

## Assignment - Data Transform Activity (20%)

- 1) a) Create a new table for your analysis called “employee\_sales”.
- b) Load the table “employee” into this table.
- c) Select these columns: Attrition, Department, JobSatisfaction & MonthlyIncome.

The screenshot shows the Ambari Hive View 2.0 interface. On the left is a sidebar with navigation links: Dashboard, Services, HDFS, YARN, MapReduce2, Tez, Hive, Oozie, ZooKeeper, Ambari Metrics, WebHcat, Hosts, Alerts, and Cluster Admin. The main panel is titled 'HIVE' and has tabs for QUERY, JOBS, TABLES, SAVED QUERIES, UDFs, and SETTINGS. A green notification bar at the top says 'Query has been submitted. (details)'. Below the tabs, there's a 'Worksheet2' editor. The 'DATABASE' dropdown is set to 'default'. The query editor contains the following SQL code:

```
1 CREATE TABLE employee_sales
2 (
3   Attrition string,
4   Department string,
5   JobSatisfaction int,
6   MonthlyIncome int
7 );
8
```

At the bottom of the editor are buttons for 'Execute', 'Save As', 'Insert UDF', and 'Visual Explain'. On the right side, there's a 'Tables(2)' panel showing a search bar and a list of tables: 'employee' and 'hivesampletable'.

The screenshot shows the Ambari Hive View 2.0 interface after the first query. The 'DATABASE' dropdown is still 'default'. The query editor now contains the following SQL code:

```
1 INSERT OVERWRITE TABLE employee_sales
2 SELECT Attrition, Department, JobSatisfaction, MonthlyIncome
3 FROM employee;
4
```

The 'Execute' button is highlighted in green. On the right side, the 'Tables(3)' panel now shows three tables: 'employee', 'employee\_sales', and 'hivesampletable'. The 'employee\_sales' table is highlighted in yellow.

2) Round the data found in the “MonthlyIncome” column to the nearest \$1000.

The screenshot shows the Ambari Hive View 2.0 interface. The left sidebar contains navigation links for Dashboard, Services, HDFS, YARN, MapReduce2, Tez, Hive, Oozie, ZooKeeper, Ambari Metrics, WebHCat, Hosts, Alerts, Cluster Admin, Stack and Versions, Service Accounts, and Service Auto Start. The main panel displays the Hive query editor. A green notification bar at the top states "Query has been submitted. (details)". The query editor shows a SQL query: 

```
1 INSERT OVERWRITE TABLE employee_sales
2 SELECT Attrition, Department, JobSatisfaction, ROUND(MonthlyIncome,-3)
3 FROM employee_sales;
4
```

 The right sidebar shows a table list for the 'default' database, including 'employee', 'employee\_sales' (highlighted), and 'hivesampletable'. The bottom of the editor has buttons for 'Execute', 'Save As', 'Insert UDF', and 'Visual Explain'.

The screenshot shows the Ambari Hive View 2.0 interface with the query results displayed. The query editor shows the query: 

```
1 SELECT * FROM employee_sales LIMIT 10;
```

 The 'RESULTS' tab is active, showing a table with 4 columns: 'employee\_sales.attrition', 'employee\_sales.department', 'employee\_sales.jobssatisfaction', and 'employee\_sales.monthlyincome'. The table contains 10 rows of data. The right sidebar shows the same table list as the previous screenshot, with 'employee\_sales' highlighted. The bottom of the editor has buttons for 'Execute', 'Save As', 'Insert UDF', and 'Visual Explain'.

employee_sales.attrition	employee_sales.department	employee_sales.jobssatisfaction	employee_sales.monthlyincome
Yes	Sales	4	6000
No	Research & Development	2	5000
Yes	Research & Development	3	2000
No	Research & Development	3	3000
No	Research & Development	2	3000
No	Research & Development	4	3000
No	Research & Development	1	3000
No	Research & Development	3	3000
No	Research & Development	3	10000
No	Research & Development	3	5000

3) Filter the data to only look at those items in the “Sales Department”.

The top screenshot shows the Ambari Hive View 2.0 interface. The left sidebar contains navigation links for Dashboard, Services, HDFS, YARN, MapReduce2, Tez, Hive, Oozie, ZooKeeper, Ambari Metrics, WebHCat, and Hosts. The main panel displays the HIVE query editor. A green notification bar at the top indicates "Query has been submitted. (details)". The query editor shows a query to insert and overwrite the employee\_sales table with data from the employee\_sales table, filtered by department like "%Sales%". The right sidebar shows a list of tables in the default database: employee, employee\_sales, and hivesampletable.

The bottom screenshot shows the same Ambari Hive View interface, but the query editor now displays a query to select all data from the employee\_sales table, limited to 10 rows. The right sidebar shows the same list of tables. Below the query editor, the "RESULTS" tab is active, displaying a table with the following data:

employee_sales.attrition	employee_sales.department	employee_sales.job satisfaction	employee_sales.monthly income
Yes	Sales	4	6000
No	Sales	4	15000
Yes	Sales	1	3000
No	Sales	2	7000
No	Sales	1	19000

4) Order the data by “JobSatisfaction” from highest to lowest.

The top screenshot shows the Ambari Hive View interface. The left sidebar contains navigation links for Dashboard, Services, HDFS, YARN, MapReduce2, Tez, Hive, Oozie, ZooKeeper, Ambari Metrics, WebHCat, Hosts, Alerts, Cluster Admin, Stack and Versions, Service Accounts, and Service Auto Start. The main panel displays the 'HIVE' section with a 'QUERY' tab selected. A green notification bar at the top indicates 'Query has been submitted. (details)'. The query editor shows the following HiveQL:

```
1 INSERT OVERWRITE TABLE employee_sales
2 SELECT *
3 FROM employee_sales
4 ORDER BY JobSatisfaction DESC;
5
```

The bottom screenshot shows the same interface with the 'RESULTS' tab selected. The query results are displayed as a table with the following columns: employee\_sales.attrition, employee\_sales.department, employee\_sales.jobsatisfaction, and employee\_sales.monthlyincome. The results are ordered by jobsatisfaction in descending order.

employee_sales.attrition	employee_sales.department	employee_sales.jobsatisfaction	employee_sales.monthlyincome
No	Sales	4	5000
No	Sales	4	9000
No	Sales	4	18000
No	Sales	4	10000
No	Sales	4	7000
No	Sales	4	8000
Yes	Sales	4	1000
No	Sales	4	5000