Abstract

In this paper, a new set of approaches is presented for analyzing datasets, which are known as Big data. The term 'big data' is often used incorrectly, especially when discussing computer science. It involves the gathering, organizing, and analyzing of large datasets to derive information from these steps. With these approaches, one is not able to do previously because there were challenges across one or more of the 3V's of big data: volume (too much data), variety (too complex data), and velocity (fast data). With the help of Big data, the performance of organizations is improved. As Hadoop is a prominent framework for big data implementation, this paper presents an overview of the architecture of Hadoop with details of its components.

References

Maryland at College Park.
6. Sameer Agarwal †, Barzan Mozafari X, Aurojit Panda †, Henry Milner †, Samuel Madden X, Ion Stoica "BlinkDB: Queries with Bounded Errors and Bounded Response Times on Very Large Data" Copyright © 2013 ACM 978-1-4503-1994-2/13/04
8. Shadi Ibrahim ⋆ Hai Jin Lu Lu "Handling Partitioning Skew in MapReduce using LEEN"

Index Terms

Computer Science Information Sciences

Keywords

Big data. HDFS