Unit -4 Spark (Work Report)

Submitted By:

    2021UCA1536 - Aparna Srivastava

    2021UCA1804 - Pulkit Batra

    2021UCA1814 - Muskan Sinha

In Unit 4, we delve into Spark, a versatile distributed computing framework designed for speed and ease of use. Spark enables processing of large-scale data sets across clusters of computers, providing capabilities for data manipulation, machine learning, and real-time analytics. Throughout this unit, we will learn how to harness the power of Spark to handle big data efficiently, perform complex data transformations, and implement advanced analytics algorithms to derive valuable insights from massive datasets.

[1. RDD Exploration](#)

[2.1 Spark with Sales Data](#)

[2.2 Spark with BigData (Google Play Store Data)](#)