Abstract

Data mining mostly use for data analysis and identifying frequent dataset. Now a days cloud computing is used for data storage and many other data operations like data mining, data retrieval, data distribution etc. As data increasing very rapidly on server day by day, many complications are introduced. Most common problems are load balancing on server and time optimization. To overcome these limitations parallel frequent dataset mining is very effective method. Fidoop parallel frequent dataset mining algorithm which is based on mapreduce framework helps to improve load balancing and FiDoop-HD, speed up the mining performance for high-dimensional data analysis. Fidoop is very efficient and scalable algorithm for large clusters of data.

References


Index Terms

Computer Science Information Systems

Keywords

Frequent item sets, Frequent Items Ultrametric trees, Hadoop, MapReduce.