers to the size of the data. Velocity refers to the speed at which data is accessible (streaming of data). Variety refers to the type of the data (Zhang et al., 2018). Meanwhile, big data has different forms of structured (e.g., database and oracle), unstructured (e.g., videos, audio, images), and semi-structured data (e.g., commerce) (Shukla, Radadiya and Akotiya, 2015). The big data characteristics are significant to understand the nature of data and use the available tools accordingly. The existing tools have potential to handle the characteristics of big data (Landset et al., 2015; Zhang et al., 2018). Now, many tools such as Hadoop, HDFS, MapReduce, YARN, Hbase, Hive, Mahout, and Zookeeper have appeared out to manage the large data (Simovic, 2018). However, these tools application datasets matched with the characteristics of big data (Oussous et al., 2018). Conversely, each of the tool has advantages and disadvantages. Currently, a lot of research in big data has been on revolution, significance, analyzing the factors related to acceptance of technology (Mayer and Cukier, 2013; Jin et al., 2015; Marz and Warren, 2015; Zhang et al., 2018). However, none of the study focuses on big data potentiality and analyzes the advantages and

disadvantages of big data tools accordingly. Therefore, this paper aims to present a short review of recent research on big data with a specific focus on big data potentiality, tools, and related advantages and disadvantages. The main research questions of this study are:

1) What are the most popular big data potentiality and tools?
2) What are the advantages and disadvantages of big data tools?

The rest of the paper is structured as follows. The research methodology is presented in Section 2. Section 3 discusses the findings. Conclusion is given in the Section 4.

2. Methodology

The aim of this study is to find out the big data tools and its advantages and disadvantages. For this purpose, initially various databases were explored to find out the relevant studies. During first search stage (S1), it has been observed that majority of studies were exist in IEEE Xplore, Emerald Insight, Scopus Elsevier, Springer Link and ACM. Additionally, keywords were searched to find as many articles. The key strings of 'big data tools', combine with advantages and disadvantages were explored. The inclusion and exclusion criteria were also considered for final selection of studies. This study considered the articles that published in English language from January 2011 to