Abstract

Hadoop mapreduce could be a powerful data processing technique giant for large information analysis on distributed artifact between clusters like clouds. In this paper we proposed the improved capacity scheduler to improve the existing scheduler issues that help the scheduler to execute the task in less time. We introduced pipeline and queue management in our proposed work for improving the performance of hadoop. Our experimental result show that our strategies can result in about 29 to 50 % decrease in average response time.

Index Terms

Computer Science

Databases

Keywords

Mapreduce, HDFS, Hadoop, Scheduler.