Chapter 56
A Study of Big Data Processing for Sentiments Analysis

Dinesh Chander
Panipat Institute of Engineering and Technology, India

Hari Singh
Jaypee University of Information Technology, India

Abhinav Kirti Gupta
Jaypee University of Information Technology, India

ABSTRACT

Data processing has become an important field in today's big data-dominated world. The data has been generating at a tremendous pace from different sources. There has been a change in the nature of data from batch-data to streaming-data, and consequently, data processing methodologies have also changed. Traditional SQL is no longer capable of dealing with this big data. This chapter describes the nature of data and various tools, techniques, and technologies to handle this big data. The chapter also describes the need of shifting big data on to cloud and the challenges in big data processing in the cloud, the migration from data processing to data analytics, tools used in data analytics, and the issues and challenges in data processing and analytics. Then the chapter touches an important application area of streaming data, sentiment analysis, and tries to explore it through some test case demonstrations and results.

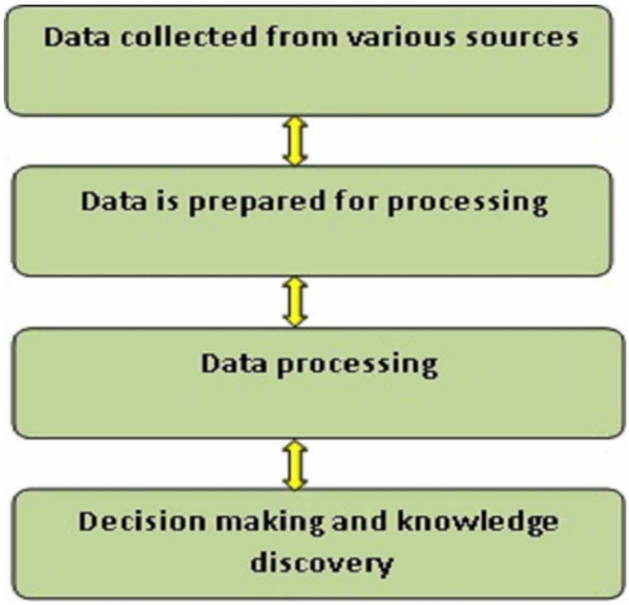
DATA PROCESSING

Since last decade, rapid development of Internet enabled services such as social media, Internet of Things, and cloud based services have led to tremendous growth of data termed as big data. This data has become very difficult to be handled and managed for further processing (Jin et al., 2015). It has been estimated that around 2.5 quintillion bytes of new data is generated per day and expected to be more in near future as the number of internet users are growing unprecedentedly. This exponential growth of data has posed many challenges in front of researchers, academia and Industry across the globe. Moreover,

DOI: 10.4018/978-1-6684-3662-2.ch056

the big data is unstructured: it varies in volume, velocity, veracity and variety makes (4Vs) it more challenging to manage and process (Mishra, R. K., & Mishra, R. K., 2018). This sudden explosion of data in terabytes, petabytes and exabytes could not be handled by the traditional database such as SQL led to the emergence of new tools and techniques to process the big data (Storey, V. C., & Song, I. Y., 2017).

Figure 1. Big data chain



Big data processing and analysis have become very crucial for better decision making, knowledge discovery, business intelligence and actionable insights. The Fig-1 represents the big data chain i.e. from data collection to decision making (Janssen, M., van der Voort, H., & Wahyudi, A., 2017). Big data is collected in raw form from various sources of interest which need to be prepared for processing. Next the quality data sets are prepared for further processing using data cleansing and standardization. After that, data processing takes place which includes transformation, aggregation and pattern generation. Once the data processing is completed, various reports are generated and analyzed for better decision making, knowledge discovery and insight or trends. Analysis of data could be classified as descriptive, diagnostic, predictive and prescriptive (Perwej, Y., 2017).

This book chapter proposes to show various tools, techniques, and technologies of data processing and analytics. Later, the use streaming data for sentiment analysis through executable test cases is presented. Sentiment analysis is performed on run-time tweets with Python using twitter API “tweepy” and obtained results are presented through plots.

A survey on various sentiment analysis methods used by researchers is also presented. This would also help in identifying the best one and possibly may be in predicting a newer one.