Abstract

In the recent years, Big Data has become ubiquitous and various big data tools are greatly in use to accelerate the computing and analytics in various fields. Various algorithms in Computer Science use large and heterogeneous data sets; and hence could be explored with Big Data platforms. One such class of algorithms is stream clustering algorithms; dealing with large scale processing of incremental data. This motivation of using Big Data tools may lead to improved efficacy of running the algorithms. Hadoop, the most popular open source implementation of MapReduce, has been utilized and modified for catering the needs of numerous clustering problems. But various scientific and computing fields are also using MongoDB, a document oriented NoSQL store supporting Map Reduce. The main purpose of this paper is to try and judge the usage of MongoDB as a Big Data platform for implementing a stream clustering algorithm using MapReduce programming model to study the factors relating Map Reduce and MongoDB together.

References

- Marr, B. (Feb 2014). A Talk on Big Data- the 5 Vs Everyone must know.

Index Terms

Computer Science  Algorithms

Keywords

Big Data  MapReduce  Sharding  Clustering  Grid