Abstract

A MapReduce is a framework that allows processing the very big amounts of formless data in parallel across a distributed cluster of processors or individual computers. The MapReduce framework is mostly used to analyze the large amount of datasets in clustering environments. MapReduce has become a dominant parallel computing paradigm for big data. This paper describes well known strategies in MapReduce, and present comprehensive comparative algorithms in MapReduce in clustering environment.
References

- Guoping Wang and CheeYong Chan, "MultiQuery Optimization in MapReduce Framework", 40th International Conference on Very Large Data Bases, September 2014.
- Prajesh P Anchalia, Anjan K Koundinya, Srinath N K, "MapReduce Design of K-Means Clustering Algorithm", 2013 IEEE.
- Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", Published in Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96).

Index Terms

Computer Science Algorithms

Keywords

Big Data Clustering Algorithm Mapreduce