

## Soudabeh Rafieisakhaei

1. Download the client.txt, building.txt, and manager.txt files from the Chapter 6 data set. Files contain datasets about a hotel chain and its business clients.
  1. Download the files and upload them to HDFS under the HV/FinalExam/input

```
[hadoop@ip-172-31-26-174 ~]$ hadoop fs -mkdir HV
mkdir: `HV': File exists
[hadoop@ip-172-31-26-174 ~]$ hadoop fs -mkdir HV/FinalExam[hadoop@ip-172-31-26-174 ~]$ ls
building.txt  client.txt  manager.txt
[hadoop@ip-172-31-26-174 ~]$ hadoop fs -mkdir HV/FinalExam/input
[hadoop@ip-172-31-26-174 ~]$ hadoop fs -ls
Found 1 items
drwxr-xr-x  - hadoop hdfsadmingroup          0 2024-05-15 08:05 HV
[hadoop@ip-172-31-26-174 ~]$ hadoop fs -copyFromLocal building.txt HV/FinalExam/input
[hadoop@ip-172-31-26-174 ~]$ hadoop fs -copyFromLocal client.txt HV/FinalExam/input
[hadoop@ip-172-31-26-174 ~]$ hadoop fs -copyFromLocal manager.txt HV/FinalExam/input
[hadoop@ip-172-31-26-174 ~]$ []
```

```
[hadoop@ip-172-31-26-174 ~]$ hadoop fs -ls HV/FinalExam/input
Found 3 items
-rw-r--r--  1 hadoop hdfsadmingroup      92 2024-05-15 09:15 HV/FinalExam/input/building.txt
-rw-r--r--  1 hadoop hdfsadmingroup     183 2024-05-15 08:08 HV/FinalExam/input/client.txt
-rw-r--r--  1 hadoop hdfsadmingroup     201 2024-05-15 08:08 HV/FinalExam/input/manager.txt
[hadoop@ip-172-31-26-174 ~]$ []
```

2. Create a database named FinalExam and create the three tables.

```
hive> DROP DATABASE FinalExam;
OK
Time taken: 0.029 seconds
hive>
> CREATE DATABASE FinalExam;
OK
Time taken: 0.018 seconds
hive> USE FinalExam;
OK
Time taken: 0.011 seconds
hive> CREATE TABLE FinalExam.building (
    >     building_id STRING,
    >     no_of_floors INT,
    >     manager_id STRING
    > );
OK
Time taken: 0.214 seconds
```

```
hive> CREATE TABLE FinalExam.client (
    >     client_id INT,
    >     client_name STRING,
    >     client_industry STRING,
    >     client_location STRING
    > );
OK
Time taken: 0.05 seconds
hive>
> CREATE TABLE FinalExam.manager (
    >     manager_id STRING,
    >     manager_firstname STRING,
    >     manager_lastname STRING,
    >     manager_dob INT,
    >     manager_salary DECIMAL(10, 2),
    >     manager_bonus DECIMAL(10, 2)
    > );
OK
Time taken: 0.045 seconds
hive> []
```

3. Load the data from the HDFS files into each table, respectively.

```
hive> LOAD DATA INPATH 'HV/FinalExam/input/building.txt' INTO TABLE FinalExam.building;
Loading data to table finalexam.building
OK
Time taken: 0.672 seconds
hive> LOAD DATA INPATH 'HV/FinalExam/input/client.txt' INTO TABLE FinalExam.client;
Loading data to table finalexam.client
OK
Time taken: 0.132 seconds
hive> LOAD DATA INPATH 'HV/FinalExam/input/manager.txt' INTO TABLE FinalExam.manager;
Loading data to table finalexam.manager
OK
Time taken: 0.131 seconds
hive> []
```

4. Display the list of all managers, including manager's id, first name, and last name.

```
hive> SELECT manager_id, manager_firstname, manager_lastname FROM FinalExam.manager;
OK
```

```
manager_id      manager_firstname      manager_lastname
M230          Fred            Lee
M120          Larry           Long
M340          George          Shetuni
Time taken: 0.053 seconds, Fetched: 3 row(s)
hive> []
```

5. Display the list of all clients, including their id, name, and industry.

```
hive> SELECT client_id, client_name, client_industry FROM FinalExam.client;
OK
```

```
client_id      client_name      client_industry
2220          Skyline Airline
8880          SouthTravel    Tourism
7770          LightRecords   Music
1110          Harmony Music
Time taken: 0.054 seconds, Fetched: 4 row(s)
hive> []
```

6. Display the hotel building id, the number of floors, and the

manager's first name and last name for each hotel.

```
hive> SELECT
    >     b.building_id,
    >     b.no_of_floors AS number_of_floors,
    >     m.manager_firstname AS manager_first_name,
    >     m.manager_lastname AS manager_last_name
    > FROM
    >     FinalExam.building b
    > JOIN
    >     FinalExam.manager m ON b.manager_id = m.manager_id;
Query ID = hadoop_20240515101631_e099ac7b-812f-46af-9a63-170eb19d39f4
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1715754688638_0007)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 2 .....	container	SUCCEEDED	1	1	0	0	0	0
Map 1 .....	container	SUCCEEDED	1	1	0	0	0	0

VERTICES: 02/02 [=====>>>] 100% ELAPSED TIME: 6.24 s

OK

```
SELECT manager_firstname, manager_lastname, manager_salary, manager_yob, num_hotel_buildings
FROM FinalExam.manager m
JOIN FinalExam.building b ON m.manager_id = b.manager_id
WHERE manager_salary > 70000
GROUP BY manager_firstname, manager_lastname, manager_salary, manager_yob;
```

7. Display the manager's first and last name, as well as salary, date of birth, and the number of hotel buildings that the manager manages for all managers with a salary higher than \$70,000.

```
hive> SELECT m.manager_firstname,
    >     m.manager_lastname,
    >     m.manager_salary,
    >     m.manager_dob,
    >     COUNT(b.building_id) AS num_hotel_buildings
    > FROM FinalExam.manager m
    > JOIN FinalExam.building b ON m.manager_id = b.manager_id
    > WHERE m.manager_salary > 70000
    > GROUP BY m.manager_firstname, m.manager_lastname, m.manager_salary, m.manager_dob;
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1 .....	container	SUCCEEDED	1	1	0	0	0	0
Map 2 .....	container	SUCCEEDED	1	1	0	0	0	0
Reducer 3 .....	container	SUCCEEDED	2	2	0	0	0	0

VERTICES: 03/03 [=====>>>] 100% ELAPSED TIME: 6.53 s

OK

```
SELECT manager_firstname, manager_lastname, manager_salary, manager_yob, num_hotel_buildings
FROM FinalExam.manager m
JOIN FinalExam.building b ON m.manager_id = b.manager_id
WHERE manager_salary > 70000
GROUP BY manager_firstname, manager_lastname, manager_salary, manager_yob;
```