**Mock Test 1 - Big Data Question no. 2**

**2.**Given a PySpark DataFrame named "sales\_data" with columns "product\_name" and "revenue", write a code to calculate the total revenue for each product and display the result in descending order.

**Sol.**

from pyspark.sql import SparkSession

from pyspark.sql.functions import sum, desc

spark = SparkSession.builder.getOrCreate()

# Assuming sales\_data is the PySpark DataFrame with columns "product\_name" and "revenue"

# If sales\_data is not already a DataFrame, you can create it from a list of tuples or a Pandas DataFrame using spark.createDataFrame()

total\_revenue\_df = sales\_data.groupBy("product\_name").agg(sum("revenue").alias("total\_revenue"))

# Sort the DataFrame by total revenue in descending order

sorted\_df = total\_revenue\_df.orderBy(desc("total\_revenue"))

sorted\_df.show()