WORK PLAN

Our proposed work can be fast distributed data processing and it is famous for its speed with which it processes the data because of its in-memory feature, and minimizing space complexity when the data source is too large and minimum support is higher for mining frequent itemset. Setting up multiple minimum support values for the items. Generation of only user interested association rules. The process of producing association rules involves the task of finding the set of all the frequent itemsets and generating promising rules.

This research work basically deals with effective recommendation system for the supermarket or E-Shoping and agriculture for identifying the most frequently purchased items and suitable crop respectively using association rule techniques. These proposed techniques are much useful in both supermarket and agriculture
6. RESEARCH METHODOLOGY

In order to overcome exiting algorithm problems like time consuming and memory related issues in this work we are using apache spark and scala. Apaches spark is a cluster computing framework which runs on top of Hadoop and process different types of data. Spark is a framework like Hadoop and scala is programming language and other than scala language python and java also used.

Apache Spark performs in-memory operations by copying the data from distributed storage into RAM memory which is much faster. As a result of this, the time consumed to read and write from cluster is reduced. Using Spark we can mine frequent itemset and we can find correlations or association rules. Using spark we can overcome problems which are identified in Apriori, FP-Growth and Hadoop.

- Using spark no need to generate candidate itemset like Apriori algorithm.
- When data size is too large for mining frequent itemset as well as to generate relations between items will take less memory computation and cost compare to Apriori algorithm.
- Distributed file processing and in memory process whenever reading and writing data form cluster and this process take less CPU usage.