# **Customer Orders Dataset Analysis**

The dataset consists of customer order records with the following columns:

|  |  |  |
| --- | --- | --- |
| **Column** | **Data Type** | **Description** |
| customer\_id | int | Unique ID for each customer |
| name | chararray | Name of the customer |
| age | int | Age of the customer |
| location | chararray | Location of the customer |
| order\_id | int | Unique order ID |
| order\_date | chararray | Date the order was placed (YYYY-MM-DD) |
| amount | float | Order amount in dollars |

### **Sample Data**

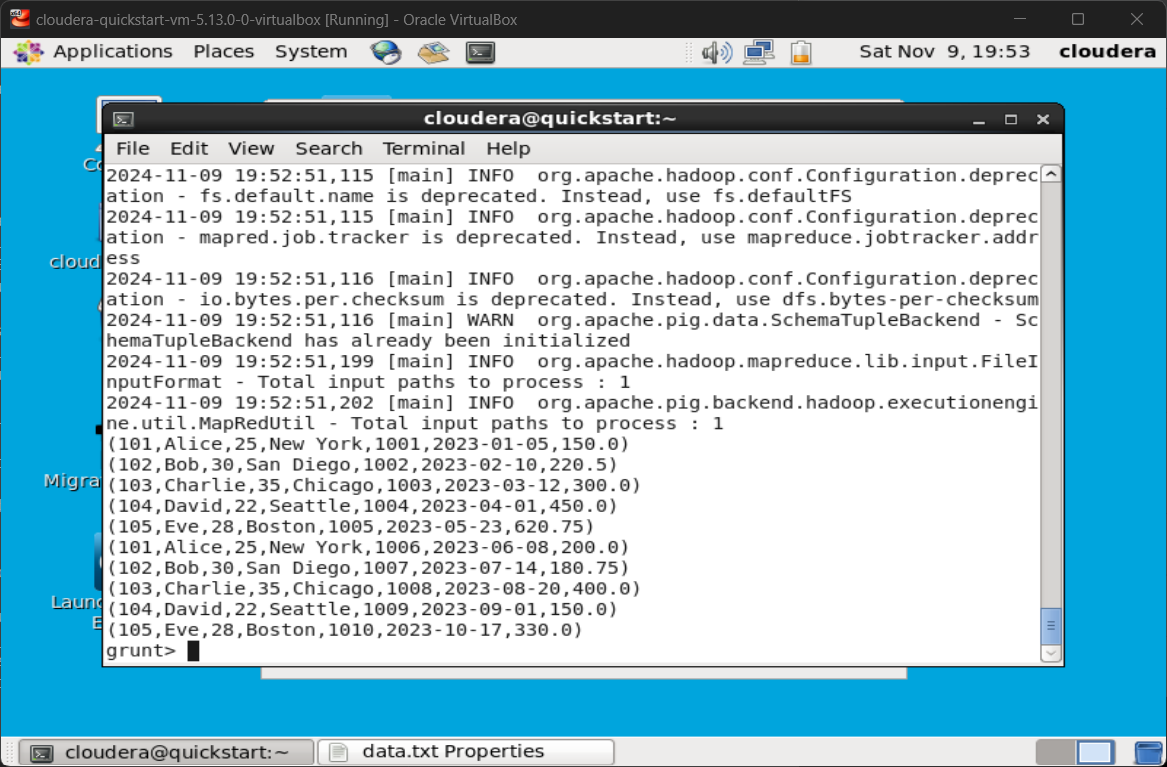
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **customer\_id** | **name** | **age** | **location** | **order\_id** | **order\_date** | **amount** |
| 101 | Alice | 25 | New York | 1001 | 2023-01-05 | 150.00 |
| 102 | Bob | 30 | San Diego | 1002 | 2023-02-10 | 220.50 |
| ... | ... | ... | ... | ... | ... | ... |

**PIG COMMANDS AND QUERIES**

### **1. Filtering Customers by Age**

**Scenario: Analyze customers who are adults (age 18 and above).**

* **Query: Load the dataset and filter out customers with age < 18.**

****

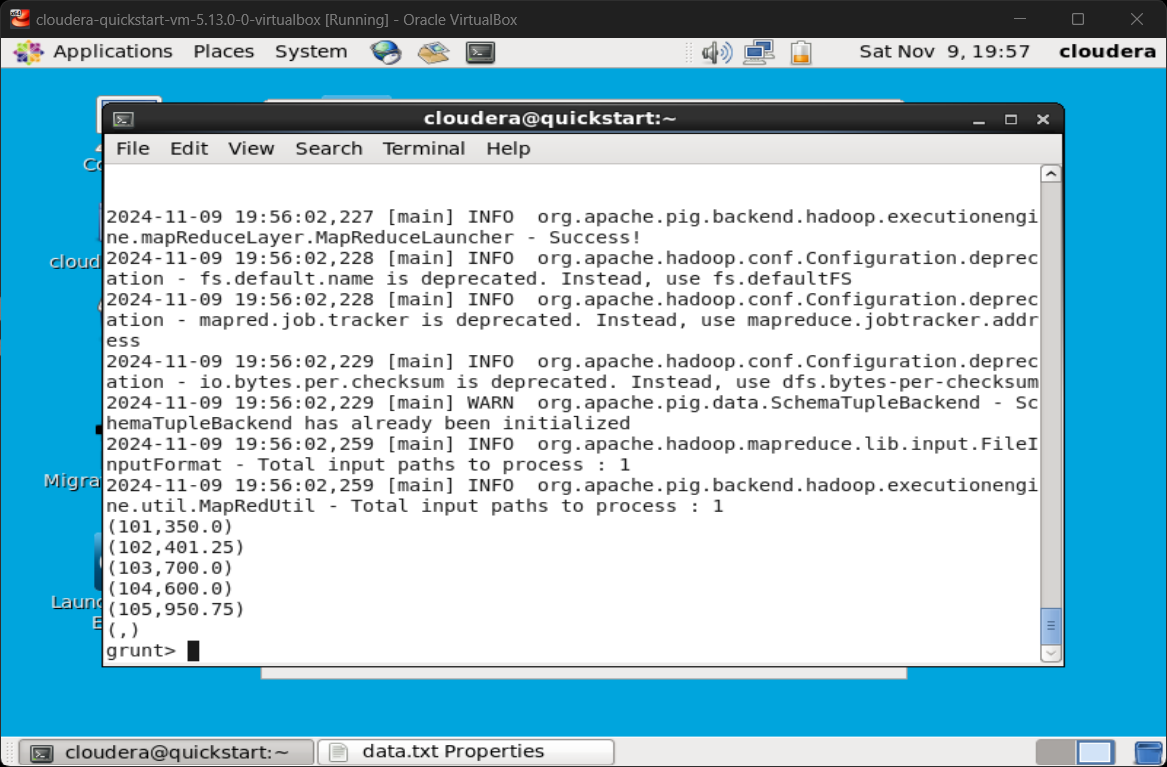
### 

### 

### **2. Grouping by Customer and Calculating Total Sales**

**Scenario: Calculate the total order amount for each customer.**

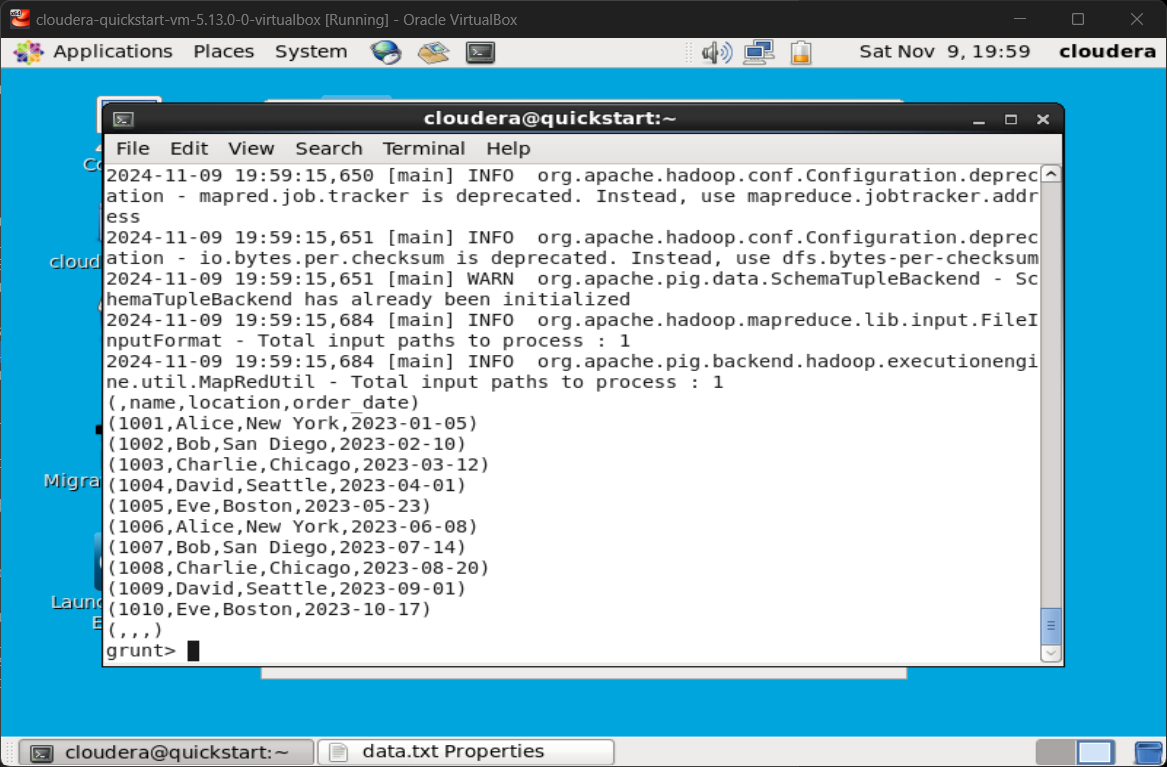
* **Query: Group the data by customer\_id and calculate the total sales (sum of amount) for each customer.**

****

### **3. Joining Orders with Customer Information**

**Scenario: Combine customer details with order information for a detailed view.**

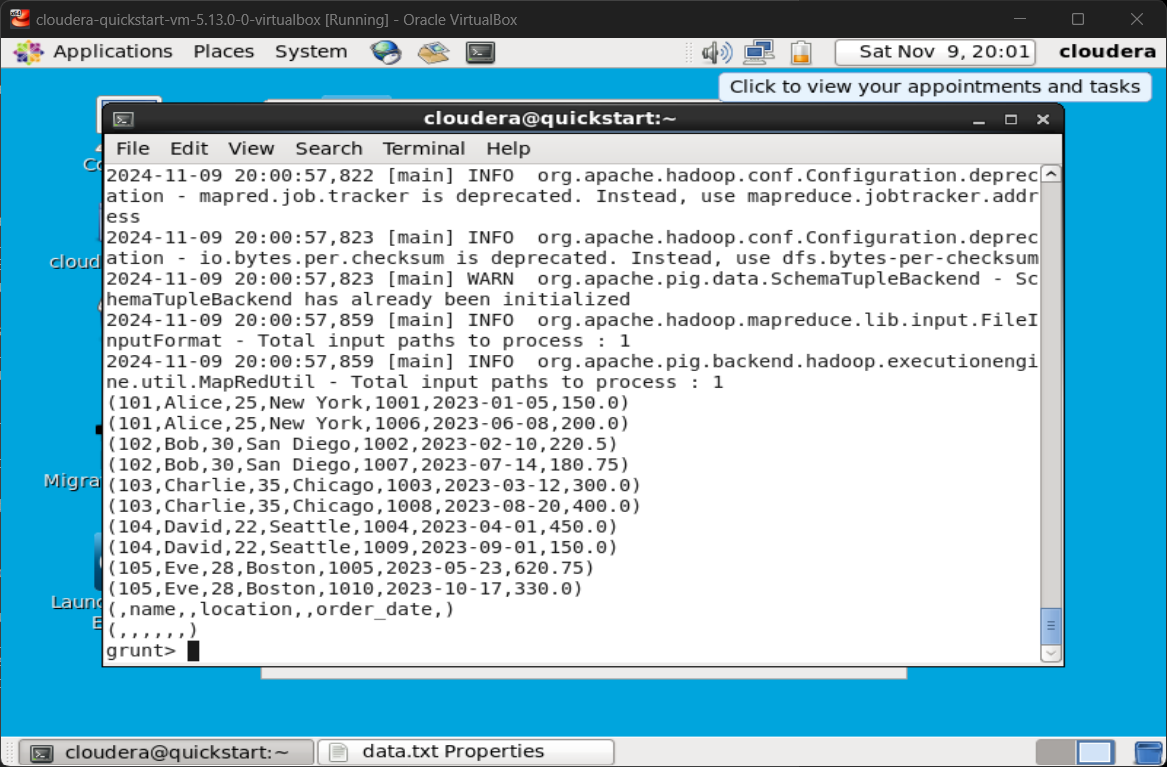
* **Query: Join the dataset with itself on customer\_id to display order\_id, name, location, and order\_date.**

****

### **4. Removing Duplicate Orders**

**Scenario: Remove duplicate records due to potential system errors.**

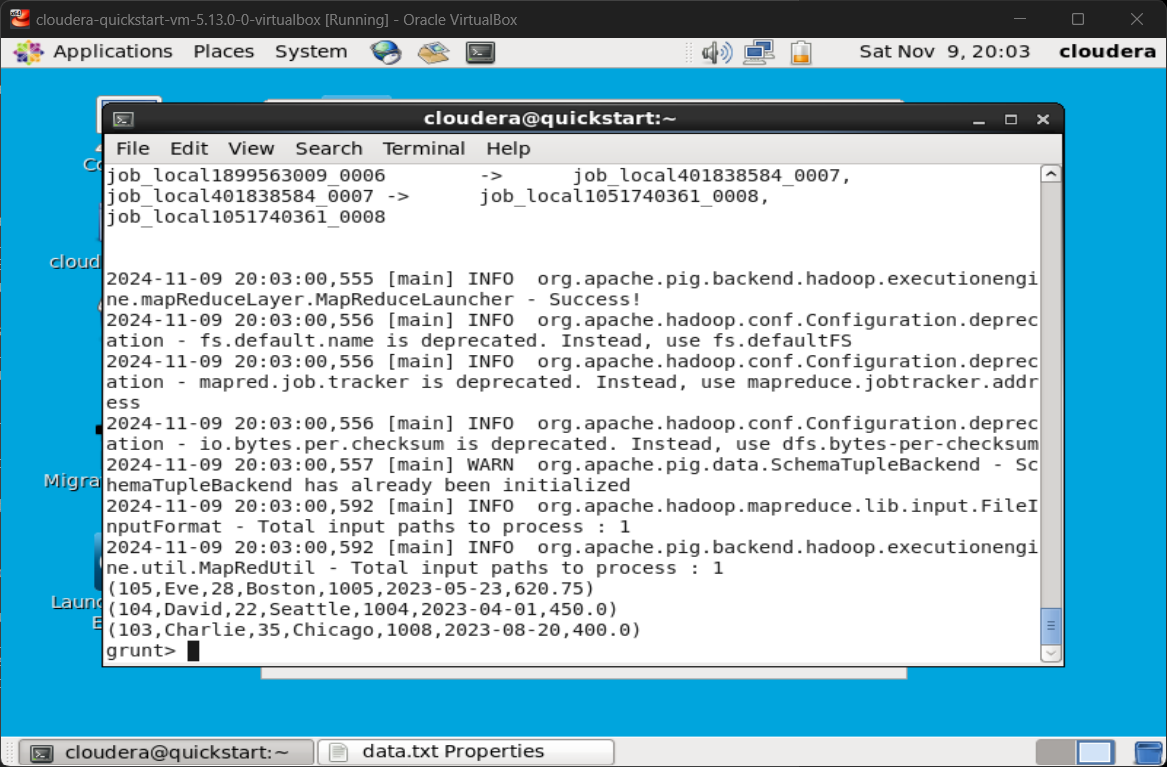
* **Query: Load the dataset and remove any duplicate entries based on order\_id.**

****

### **5. Sorting Orders by Amount**

**Scenario: Identify the highest value orders.**

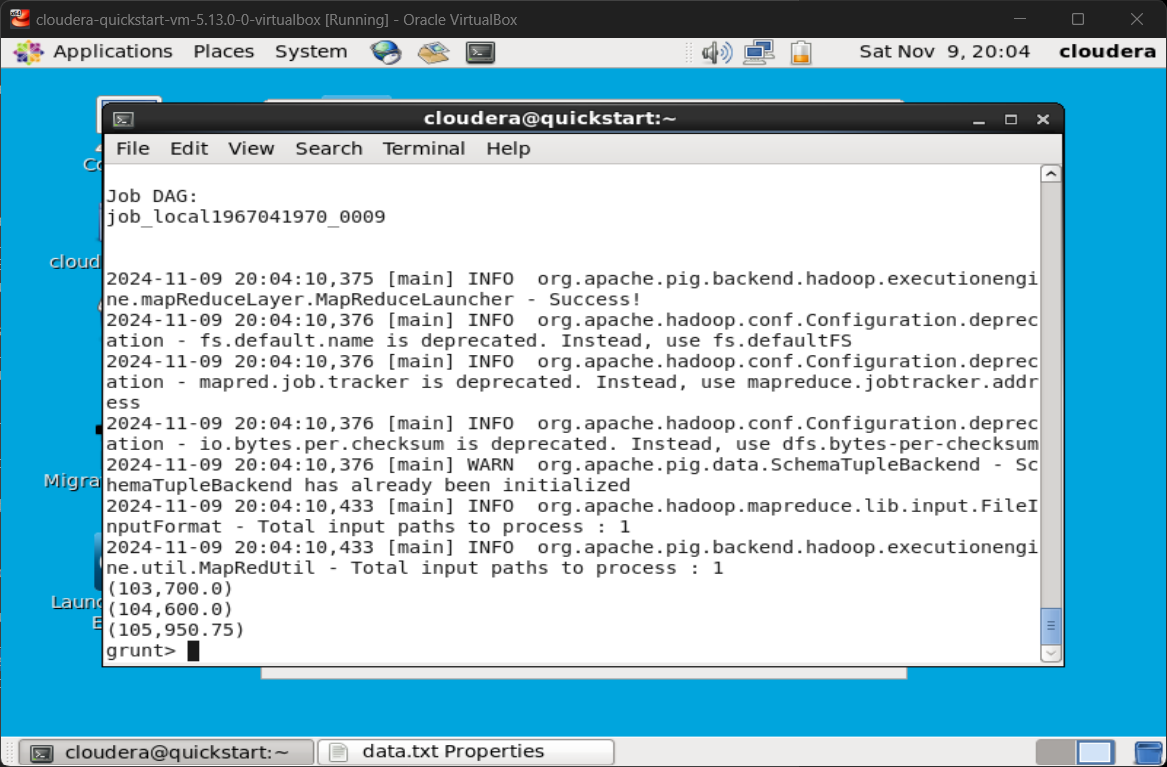
* **Query: Load the dataset, sort by amount in descending order, and display the top 3 highest orders.**

****

### **6. Aggregating and Filtering Data (Total Purchases Above $500)**

**Scenario: Identify customers whose total purchases exceed $500.**

* **Query: Group orders by customer\_id, calculate the total purchase amount, and filter for customers who spent more than $500.**

****