Problem1

Instructions:

Connect to mySQL database using sqoop, import all orders that have order_status as COMPLETE

Data Description:

A mysql instance is running on the localhost. In that instance, you will find orders table that contains order's data.

> Installation: localhost

> Database name: retail_db

> Table name: Orders

> Username: root

> Password: cloudera

Output Requirement:

Place the customer's files in HDFS directory "/user/cloudera/problem1/orders/parquetdata"

Use parquet format with tab delimiter and snappy compression.

Null values are represented as -1 for numbers and "NA" for string

Problem2

Instructions:

Connect to mySQL database using sqoop, import all customers that lives in 'CA' state.

Data Description:

A mysql instance is running on the localhost node. In that instance, you will find customers table that contains customer's data.

> Installation: localhost

> Database name: retail_db

> Table name: Customers

> Username: root

> Password: cloudera

Output Requirement:

Place the customers files in HDFS directory "/user/cloudera/problem1/customers_selected/avrodata"

Use avro format and snappy compression.

Load only customer_id,customer_fname,customer_lname,customer_state

Problem3

Instructions:

Connect to mySQL database using sqoop, import all customers whose street name contains "Plaza" .

Data Description:

A mysql instance is running on the localhost node.In that instance you will find customers table that contains customers data.

> Installation : localhost

> Database name: retail_db

> Table name: Customers

> Username: root

> Password: cloudera

Output Requirement:

Place the customers files in HDFS directory "/user/cloudera/problem1/customers/textdata"

Save output in text format with fields seperated by a '*' and lines should be terminated by pipe

Load only "Customer id, Customer fname, Customer lname and Customer street name"

Sample Output

11942*Mary*Bernard*Tawny Fox Plaza|10480*Robert*Smith*Lost Horse Plaza|.....