

From Two-Way to Multi-Way: A Comparative Study for Map-Reduce Join Algorithms

Authors : Marwa Hussien Mohamed, Mohamed Helmy Khafagy

Abstract : Map-Reduce is a programming model which is widely used to extract valuable information from enormous volumes of data. Map-reduce designed to support heterogeneous datasets. Apache Hadoop map-reduce used extensively to uncover hidden pattern like data mining, SQL, etc. The most important operation for data analysis is joining operation. But, map-reduce framework does not directly support join algorithm. This paper explains and compares two-way and multi-way map-reduce join algorithms for map reduce also we implement MR join Algorithms and show the performance of each phase in MR join algorithms. Our experimental results show that map side join and map merge join in two-way join algorithms has the longest time according to preprocessing step sorting data and reduce side cascade join has the longest time at Multi-Way join algorithms.

Keywords : Hadoop, MapReduce, multi-way join, two-way join, Ubuntu

Conference Title : ICBDC 2018 : International Conference on Big Data Computing

Conference Location : Bangkok, Thailand

Conference Dates : February 08-09, 2018