

Hive + HBase: Customer Information System - Commands Only

a. Create Tables in Hive

```
CREATE DATABASE retail;
```

```
USE retail;
```

```
CREATE TABLE customer_info(cust_id INT, cust_name STRING, order_id INT)
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY ','
```

```
STORED AS TEXTFILE;
```

```
CREATE TABLE order_info(order_id INT, item_id INT, qty INT)
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY ','
```

```
STORED AS TEXTFILE;
```

```
CREATE TABLE item_info(item_id INT, item_name STRING, item_price FLOAT)
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY ','
```

```
STORED AS TEXTFILE;
```

b. Load Data into Hive Tables

```
LOAD DATA LOCAL INPATH '/home/cloudera/customer_info.txt' INTO TABLE customer_info;
```

```
LOAD DATA LOCAL INPATH '/home/cloudera/order_info.txt' INTO TABLE order_info;
```

```
LOAD DATA LOCAL INPATH '/home/cloudera/item_info.txt' INTO TABLE item_info;
```

c. Perform Join in Hive

```
SELECT c.cust_id, c.cust_name, o.order_id, o.qty, i.item_name, i.item_price, (o.qty *  
i.item_price) AS total_price
```

```
FROM customer_info c
```

```
JOIN order_info o ON c.order_id = o.order_id
```

```
JOIN item_info i ON o.item_id = i.item_id;
```

d. Create Index on Customer Info Table

```
CREATE INDEX idx_cust_id
```

```
ON TABLE customer_info(cust_id)
```

```
AS 'COMPACT'
```

```
WITH DEFERRED REBUILD;
```

e. Find Total and Average Sales

```
SELECT SUM(o.qty * i.item_price) AS total_sales
FROM order_info o
JOIN item_info i ON o.item_id = i.item_id;

SELECT AVG(o.qty * i.item_price) AS average_sales
FROM order_info o
JOIN item_info i ON o.item_id = i.item_id;
```

f. Find Order with Maximum Cost

```
SELECT o.order_id, MAX(o.qty * i.item_price) AS max_cost
FROM order_info o
JOIN item_info i ON o.item_id = i.item_id
GROUP BY o.order_id
ORDER BY max_cost DESC
LIMIT 1;
```

g. Create External Hive Table for HBase Integration

```
CREATE EXTERNAL TABLE hbase_customer_info(cust_id INT, cust_name STRING, order_id INT)
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH SERDEPROPERTIES ("hbase.columns.mapping" = ":key,info:name,info:order_id")
TBLPROPERTIES ("hbase.table.name" = "hbase_customer_info");
```

h. Display Records from HBase Table

```
SELECT * FROM hbase_customer_info;
```