Abstract

This paper describes the concept of Big Data, which refers to large datasets that cannot be processed by traditional computational techniques. Therefore, Hadoop technology was designed to handle Big Data. Hadoop is the platform used in businesses for Big Data processing. Hadoop is an open-source, Java-based programming framework that supports the processing and storage of extremely large datasets in a distributed computing environment. It helps in Big Data analytics by overcoming the difficulties faced in handling large datasets. Hadoop can break down large computational problems into smaller tasks, allowing smaller elements to 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).