Abstract

This paper describes the concept of Big Data which is the collection of large data sets that cannot be processed by traditional computational techniques. Therefore, Hadoop technology is designed to process Big Data. Hadoop is the platform in businesses for Big Data processing. Hadoop is an open-source, Java-based programming framework that supports the processing and storage of extremely large data sets in a distributed computing environment. It helps Big Data analytics by overcoming the difficulties that are usually faced in handling Big Data. Hadoop can break down large computational problems into smaller tasks so that smaller elements can be analyzed economically and quickly[1]. Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for various kinds of data, enormous processing power, and the ability to handle virtually limitless concurrent tasks. All these parts are analyzed in parallel, and the results of the analysis are regrouped to produce the final output.

References


Index Terms

Computer Science

Information Sciences

Keywords

Big Data, Hadoop Architecture, Apache Hadoop, Mapreduce, Hadoop Ecosystem, Hadoop Distributed File System (HDFS).