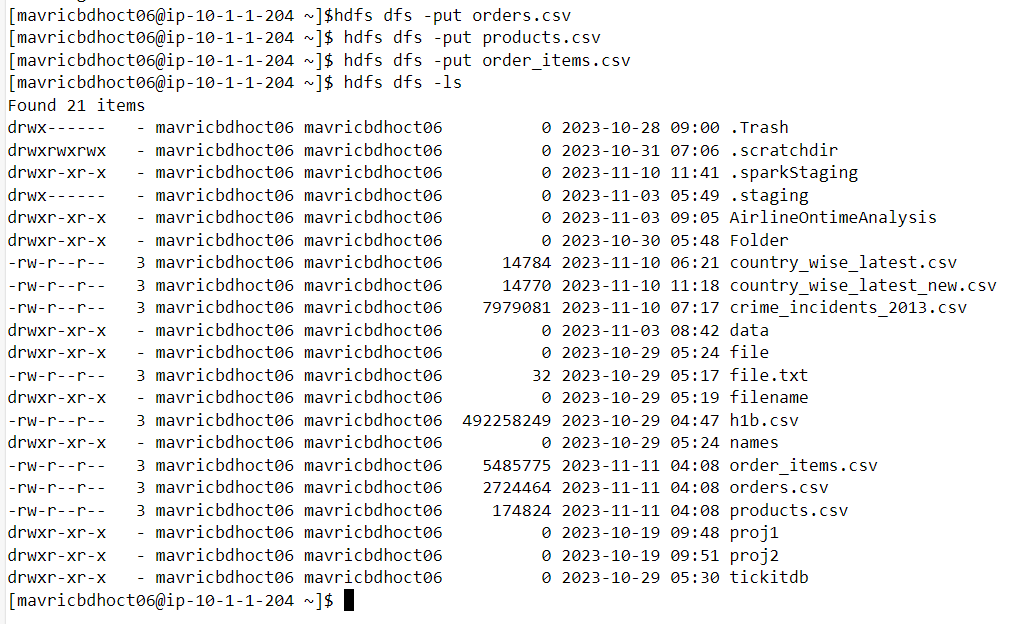
Order Data analysis (Joins)



1. Problem Statement: What is the daily product revenue for CLOSED or COMPLETE orders?

A screenshot of a computer

Description automatically generated

1. Load the required data in to DF like categories, customer,departments,order\_items,orders and products

orderitemsDF=spark.read.option("header",True).option("inferschema",True).csv("order\_items.csv")

orderitemsDF.createOrReplaceTempView("orderitems")

orderitemsDF.printSchema()

ordersDF=spark.read.option("header",True).option("inferschema",True).csv("orders.csv")

ordersDF.createOrReplaceTempView("orders")

ordersDF.printSchema()

productsDF=spark.read.option("header",True).option("inferschema",True).csv("products.csv")

productsDF.createOrReplaceTempView("products")

productsDF.printSchema()

A close-up of a document

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a list of items

Description automatically generated

1. Get the count for each order status

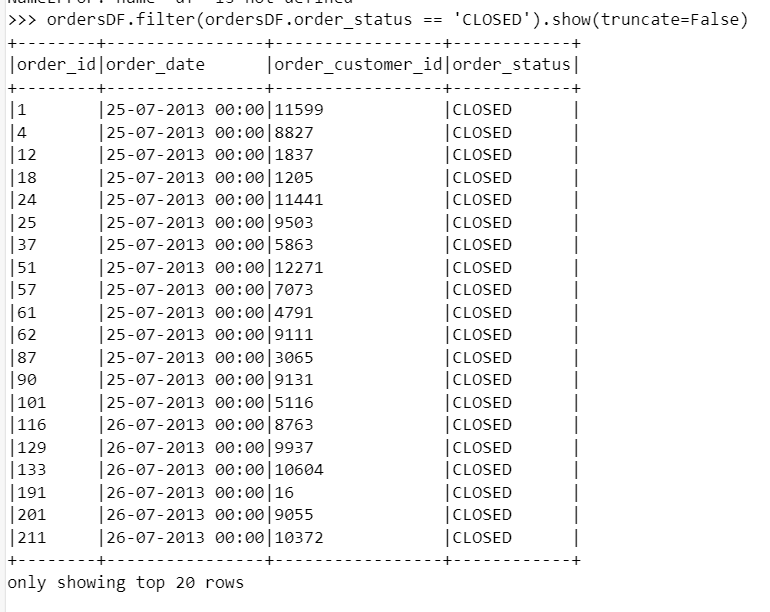
A screen shot of a computer

Description automatically generated

4. Filter only COMPLETE or CLOSED orders

A screenshot of a computer

Description automatically generated

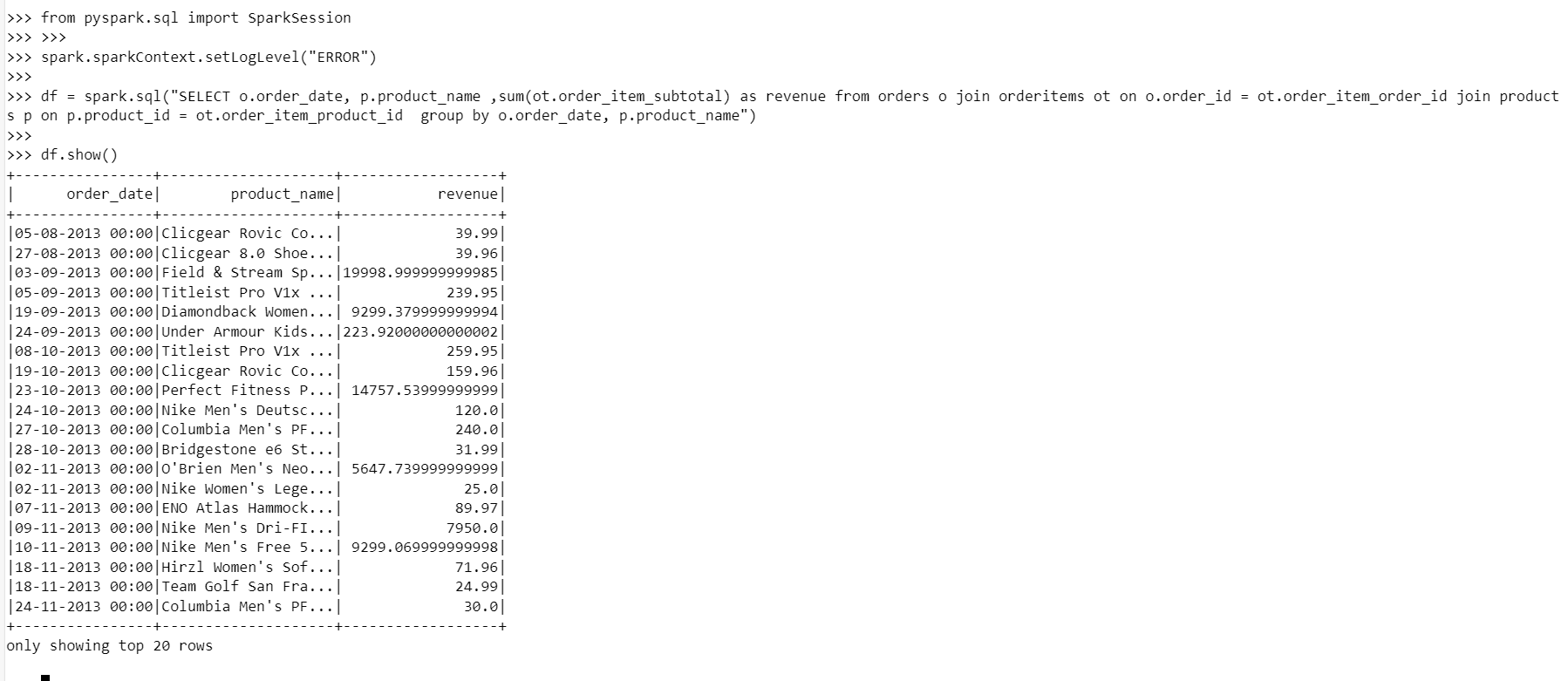


5. Join the products , order\_items and orders tables and calculate daily product revenue

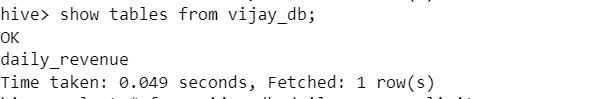
A white screen with black text

Description automatically generated

6. Write the data in to the table Daily product revenue in Hive







A screenshot of a computer

Description automatically generated