

# **Mining Large Data with Hadoop**

## **Abstract**

The rapid growth in the size of datasets that are collected from different resources has made capturing, managing, or analyzing the datasets beyond the ability of most software tools and people. Dealing with large data (big data) has shown many issues and proposed solution by using Hadoop framework which is discussed in the tutorial. Hadoop is an open source project which was implemented in Java. Hadoop provides a framework for distributed computing using MapReduce programming model. As a framework, it needs support with implementation projects such as Mahout and Hive, which are run on top of Hadoop for mining big data. Mahout and Hive projects have implementations for data mining algorithms and techniques which are presented in the tutorial. RHive package has been proposed at the end of the tutorial as a solution for analysts who are not interested in Java programming.