

Lecture Notes in Networks and Systems 179

Devendra Kumar Sharma
Le Hoang Son
Rohit Sharma
Korhan Cengiz *Editors*

Micro-Electronics and Telecommunication Engineering

Proceedings of 4th ICMETE 2020

 Springer

Read
first
21 | OriginalPaper | Chapter

An Efficient Model for High Availability Data in Hadoop 1.2.1

Authors: Anurag Bhatnagar, Venkatesh Gauri Shankar, Bali Devi, Nikhar Bhatnagar

Publisher: Springer Singapore

Published in: Micro-Electronics and Telecommunication Engineering

Abstract

Day by day the data is increasing enormously so storage of big data and analysis of big data require large storage as well as fast processing. Scaling can be done in two ways horizontal scaling and vertical scaling. Earlier horizontal scaling was approached to process big data. They tend to increase processing power each time the data is increased. Today, the vertical scaling approach is the best practice to handle big data. Privacy is also major concern, public fear of inapt use of their personal data. Apache Hadoop is one of the solution of big data. Hadoop below version 2 lack high availability in NameNode. We have found the solution of high availability of NameNode in Hadoop version 1. We have used centralized storage that is NFS server and shared that storage with two places. We have called it primary and secondary NameNode according to the vocabulary of Hadoop version 2. Primary NameNode is working as a master to the DataNodes while secondary NameNode will keep track of live status of primary NameNode. When primary NameNode goes down then secondary NameNode will replace its IP with primary NameNode's IP and then it will become master to the DataNodes. In this way, we can obtain high availability in older Hadoop version older than 2.