Abstract

In the Big Data group, MapReduce has been seen as one of the key empowering methodologies for taking care of ceaselessly expanding requests on figuring assets forced by Big Datasets yet at the same time numerous issues arrive with MapReduce keeping in mind the end goal to handle a much more extensive cluster of employments, combination into Hadoop’s native file system. The purpose behind this is the high versatility of the MapReduce worldview which takes into account hugely parallel and circulated execution over an expansive number of figuring hubs. This paper address the how supplant MapReduce with Apache Spark as the default preparing for Hadoop. Apache Spark is superior to MapReduce towards leads issues and difficulties in taking care of Big Data with the target of giving an outline of the field, encouraging better arranging and administration of Enormous Information ventures, larger amount reflection and speculation of MapReduce.


Index Terms

Computer Science

Databases

Keywords

Big Data, Big Data Analytics, MapReduce, Apache Spark