**Mock Test 1 - Big Data Question no. 10**

**10.** Retrieve the top 5 customers who have spent the highest total amount on purchases.

**Sol.**

from pyspark.sql import SparkSession

from pyspark.sql.functions import desc

# Create SparkSession

spark = SparkSession.builder \

.appName("TopSpendingCustomers") \

.getOrCreate()

# Assuming you already have a DataFrame named "purchases" containing the "Purchases" table

# Group purchases by customer and sum the total amount spent

total\_spent\_df = purchases.groupBy("customer\_id") \

.sum("amount") \

.withColumnRenamed("sum(amount)", "total\_amount")

# Sort the DataFrame by total amount spent in descending order

sorted\_df = total\_spent\_df.orderBy(desc("total\_amount"))

# Retrieve the top 5 customers

top\_5\_customers = sorted\_df.limit(5)

# Show the result

top\_5\_customers.show()

# Stop the SparkSession

spark.stop()