

## Assignment - Data Aggregation Activity (10%)

1-

Table shows how many salespeople have attrition

The screenshot shows the Ambari web interface. On the left is a sidebar with navigation links: Dashboard, Services (HDFS, YARN, MapReduce2, Tez, Hive, Pig, Sqoop, Oozie, ZooKeeper, Ambari Metrics, WebHCat), and Hosts. The main area displays a Hive query in a text editor:

```
1 select count(attrition) as attrition from employee
2 where attrition like "%Yes%";
3
4
```

Below the query editor are buttons: Execute (green), Save As, Insert UDF, and Visual Explain. Below these are tabs: RESULTS (selected), LOG, VISUAL EXPLAIN, and TEZ UI. Under the RESULTS tab, there is a 'Filter col' button and a table showing the result:

attrition
237

The table above shows the number of sales people (237) with yes attribution

**The table counts how many salespeople have no attrition**

The screenshot shows the Ambari web interface with the same sidebar as the previous image. The main area displays a different Hive query:

```
1 select count(attrition) as Non_attrition from employee
2 where attrition like "%No%";
3
4
```

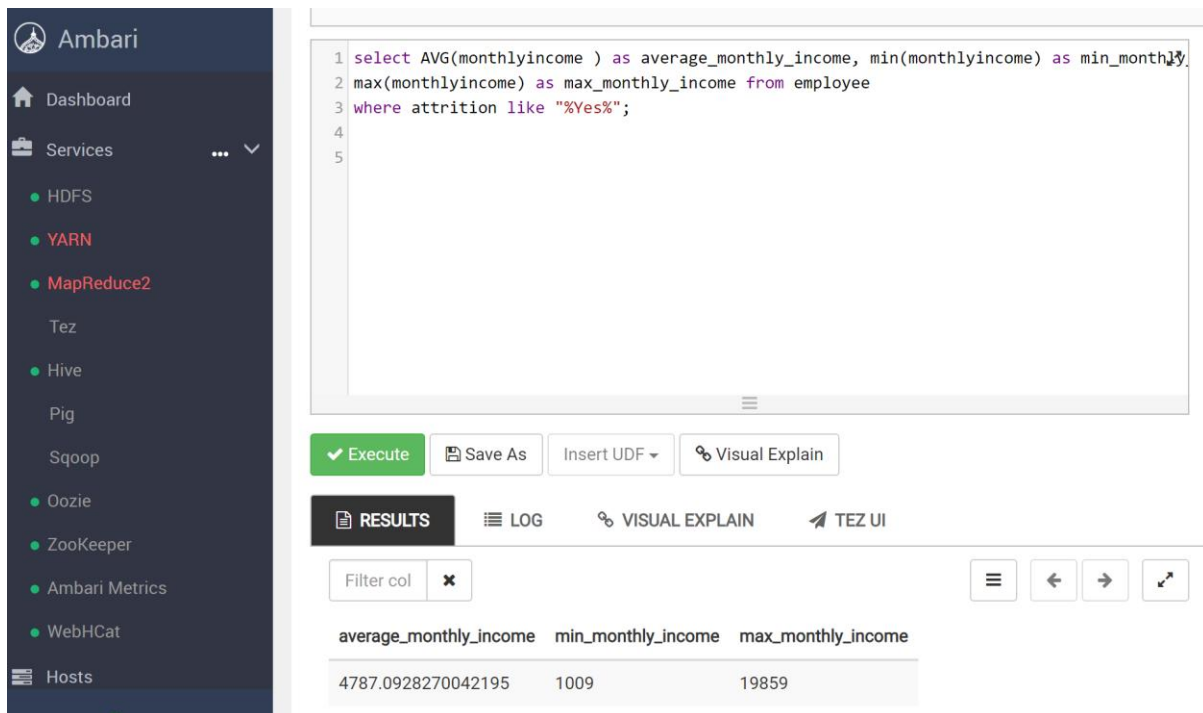
Below the query editor are the same buttons: Execute, Save As, Insert UDF, and Visual Explain. Below these are the same tabs: RESULTS (selected), LOG, VISUAL EXPLAIN, and TEZ UI. Under the RESULTS tab, there is a 'Filter col' button and a table showing the result:

non_attrition
1233

The table above shows the number of sales people (237) with NO attribution

## 2-

Table 1 shows the (average, min, max) monthly income for all salespeople with attrition



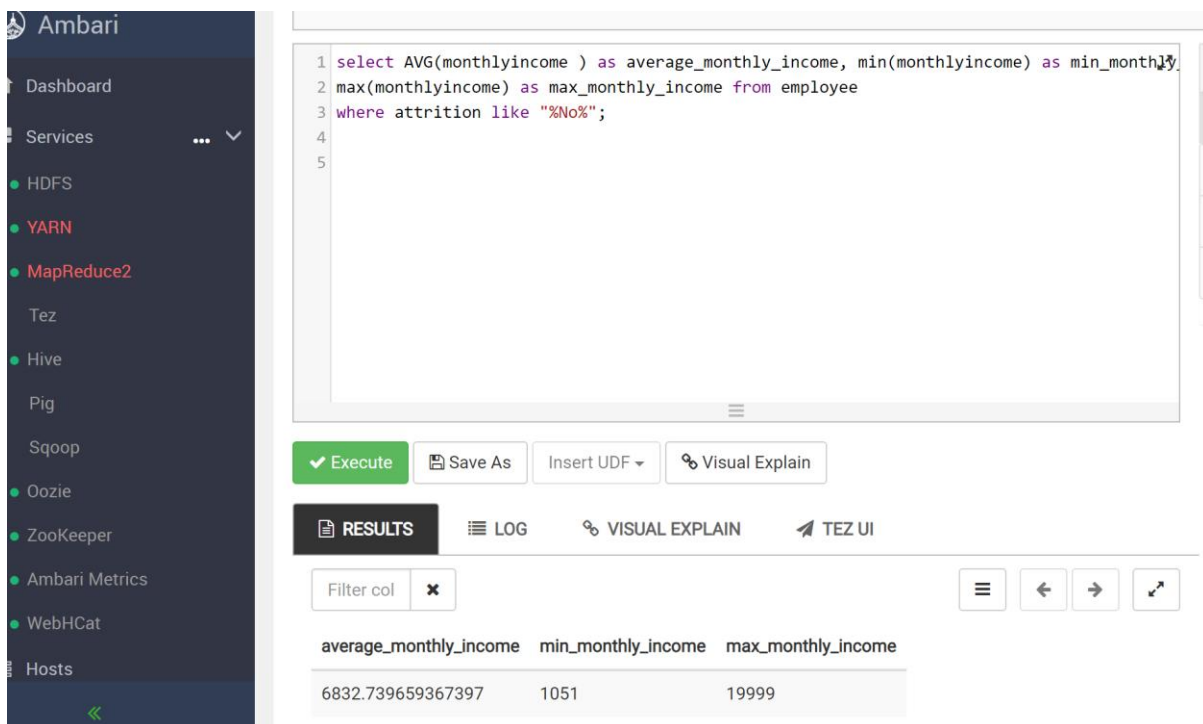
The screenshot shows the Ambari web interface. On the left is a sidebar with navigation links: Dashboard, Services (HDFS, YARN, MapReduce2, Tez, Hive, Pig, Sqoop, Oozie, ZooKeeper, Ambari Metrics, WebHCat), and Hosts. The main panel displays a Hive query editor with the following SQL:

```
1 select AVG(monthlyincome ) as average_monthly_income, min(monthlyincome) as min_monthly_income,
2 max(monthlyincome) as max_monthly_income from employee
3 where attrition like "%Yes%";
4
5
```

Below the query editor are buttons: Execute (green), Save As, Insert UDF, and Visual Explain. Below these are tabs: RESULTS (selected), LOG, VISUAL EXPLAIN, and TEZ UI. A 'Filter col' button is also present. The results table shows the following data:

average_monthly_income	min_monthly_income	max_monthly_income
4787.0928270042195	1009	19859

Table 2 shows the (average, min, max) monthly income for all salespeople without attrition



The screenshot shows the Ambari web interface with the same sidebar as the previous image. The main panel displays a Hive query editor with the following SQL:

```
1 select AVG(monthlyincome ) as average_monthly_income, min(monthlyincome) as min_monthly_income,
2 max(monthlyincome) as max_monthly_income from employee
3 where attrition like "%No%";
4
5
```

Below the query editor are buttons: Execute (green), Save As, Insert UDF, and Visual Explain. Below these are tabs: RESULTS (selected), LOG, VISUAL EXPLAIN, and TEZ UI. A 'Filter col' button is also present. The results table shows the following data:

average_monthly_income	min_monthly_income	max_monthly_income
6832.739659367397	1051	19999

### 3-

The table containing the monthly income and count of salespeople making that monthly income with attrition

The screenshot shows a web interface for a Hadoop distribution. On the left is a sidebar with navigation links: Dashboard, Services (HDFS, YARN, MapReduce2, Tez, Hive, Pig, Sqoop, Oozie, ZooKeeper, Ambari Metrics, WebHCat), and Hosts. The main area displays a SQL query editor with the following code:

```
1 select monthlyincome, count(monthlyincome) as count from employee
2 where attrition like "%Yes%"
3 group by monthlyincome ;
4
5
```

Below the editor are buttons for Execute, Save As, Insert UDF, and Visual Explain. The Execute button is highlighted in green. Below these buttons are tabs for RESULTS, LOG, VISUAL EXPLAIN, and TEZ UI. The RESULTS tab is active, showing a table with the following data:

monthlyincome	count
1009	1
1081	1
1091	1
1102	1
1118	1
1261	1
1359	1
1393	1
1416	1
1420	1
1555	1
1569	1
1601	1
1675	1
1790	1
1859	1
1878	1

The table containing the monthly income and count of salespeople making that monthly income without attrition

Services

HDFS

YARN

MapReduce2

Tez

Hive

Pig

Sqoop

Oozie

ZooKeeper

Ambari Metrics

WebHCat

Hosts

Worksheet1 \*   Worksheet2 \*   **Worksheet3 \***   +

DATABASE

Select or search database/schema

default

1 select monthlyincome,count(monthlyincome) as count from employee

2 where attrition like "%No%"

3 group by monthlyincome ;

4

5

Execute

Save As

Insert UDF

Visual Explain

RESULTS

LOG

VISUAL EXPLAIN

TEZ UI

Filter col

monthlyincome

count

1051	1
1052	1
1129	1
1200	1
1223	1
1232	1
1274	1
1281	1
1483	1
1514	1
1563	1
1611	1
1702	1
1706	1
1951	1
2001	1
2007	1
2008	2
2011	1
2013	1
2014	1
2024	1
2028	1
2029	1
2033	1
2042	1