Abstract

In this paper a new set of approaches is present for analyzing dataset which is called as Big data. The big data is most wrongly interpreted term, when one talk about big data in terms of computers it is concept which explains about gathering, organizing, analyzing the data and from these steps to get information from this data. With this approach one is not able to do previously because there had challenges across one or more of the 3V’s of bigdata. Firstly volume which means too big data second is verity which has too complex data and velocity which has fast data. With the help of Bigdata performance of organization is improved. As HADOOP is a prominent framework for big data implementation, so paper present overall architecture of Hadoop with details of its components.

References

1. https://gavinbadcock.files.wordpress.com/2013/02/hadoopschema1.jpg
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 978-1-4503-19942/13/04


8. Shadi Ibrahim, HaiJin, LuLu “Handling Partitioning Skew in MapReduce using LEEN”


**Index Terms**

Computer Science
Information Sciences

**Keywords**

Big data, HDFS