Abstract

Data analytics has been rapidly growing in a variety of application areas like mining business intellect for processing the huge amount of data. MapReduce programming paradigm adds itself well to these data-intensive analytics jobs, given its one of the well known ability to scale-out and force several machines to parallely process data. This paper introduces a detailed analysis of big data over cloud computing with several mapred techniques from system and application aspects. Here in this work we say that such Mapper and Reducer based analytics provide a better result over cloud platform. However, End-Users in this environment has the ability to use MapReduce applications to minimize the incurred cost, while obtaining the best performance. From the implementation point of view, we describe the key issues and challenges of big data on cloud and on local system as well. At last, the challenges come across in implementing MapReduce functions over Hadoop and the analysis of standalone, clustered and virtualized systems over our test-bed.
Performing Big Data over Cloud on a Test-Bed

- RedHat Enterprises virtualization workbook for student.
- Big Data Processing in Cloud Computing Environments College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
- Virtualization in Linux a Key Component for Cloud Computing
- "White Paper: Ten Things You Need to Know About Virtualization." www.datacore.com
- Implementation of MapReduce Algorithm and Nutch Distributed File System in Nutch published by IJCA 2011
- Pipeline computing. From Wikipedia en.wikipedia.org/wiki/Pipeline_(computing)

Index Terms

Computer Science
Distributed Systems

Keywords

Cloud computing  Big data  Hadoop  HDFS  Map Reduce  virtualization