



RESEARCH ARTICLE

IMPLEMENTATION OF PARALLEL APRIORI ALGORITHM ON HADOOP CLUSTER

A. Ezhilvathani¹, Dr. K. Raja²

¹P.G Student, M.E CSE, Alpha College of Engg, Chennai, India

²Dean (Academics), Alpha College of Engg, Chennai, India

Abstract— Nowadays due to rapid growth of data in organizations, large scale data processing is a focal point of information technology. To deal with this advancement in data collection and storage technologies, designing and implementing large-scale parallel algorithm for Data mining is gaining more interest. In Data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. This paper aims to extract frequent patterns among set of items in the transaction databases or other repositories. Apriori algorithms have a great influence for finding frequent item sets using candidate generation. Apache Hadoop software framework is used to build the cluster. It working is based on MapReduce programming model. It is used to improve the processing of large-scale data on high performance cluster. It processes vast amount of data in parallel on large cluster of computer nodes. It provides reliable, scalable, distributed computing.

Key Terms: - Hadoop; MapReduce; Apriori

Full Text: <http://www.ijcsmc.com/docs/papers/April2013/V2I4201396.pdf>