Abstract

Hadoop is a 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

- J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large
  In-Depth Study," Proc. VLDB Endowment,
- T. Condie, N. Conway, P. Alvaro, J. M. Hellerstein, K. Elmeleegy, and R. Sears,
  and Implementation (NSDI), pp. 312-328, Apr. 2010.
- J. Dean and S. Ghemawat. MapReduce: Simplified data processing on large clusters.
  Sixth Symp. On Operating System Design and Implementation (OSDI), pages 137–150,
  December 2004.
- Dawei Jiang, Beng Chin Ooi, Lei Shi, and Sai Wu. The performance of MapReduce: An
  in-depth study. In Proceedings of the 36th International Conference on Very Large Data Bases
- Yandong Mao, Robert Morris, and Frans Kaashoek. Optimizing MapReduce for
- Konstantin Shvachko, Hairong Kuang, Sanjay Radia, and Robert Chansler. The hadoop
  Computer Society.
- Yandong Wang, Xinyu Que, Weikuan Yu. Hadoop Acceleration through Network
- Weikuan Yu, Member, IEEE, Yandong Wang, and Xinyu Que. Design and Evaluation of
  Network-Levitated Merge for Hadoop Acceleration: IEEE TRANSACTIONS ON PARALLEL
Computer Science

Index Terms

Networks

Keywords

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

And Reduce.