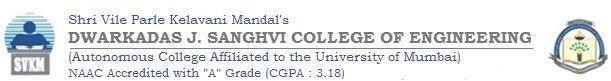
`

Department of Information Technology A.Y.

2024-2025

Class: TY BTech-IT, Semester: VI Subject: Big Data Lab

NAME: Anish Sharma SAP:60003220045

# Experiment – 11

1. Aim: To implement Big Data Technologies for real world applications.

Procedure:

CODE:

from pyspark.sql import SparkSession from pyspark.sql.functions import col

# Initialize the Spark session spark = SparkSession.builder.appName("ECommerceAnalysis").getOrCreate()

# Sample data (assuming this is a CSV file with transactional data) data = [

(1, 101, 2, 20, "2025-04-10 10:15:00"),

(2, 102, 1, 50, "2025-04-10 10:20:00"),

(3, 103, 4, 15, "2025-04-10 10:25:00"),

(4, 104, 1, 30, "2025-04-10 10:30:00"),

(5, 105, 3, 25, "2025-04-10 10:35:00")

]

# Define schema

columns = ["TransactionID", "ProductID", "Quantity", "Price", "Timestamp"]

# Create DataFrame

df = spark.createDataFrame(data, columns)

# Show the loaded data df.show()

# Filter transactions where quantity is greater than 2

filtered\_df = df.filter(col("Quantity") > 2)

# Add a new column for the total value (Quantity \* Price)

transformed\_df = filtered\_df.withColumn("TotalValue", col("Quantity") \* col("Price"))

# Show the transformed data transformed\_df.show()

# Aggregate total sales per ProductID sales\_per\_product\_df =

transformed\_df.groupBy("ProductID").sum("TotalValue").withColumnRenamed("sum(TotalValue)"

, "TotalSales")

# Show the aggregated results sales\_per\_product\_df.show()

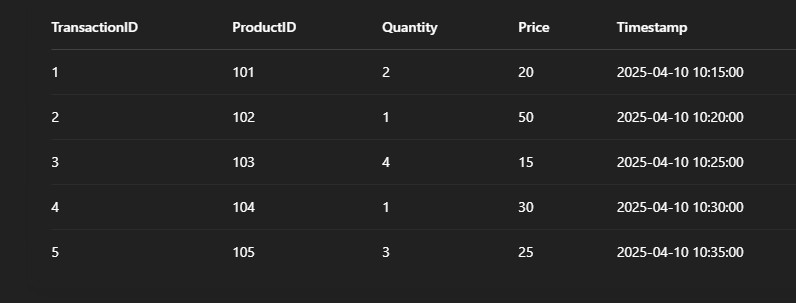
# Optional: Write the result to a CSV file

# sales\_per\_product\_df.write.csv("total\_sales\_per\_product.csv", header=True)

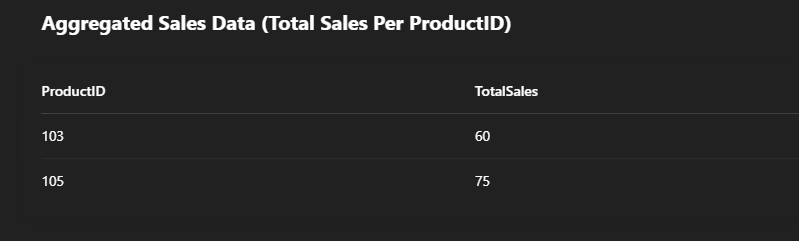
# Stop the Spark session spark.stop()

1. Requirements: PC, Internet

OUTPUT:







1. Conclusion: Thus, in this experiment, we implemented Big Data technologies for real world applications.