Cricket Analysis Using HDFS and MapReduce APIs

For the fulfilment of Big Data Analytics J component

Abstract

This project aims to perform cricket data analytics using Hadoop Distributed File System (HDFS) and MapReduce APIs. It involves the analysis of a cricket cup played by countries, leveraging the power of Hadoop as a distributed file system. The project utilizes Java-based MapReduce APIs, specifically MapReduce-2, implemented through the Mapper-Reducer method controlled by a driver program. The project focuses on several key tasks for cricket analysis: Calculate Team-wise Total Score, Show Full Scorecard Team-wise, Count Total Number of Sixes in a Match, Count Total Number of Fours in a Match, Find Team-wise Top Three Scorers, Find Team-wise Top Three Highest Wicket Takers, Find the Name of Players who Hit Century and Find the Name of Players who Hit Half-Century.

The project leverages the scalability and parallel processing capabilities of Hadoop's MapReduce framework to efficiently handle large volumes of cricket data. By utilizing HDFS, the project ensures fault tolerance and reliability in storing and processing the data. The MapReduce APIs provide a flexible and scalable solution for performing complex analytics on cricket data.

Keywords: Cricket analysis, HDFS, MapReduce APIs, Java, Mapper-Reducer method, Team-wise total score, Full scorecard, Sixes, Fours, Top scorers, Highest wicket takers, Century, Half-century.

Group Members:

1. Rajat Singh 22MCA0139 Ph: 9038897689
2. Sambit Basu 22MCA0240
3. Arjun Mitra 22MCA0167
4. Srijan Paria 22MCA0266