Abstract

In recent years, there are various sources that generates data in petabyte and terabyte, known as big data and its generated by human, machine, sensor etc. So the solution leds with big data is Apache Hadoop has attracted strong attention because of its applicability on processing for the large data sets. This paper present the review about the big data and it’s characteristics and also the types of the open source tools environment like HADOOP. The objective of paper to identify the power of the Hadoop on the big data and motivation behind the new research and outlines to address to Apache Hadoop also includes the programming paradigm that is Map-Reduce.

References

1. IvaniltonPolato a,b,n, ReginaldoRé b, AlfredoGoldman a, FabioKon a A comprehensive review of Hadoop research—A systematic 46 (2014)1–25
2. Seyed Reza Pakize, A Comprehensive View of Hadoop MapReduce Scheduling
4. Kamble Ashwini, Kanawade Bhavan, A Brief on MapReduce Performance Volume 1 Issue 1 (April 2014)
5. Lu Lu, Xuanhua Shi *, Hai Jin, Qiuyue Wang, Daxing Yuan, Song Wu, Morpho: A decoupled MapReduce framework for elastic cloud computing, 36 (2014) 80–90
7. Lizhe Wanga,b,*, Jie Taoc, Rajiv Ranjan d, Holger Martenc, Achim Streit c, Jingying Chene, Dan Chena,**, G-Hadoop: MapReduce across distributed data centers for data-intensive computing 29 (2013) 739–750
8. Vidyasagar S.D, A Study on “Role of Hadoop in Information Technology era” Volume : 2 | Issue : 2 | Feb 2013 • ISSN No 2277 - 8160
9. Shaochun Wu , Xiang Shuai, Liang Chen, Ling Ye, Bowen Yuan, A replica pre-placement strategy based on correlation analysis in cloud environment (CCIS 2013)
12. Hortonworks, Community Driven Apache Hadoop Apache Hadoop Basics May 2013 ©
13. S. Chandra Moulishwaran And Shyam Sathyan*, Study On Replica Management And High Availability In Hadoop Distributed File System (Hdfs), Vol 2 / Issue 2 / 2012 / 65-70
16. J. Gerard Wolff, (Member, Ieee) Cognitionresearch.Org, Menai Bridge, U.K. (Jgw@cognitionresearch.org) Big Data And The Sp Theory Of Intelligence ,Received October 20, 2013, Accepted March 27, 2014, Date Of Publication April 2, 2014, Date Of Current Version April 15, 2014. Digital Object Identifier 10.1109/Access.2014.2315297
19. Md. Rezaul Karim1, Azam Hossain1, Md. Mamunur Rashid1, Byeong-Soo Jeong1, and Ho-Jin Choi2,” An Efficient Market Basket Analysis Technique with Improved MapReduce Framework on Hadoop: An E commerce Perspective
21. S. Chandra Moulishwaran And Shyam Sathyan, “Study On Replica Management And High Availability In Hadoop Distributed File System (Hdfs)” S. Chandra Moulishwaran And Shyam
Sathyan. Et Al. / Journal Of Science / Vol 2 / Issue 2 / 2012 / 65-70


24. Ivan Baev† Rajmohan Rajaraman‡ Chaitanya Swamy§,”Approximation Algorithms for Data Placement Problems

25. R.Jemina Priyadarsini1, Dr.L.Arockiam2,” An Extensive Analysis On Task Scheduling Algorithms In Cloud Environments” (Ijetcas)"


28. Phokham Nonava October 2014,” HDFS Blocks Placement Strategy


30. George Porter UC San Diego La Jolla,” Decoupling Storage and Computation in Hadoop with SuperDataNodes”.

Index Terms

Computer Science Information Sciences

Keywords

Bigdata, Apache Hadoop, Map-Reduce, distributed system HDFS, MPI, G-Hadoop, Gfram, 3V Big Data model, 5 V Big Data model.