Abstract

Big data: Everyone just talking about Big data, but what is meant by big data actually? How is it changing the point of view in different fields such as researchers of the science or at companies, non-profits, governments, institutions, and other organizations are learning about big data that is nothing the world around them? Where this data is coming from, how is it being processed, and how are the results being stored and used for their future work? And why is open source so important to answering these questions? In this paper we will discuss all above points to clear actually what Big data means and how it deals in our day to day life. In today’s 21st century, the most important area is social media which shares, search and shares the information and generates huge of data everyday. So the importance of big data is more as millions and billions of peoples are using this media to share and store the information. Nowadays many projects are developing under social media, sensor data, stock exchange,
Transport data, and in the field of science where data is most important factor to store and retrieve. So we need new technology which is Big data and Hadoop to handle this huge amount of data which is not possible to handle by RDBMS. Big data has very basic important characteristics such as volume, variety, veracity and velocity. Big data handles the large amount of data with management, analysis, storage and processed data very fast within the time span. In this paper discusses, the important characteristics, types of data which is used in big data, what are the various sources of big data in our day to day life, introduction to big data and Hadoop with explanation, Structure of Hadoop core components, role of Namenode and data node, function of job tracker and task tracker, and Hadoop Ecosystem is explained in detail.

References

15. http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data
17. http://searchcloudcomputing.techtarget.com/definition/Hadoop


22. A. Foundation, “Yarn,”
https://hadoop.apache.org/docs/r0.23.0/hadoopyarn/hadoop-yarn-site/YARN.html


Index Terms

Computer Science
Information Systems

Keywords

Big Data, Hadoop, HDFS, MapReduce, Hadoop Ecosystem, Namenode, Datanode.