Abstract

In the present world, where more and more users upload data to the internet, the overall size of data that need to be stored and analyzed exceeds the capacity of traditional storage and analysis techniques. Hence, it is necessary to introduce new and efficient methods for the analysis of Big Data in order to extract useful information from them. Big Data has tremendous importance in almost all areas such as education, healthcare, etc. Big data is defined by its three main characteristics which are high volume, high velocity and huge variety. Hadoop is provides an efficient platform for the analysis of Big Data. It consists of a distributed storage system, HDFS for the storage of large volumes of data and a programming framework,
MapReduce for the analysis of data.

References

- Divyakant Agrawal et al., "Challenges and Opportunities with Big Data," in Cyber Center Technical Reports, Purdue e-Pubs, Purdue University, 2011.

Index Terms
Keywords
Hdfs  Mapreduce