

Write an application using HBase and HiveQL for OnlineRetail Dataset which will include

i. Create and Load table with Online Retail data in Hive.

```
CREATE TABLE online_retail (  
  invoice_no STRING,  
  stock_code STRING,  
  description STRING,  
  quantity INT,  
  invoice_date STRING,  
  unit_price FLOAT,  
  customer_id STRING,  
  country STRING  
)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY '\t'  
STORED AS TEXTFILE;
```

```
LOAD DATA LOCAL INPATH '/home/cloudera/online_retail.csv' INTO TABLE online_retail;
```

j. Create Index on Online Retail Table in Hive.

```
CREATE INDEX idx_customer_id  
ON TABLE online_retail (customer_id)  
AS 'COMPACT'  
WITH DEFERRED REBUILD;
```

```
ALTER INDEX idx_customer_id ON online_retail REBUILD;
```

k. Find the total, average sales in Hive

```
SELECT  
  SUM(quantity * unit_price) AS total_sales,  
  AVG(quantity * unit_price) AS average_sales  
FROM online_retail;
```

l. Find Order details with maximum cost.

```
SELECT  
  invoice_no,  
  customer_id,  
  SUM(quantity * unit_price) AS order_total  
FROM online_retail  
GROUP BY invoice_no, customer_id  
ORDER BY order_total DESC  
LIMIT 1;
```

m. Find Customer details with maximum order total.

```
SELECT
  customer_id,
  SUM(quantity * unit_price) AS total_spent
FROM online_retail
GROUP BY customer_id
ORDER BY total_spent DESC
LIMIT 1;
```

n. Find the Country with maximum and minimum sale.

```
SELECT
  country,
  SUM(quantity * unit_price) AS country_sales
FROM online_retail
GROUP BY country
ORDER BY country_sales DESC
LIMIT 1;
```

```
SELECT
  country,
  SUM(quantity * unit_price) AS country_sales
FROM online_retail
GROUP BY country
HAVING country_sales > 0
ORDER BY country_sales ASC
LIMIT 1;
```

o. Creating an external Hive table to connect to the HBase for OnlineRetail.

```
hbase shell
create 'online_retail_hbase', 'info'
put 'online_retail_hbase', '1001', 'info:customer_id', '12345'
put 'online_retail_hbase', '1001', 'info:country', 'United Kingdom'
put 'online_retail_hbase', '1001', 'info:total', '250.00'
```

```
CREATE EXTERNAL TABLE hbase_online_retail (
  key STRING,
  customer_id STRING,
  country STRING,
  total STRING
)
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH SERDEPROPERTIES (
  "hbase.columns.mapping" = ":key,info:customer_id,info:country,info:total"
)
TBLPROPERTIES("hbase.table.name" = "online_retail_hbase");
```

p. Display records of OnlineRetail Table in Hbase.

```
scan 'online_retail_hbase'
```