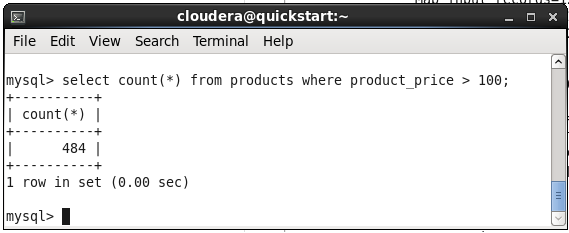
# Aula 2 – Ingestão de Dados com Sqoop

**334409 - Charles Cavalcante**

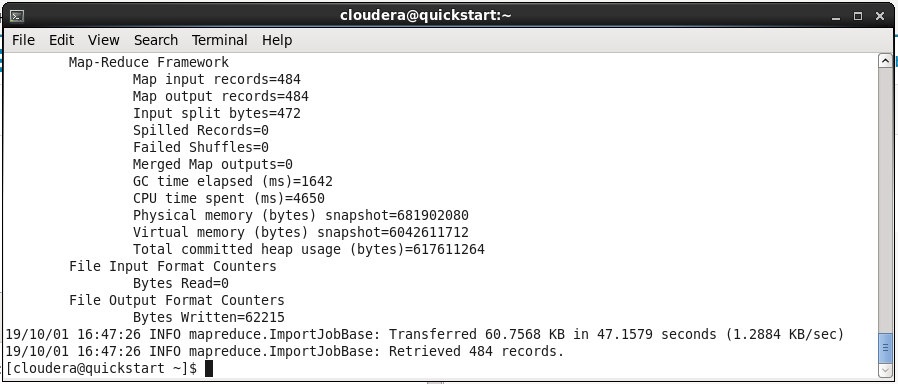
# **Exercício 1: Reconhecimento do Ambiente e Importação – MySQL 🡺 HDFS**

sqoop import --connect jdbc:mysql://quickstart:3306/retail\_db   
--table products --warehouse-dir /user/mysql --delete-target-dir   
--username root --password cloudera --where "product\_price > 100"

# Número de registros no MySQL

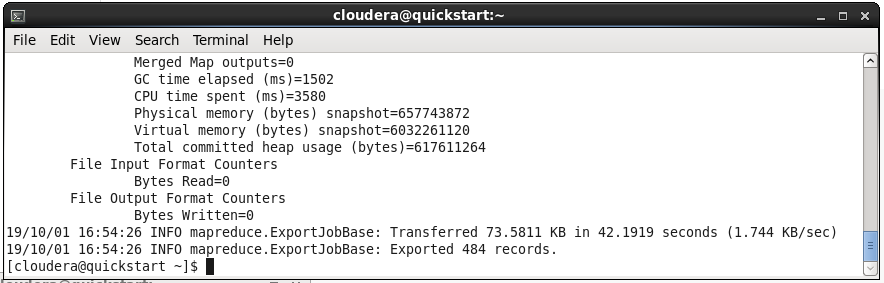


# Saída do sqoop

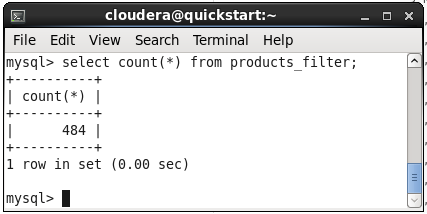


# **Exercício 2: Exportação – HDFS 🡺 MySQL**

# Saída do sqoop



# Número de registros no MySQL



# **Exercício 3: Otimização do processo de importação**

### Importação da tabela “customers” (52.056 segundos)

### 

### Importação da tabela “customers” utilizando a opção “split-by” (48.9813 segundos)

### 

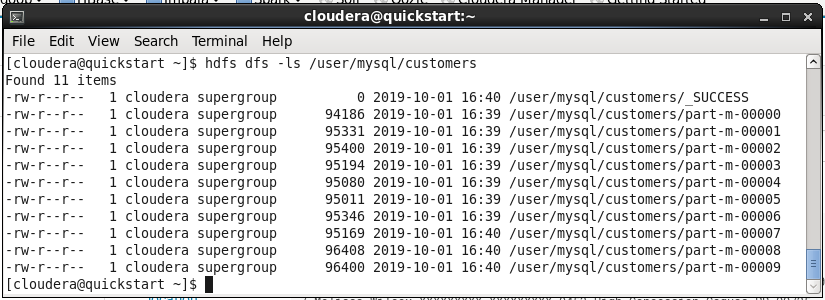
### Foram gerados 4 mappers

### 

### Importação da tabela “customers” utilizando a opção num-mappers=10 (98.7485 segundos)

### 

### Foram gerados 10 mappers

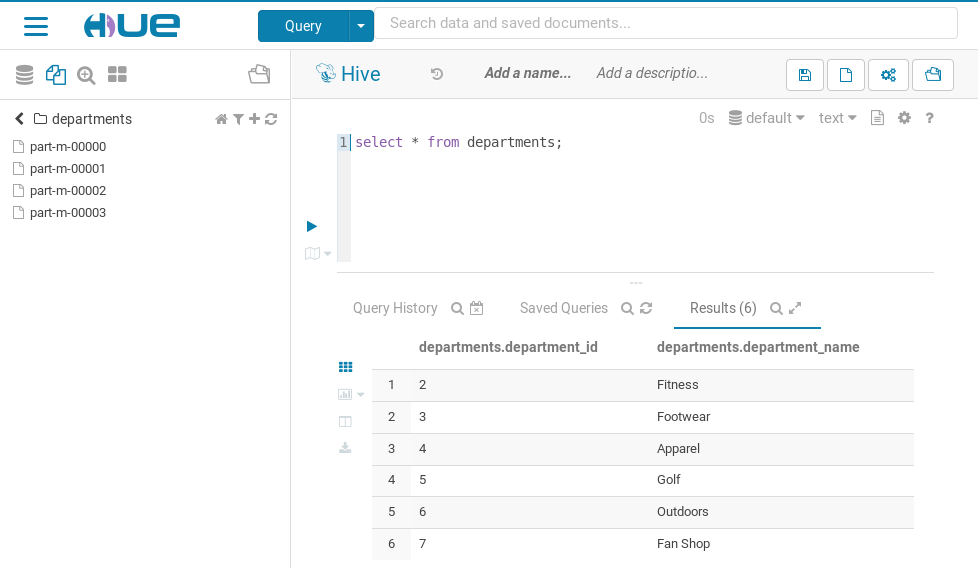


# **Exercício 4: Importação com camada de metadados no HIVE**

# Saída do sqoop

# 

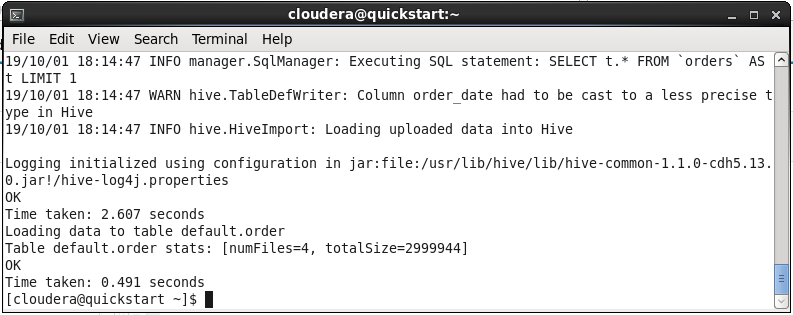
# Verificação no Hive



# **Desafio**

# Importação da tabela orders para o Hive com split pela chave da tabela (order\_id)

sqoop import --connect jdbc:mysql://quickstart:3306/retail\_db   
--table orders --target-dir /user/mysql/order\_groups   
--hive-import --hive-table orders --delete-target-dir --split-by order\_id   
--username root --password cloudera



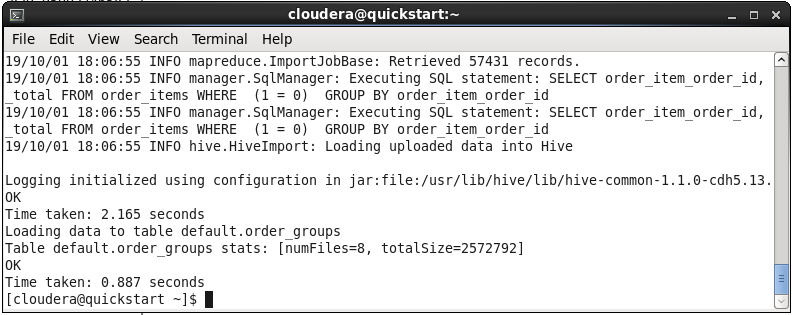
# Importação via sqoop de uma tabela nova chamada orders\_group, oriunda de uma agregação feita na order\_items, sumarizando o valor da venda. Faça o split pela chave estrangeira order\_item\_order\_id

#### **query que será utilizada:**

SELECT order\_item\_order\_id, SUM(order\_item\_subtotal) AS order\_item\_total   
FROM order\_items   
GROUP BY order\_item\_order\_id;

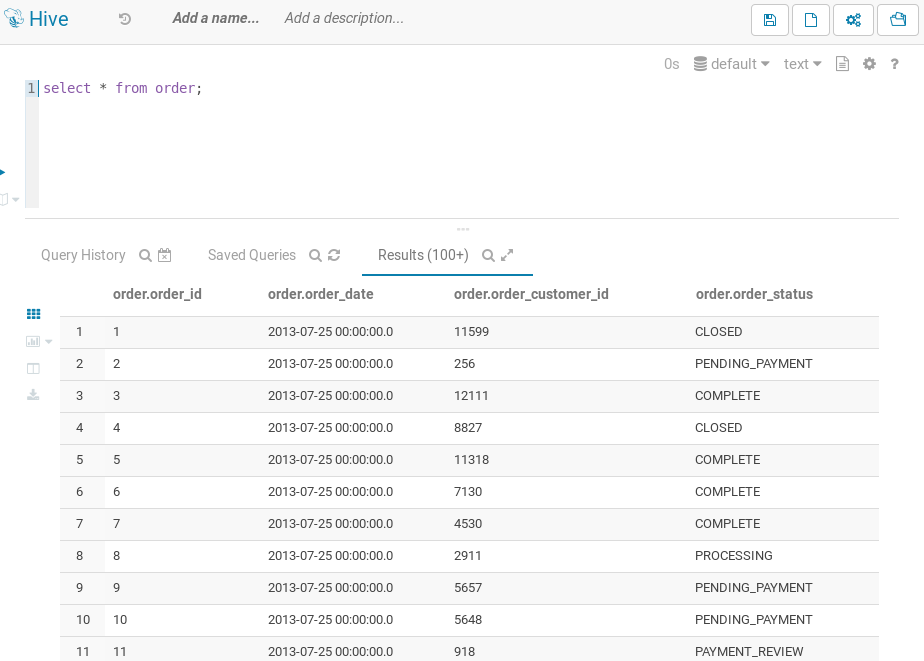
#### **comando no sqoop**

sqoop import --connect jdbc:mysql://quickstart:3306/retail\_db --query 'SELECT order\_item\_order\_id, SUM(order\_item\_subtotal) AS order\_item\_total FROM order\_items WHERE $CONDITIONS GROUP BY order\_item\_order\_id' --hive-import --hive-table order\_groups --delete-target-dir --split-by order\_item\_order\_id --username root --password cloudera



# Resultados Esperados

# 1) Tabela orders



# 2) Tabela orders\_group

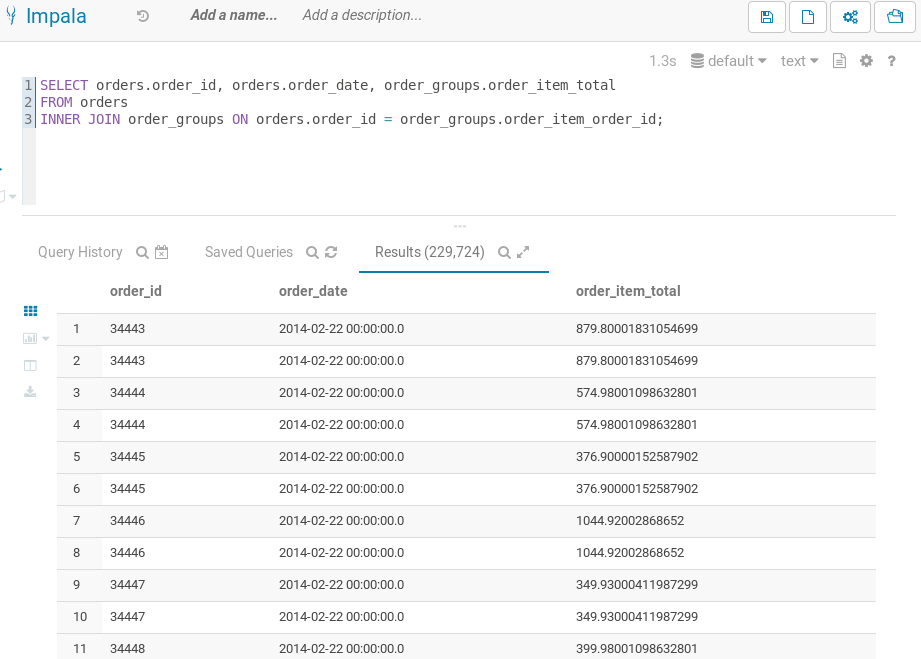
# 

# 3) Pedidos agrupados com a data em que ocorreram

SELECT orders.order\_id, orders.order\_date, order\_groups.order\_item\_total

FROM orders

INNER JOIN order\_groups ON orders.order\_id=order\_groups.order\_item\_order\_id;



P.S. a consulta no Impala demorou 1.2 segundos, a mesma consulta no Hive demorou 38.4 segundos.