Chapter 55
Insight Into Big Data Analytics: Challenges, Recent Trends, and Future Prospects

Mohd Vasim Ahamad
Aligarh Muslim University, India

Misbahul Haque
Aligarh Muslim University, India

Mohd Imran
Aligarh Muslim University, India

ABSTRACT

In the present digital era, more data are generated and collected than ever before. But, this huge amount of data is of no use until it is converted into some useful information. This huge amount of data, coming from a number of sources in various data formats and having more complexity, is called big data. To convert the big data into meaningful information, the authors use different analytical approaches. Information extracted, after applying big data analytics methods over big data, can be used in business decision making, fraud detection, healthcare services, education sector, machine learning, extreme personalization, etc. This chapter presents the basics of big data and big data analytics. Big data analysts face many challenges in storing, managing, and analyzing big data. This chapter provides details of challenges in all mentioned dimensions. Furthermore, recent trends of big data analytics and future directions for big data researchers are also described.

INTRODUCTION

Big data analytics is the process of extracting hidden patterns and correlations, consumer behavior and preferences, market trends and decision making, by examining huge data sets coming from various sources such as web log files, social media, satellites and sensors, GPS data, IoT (Internet of Things) enabled devices, etc. When we click on a website, a large data is saved in the form of web log files. Which can

DOI: 10.4018/978-1-6684-3662-2.ch055

be used in recommender services in future transactions. Facebook, Tweeter, Instagram and various other social media are generating very huge data every day in terms of contents, tweets, photos etc. They must be saved for further processing. Sensor can be embedded in machines that senses the inputs from the outer world and provide it to the machine for further analysis. Hence, sensors can generate a large volume of data. There many handheld and IoT enabled devices which generated huge data. To extract meaningful pattern from big data, we need to apply application specific analytical methods. As, enormous data are coming from thousands of sources in structured, unstructured and semi-structured formats, it's a very challenging task to analyze it.

There are following challenges with respect to data storage, data management, analyzing and processing the big data, scalability, privacy and security. Huge data are coming from thousands of sources in different formats, it is a big challenge to store them in an efficient, unambiguous and scalable form. Big data is in the scale of Exabyte. Big data requires special kind of techniques to handle the data. It is not possible for the traditional tools to process the big data. To process them, we need a cluster of machines that can process the data in parallel. So, we use some big data analytics technologies such as Hadoop, Spark, Pig, Hive, etc., to manage and process the big data.

In big data analytics, we deal with huge amount of data with different format, inconsistent, noisy and incomplete data, which generates following challenges. Do all the data need to be analyzed? Do the stored data suitable for analysis? How to find interesting patterns from such a huge, multi-formed, inconsistent, incomplete, uncertain and noisy data? etc. It is much possible that the approach used for big data analytics provides good results on "small" big data but performance degrades rapidly for comparatively larger datasets. It's a challenging task to produce high quality of information from huge datasets with minimum time, resources and cost.

In recent years, large numbers of techniques and tools & technologies have been developed to analyze the big data. The techniques used for big data are clustering, classification, machine learning, neural networks, topic modelling, etc. To incorporate these techniques to analyze the big data, we have technologies such as Hadoop, Spark, Cassandra, Pig, Hive, NoSQL, HBase, MapReduce, etc. In future, advanced analytics and visualization techniques will be applied on real time business intelligence. To get high performance, in-memory datasets usage will be accelerated.

BACKGROUND

Big Data

Before the evolution of Big Data, data around the world was not so huge with limited types. Analytics methods were developed to deal with the less amount of data, some defined sources and limited types of data. Now, data is in the scale of Petabytes and Exabytes (Akhtar *et al,* 2015).

The big data is different from the traditional types of data in many dimensions. These dimensions are the foundation for defining the term Big Data. Doug Laney defined the term big data with three Vs: Volume, Velocity and Variety (Doug Laney, 2001). Some literatures also supports two other dimensions that characterizes the big data namely Variability and Value (A. Katal *et al,* 2013).