Lab 2 Evaluations

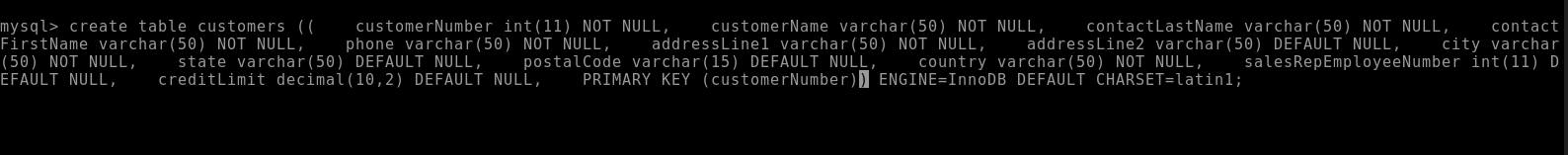
==================

==================

Create 4 tables in mysql databases as mentioned below,

=======================================================

1.Create one customer table and whith (USA, UK, France data in mysql)



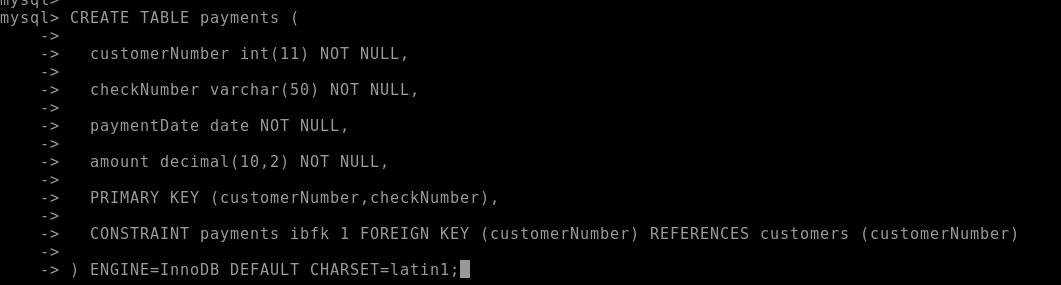
2.Create Order and order details andpayment details tables in mysql

Text

Description automatically generated

Text

Description automatically generated



3.What is Sqoop and list all the commands and its purpose in sqoop

Sqoop is a framework used to load tables from sql to Hadoop and vice versa.

Commands: codegen

Import : used to import tables to Hadoop

Export : export tables from Hadoop

List\_tables: lists all tables in database

List\_databases : lists all databases

Eval : used to perform particular queries from sql table

Import\_all\_tables: imports all the tables in database to hadoop

4.Move all the four mysql tables into hadoop file to hadoop load format and Hive table (any 5 customers details maybe same country or maynot..)









5.Move all the four mysql tables into hadoop Mysql table to hive table directly with (any 5 customers details maybe same country or maynot.. )

sqoop import --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table customers --hive-table exam2.customers1 --create-hive-table --hive-import --m 1

sqoop import --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table orders --hive-table exam2.orders1 --create-hive-table --hive-import --m 1

sqoop import --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table ordersdetails --hive-table exam2.ordersdetails1 --create-hive-table --hive-import --m 1

sqoop import --connect jdbc:mysql://cdhserver/test --username root -password labuserbdh --table payments --hive-table exam2.payments1 --create-hive-table --hive-import --m 1

6.What is Hbase and it advantages

Hbase is nosql database that is built on top of HDFS. Compared to rdbms has very simple and limited api.

characterstics:-

Horizontally scalable

Strongly consistent reads and writes

Automatic fail-over

Simple java api

Integration with map-reduce framework

7.What is Hbase differece beween rdbms and habase

Hbase is good for large amounts of data and compared to rdbms has very simple and limited api.

8.What is Streaming

9.What is Kafka and its component

kafka is a leading general purpose publish-subscribe distributed messaging system written in scala,which offers strong durability,scalability and fault-tolerance support.It is designed for handling of real time activity stream data such as logs,metrics collection. An apache project intially devloped at linkden at the year 2007. It is not specifically designed for hadoop rather hadoop ecosystem is just one of its possible consumers.

components:-

Broker(channel):A single kafka server. It receives message from producers assigns offset to them and commits the message to storage on disk. It serves consumer responding with the messages based on the partitions

Partition: horizontal division of topics, scaled across multiple broker folders.

Topics: actual message. Category of messages (tables/folder), spread across all broker/nodes

producer(source): creates messages and maintains message key to the particular partition

10.Draw the Kakfa and Hbase Architechture.