

BDA - Experiment 1

Title: To study Hadoop Ecosystem and to demonstrate Basic Hadoop Commands

Objective: Explore Hadoop Ecosystem and its components and demonstrate Hadoop commands for effective data management and preprocessing

Outcome: Successfully install Hadoop Ecosystem and execute Basic Hadoop Commands on it.

Problem Statement: Install Cloudera CDP and learn how to use Hadoop commands for managing and preprocessing Big Data.

Theory: Hadoop is an open-source framework designed for processing and managing large volumes of data across distributed computing clusters. It was initially developed by the Apache Software Foundation and is widely used in the field of big data analytics. Hadoop is used to store and process massive amounts of data in a scalable and reliable manner. It is well-suited for handling unstructured or semi-structured data, such as text, images, videos and log files. Hadoop is employed for tasks like batch processing, data warehousing, data transformation, and more. Its commonly used in industries like finance, healthcare, e-commerce and social media to extract valuable insights from vast datasets.

Conclusion: learned about Hadoop's important features and structures and practiced basic Hadoop commands, like making folders, moving files and checking data. These actions showed us how Hadoop can handle big data tasks effectively.