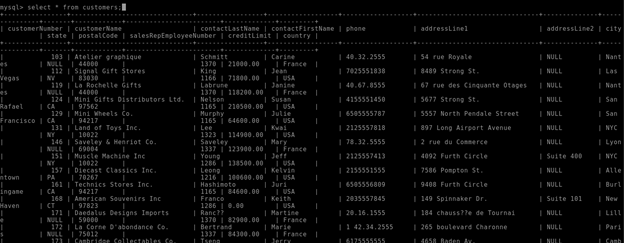
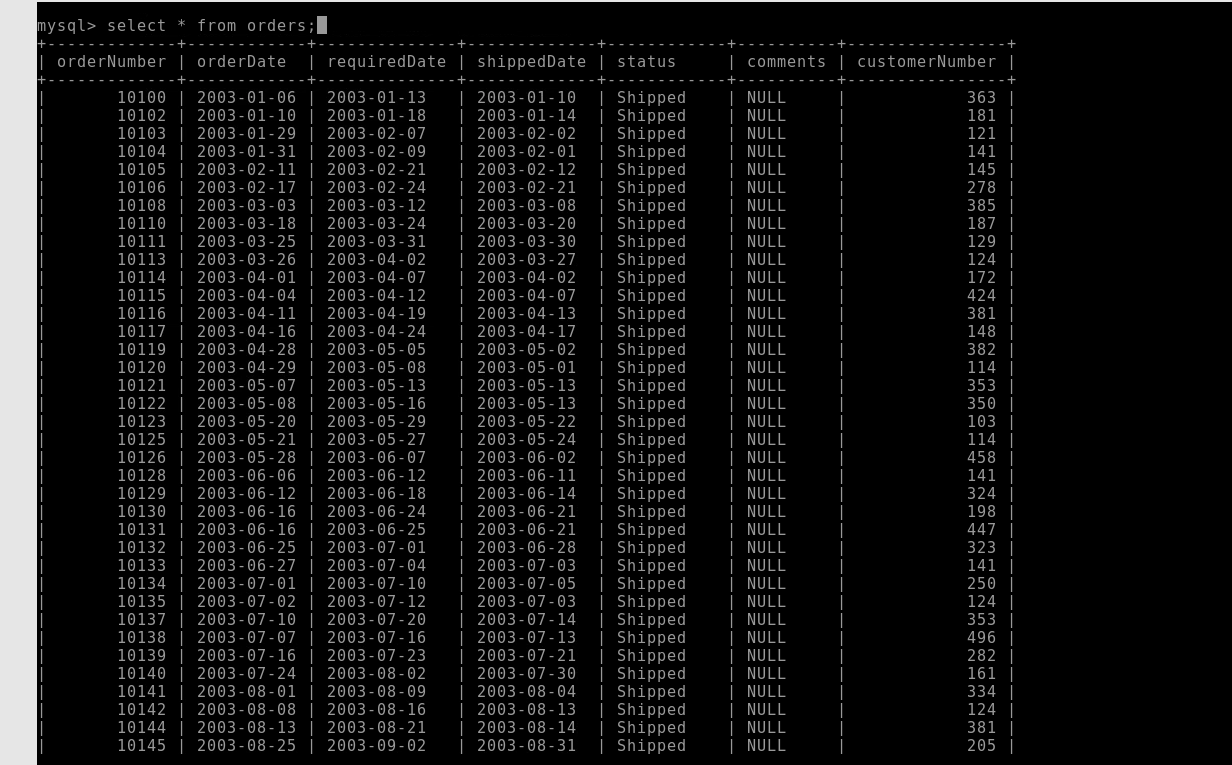
1. Created table with usa,uk and France records

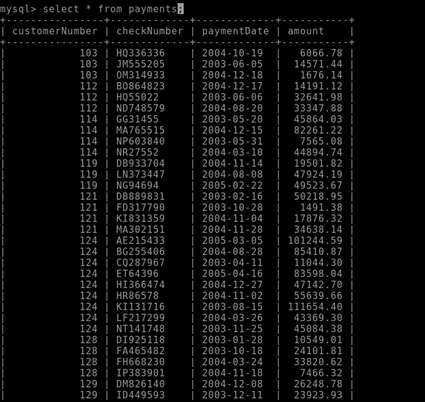
Select \* from customers;



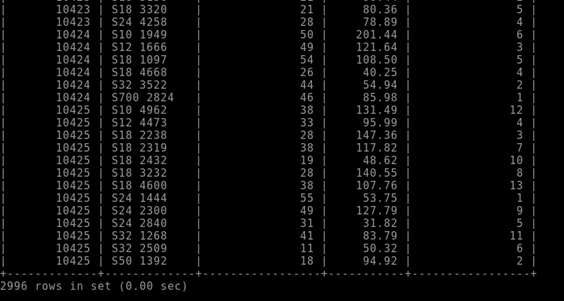
2.created orders table-(select \* from orders)



Created payments table(select \* from payments)



Created a order details table

Select \* from orderdetails;

3.what is sqoop

Apache Sqoop is designed to efficiently transfer enormous amount of data between Apache Hadoop and structured datastores such as relational databases. It helps to offload certain tasks, such as ETL processing at a much lower cost.

1.sqoop list-tables --connect jdbc:mysql://cdhserver/mysql --username root -password labuserbdh-

used to list all tables present

2. sqoop eval --connect jdbc:mysql://cdhserver/ --username root -password labuserbdh --query "SELCT \* from test.employees;" (used to execute a query using sqoop we use eval to do it)

3. sqoop import --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table employees --delete-target-dir --target-dir 'tmp/t1/sqoop' (used to have a target directory and use it with outr local or external table

4. sqoop import --connect jdbc:mysql://cdhserver/test --username root --password labuserbdh --table employees --delete-target-dir --num-mappers 2 --where "salary>15000" --target-dir '/tmp/t1/sqoop/where' (we use this this import to divide the columns with respect to mappers with a condition)

4.moving all tables into Hadoop

sqoop import-all-tables --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table customers --delete-target-dir --target-dir 'tmp/Vivek/d'

here we have imported all four tables into test database

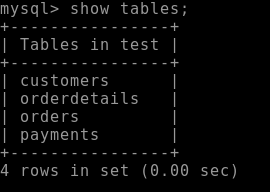
Text

Description automatically generated

Text

Description automatically generated

sqoop import-all-tables --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table orders --delete-target-dir --target-dir 'tmp/Vivek/d1'



sqoop import-all-tables --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table orders --delete-target-dir --target-dir 'tmp/Vivek/d2'

Text

Description automatically generated

5.sql to hive table moved all the 4 tables

1.customers

2.orders

3.orderdetails

4.payments

1.

sqoop import --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table customers --hive-table test.vvk --create-hive

-table --hive-import-m 1

2

sqoop import --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table orders--hive-table test.vvk --create-hive

-table --hive-import-m 1

3

sqoop import --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table orderdetails --hive-table test.vvk --create-hive

-table --hive-import-m 1

4

sqoop import --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table payments--hive-table test.vvk --create-hive

-table --hive-import-m 1

6.what is hbase

Hbase framework(NO sql database

Column oriented datastore known as hadoop database

Supports random real time crud operations (unlike hdfs)

Bigdata derived from gfs(google file systems)

Hbase use zookeeper extensively for region assignment

Hbase can manage zookeper daemons for you or you can install/manage them seperate

7. hbase vs rdbms

The goal of HBase is to store and process large amounts of data, specifically to handle large amounts of data consisting of thousands of rows and columns using only standard hardware configurations.

Hbase is highly scalable

Hbase is no sql type

Hbase is dynamic in nature

Hbase is flexible

Rdbms is not scalable

Rdbms is sql type

Rdbms is static in nature

rdbms cannot work as efficient as hbase

8. What is streaming

streaming (process of bidirectional transfer between clusters )is a utility that comes with the Hadoop distribution. This utility allows you to create and run Map/Reduce jobs with any executable or script as the mapper and/or the reducer.

It has mappers reducers for example Netflix

2.youtube

3.rapido (when we book a cab it will show the location and the path of captain which is an example of streaming

9.kafka components

Kafka is highly scalable consumption model it helps with the server it has producers consumers

Components of kafka

1.Broker

Single kafka server

The broker receives messages from procedures, assigns offsets to them

And commits the messages to storage on disk

2.PRODUCER

Creates messages

Maintains message key to deliver message to particular partitions

3.CONSUMER

Consume messages

Subscribe messages

Inside the broker partition and topic

Partition horizontal division of topics

Scaled across multiple brokers/nodes

4.Topic

Category of messages ex; table/folder

Scaled across all brokers/nodes\

5.REPLICAS

Partitions in the topics are replicated

10.kafka architecture

