Abstract

Hadoop is popular large scale open source software framework which is written in JAVA programming for securely distributes storage and it is the master implementation of Map-Reduce programming used for cloud computation [1]. Now a days, hadoop faces a lot of problems to obtain the best outcomes from underlying system. The issue includes a serialization needs to gain quality performance which setback the aspect. Disk access and repetitive merges causes to current speedy interconnections that increases the volume of data sets. To stay with increasing volume of data sets, Hadoop also requires I/O ability from the underlying system nodes to process and examine data. So, for this HADOOP-A [12] architecture is formed. Hadoop-A is an enhancement of framework that minimizes hadoop
with peripherals for speedily data movement and bounding the existing limits to keep updating the architecture. A novel network algorithm for merging the data is explained in this paper. In supplementary, a full pipeline which is designed to overlay the shuffle, minimize phases and merge. The experimental results which shows that HADOOP-A is intensely speeds up data processing in Map – Reduce and extends the hadoop’s throughput as double. HADOOP-A is significantly helps to optimize disk accesses which are caused by intermediate data.

References


Index Terms

Computer Science  Networks

Keywords

Serialization  Repetitive Merges  Disk Access  Network Portability  Network-levitated Pipelined Merge
Shuffle
Merge
And Reduce.