Abstract

A large amount of data can be referred as BigData. A vast size of data requires special kind of methodology to process and store. BigData research consortium team developed a distributed server known as Hadoop Server, to divide and partition large data into multiple pieces for fast and efficient processing. Hadoop is an open source solution developed by Google Corporation for large data processing. It supports variety of components and distributed file system. MapReduce, Pig, Hive are the components used for efficient development of software, together with Hadoop Distributed File System which is responsible for storing and processing large data with multiple nodes. The complete study observes that advance level of processing is required for large data scale, thereby to accomplish level of concert. In order to circumvent problem of privacy leakage and access maintenance, an elucidated security model has been developed for BigData application. This paper describes the security issue along with its solution. The proposed solution is implemented with Hadoop server in single node and multinode environment.
References

1. Interactions with Big Data Analytics, Danyel Fisher Microsoft Research | danyelf@microsoft.com.
2. https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html#Overview
4. The Hadoop Distributed File System Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler Yahoo! Sunnyvale, California USA {Shv, Hairong, SRadia, Chansler}@Yahoo-Inc.com
5. Introduction to MapReduce and Hadoop Matei Zaharia UC Berkeley RAD Lab matei@eecs.berkeley.edu
6. A Solution For Privacy Protection In MapReduce Quang Tran, Hiroyuki Sato Graduate School of Engineering, The University of Tokyo.
17. Wang, et.al, Federated MapReduce to Transparently Run Applications on Multicluster Environment, 2014 IEEE International Congress on Big Data
19. Access Control for Sensitive Data in Hadoop Yenumula B. Reddy Department of Computer Science Grambling State University, USA
20. Ron Rivest, Adi Shamir, Leonard Adlema. “RSA algorithm,” and approached by Avi Kak (kak@purdue.edu) February, 2016 Avinash Kak, Purdue University.
21. Marko Grobelnik “Introduction to big data” marko.grobelnik@ijs.si Jozef Stefan Institute Ljubljana, Slovenia.
22. Yenumula B Reddy “Access Control Mechanisms in Big Data Processing” Department of Computer Science Grambling State University, Grambling, LA 71245, USA.
29. Xianfeng Yang and Liming Lian, et. al., 2014, “A New Data Mining Algorithm based on MapReduce and Hadoop,” Xinxiang University, Xinxiang Henan, P.R.CHINA.
32. Privacy-preserving Anonymization of Setvalued Data Manolis Terrovitis Dept. of Computer Science University of Hong Kong rrovitis@cs.hku.hk Nikos Mamoulis Dept. of Computer Science University of Hong Kong nikos@cs.hku.hk Panos Kalnis Dept. of Computer Science National University of Singapore kalnis@comp.nus.edu.sg

Index Terms

Computer Science

Databases

Keywords

Big Data, Hadoop, Hive, Sqoop, MapReduce, RSA Cryptographic.