



**Hope Foundation's  
Finolex Academy of Management and Technology, Ratnagiri  
Information Technology Department**

Subject name: Big Data Lab		Subject Code: ITC801	
Class	BE IT	Semester – VIII (CBGS)	Academic year: 2019-20
Name of Student	Kazi Jawwad A Rahim		QUIZ Score :
Roll No	28	Assignment/Experiment No.	02
<b>Title:</b> Execution of Hive SQL Queries on Hadoop by using HUE interface			
<b>1. Course objectives applicable COB1.</b> To understand main business drivers and key issues of BDA <b>COB2.</b> To acquire knowledge about fundamentals of Big Data Analytics <b>COB4 –</b> To handle larger database through BDA framework			
<b>2. Course outcomes applicable:</b> <b>CO1 :</b> Understand the key issues in big data management and its associated applications in intelligent business and scientific computing. <b>CO2 -</b> Acquire fundamental enabling techniques and scalable algorithms like Hadoop, Map Reduce and NO SQL in big data analytics. <b>COB4.</b> Implement use of combiners to consolidate results and ability to handle larger datasets			
<b>3. Learning Objectives:</b> <ol style="list-style-type: none"><li>1. To understand the concept of Hadoop User Interface</li><li>2. To understand the functioning of Hive SQL, Pig</li><li>3. To Execute Hive SQL Query</li><li>4. To Execute Pig Script</li></ol>			
<b>4. Practical applications of the assignment/experiment:</b> Hue is a widely used GUI dashboard for Hadoop			
<b>5. Prerequisites:</b> <ol style="list-style-type: none"><li>1. Knowledge of Hadoop Ecosystem</li><li>2. Knowledge of basic SQL queries</li></ol>			
<b>6. Hardware Requirements:</b> <ol style="list-style-type: none"><li>1. PC with 4GB RAM, 500GB HDD</li></ol>			
<b>7. Software Requirements:</b> <ol style="list-style-type: none"><li>1. Ubuntu / Windows , access to internet www.gethue.com</li></ol>			
<b>8. Quiz Questions (if any): (Online Exam will be taken separately batchwise, attach the certificate/ Marks obtained)</b> <ol style="list-style-type: none"><li>1. What is a Hadoop?</li><li>2. What is SQL?</li><li>3. What is Pig Latin?</li><li>4. What is Hue?</li><li>5. In which language Hue interface is programmed?</li></ol>			
Sr. No.	Parameters	Marks obtained	Out of
1	Technical Understanding (Assessment may be done based on Q & A <u>or</u> any other relevant method.) Teacher should mention the other method used -	05	6
2	Neatness/presentation	01	2
3	Punctuality	02	2
Date of performance (DOP)	15/1/20	Total marks obtained	08
Date of checking (DOC)	5/2/20	Signature of teacher	



## 1. Theory:

Hue (Hadoop User Experience) is an open-source web experience interface that supports Apache Hadoop and its ecosystem, licensed under the Apache V2 license.

Stable release: 4.1.0 / September 4, 2017;

Repository: <https://github.com/cloudera/hue>

Written in: Python

Type: Web platform

License: Apache License 2.0

Website: [gethue.com](http://gethue.com)

## Features:

Hue is an open-source Analytics Workbench for browsing, querying and visualizing data.

## Applications:

Editors for Hive, Impala, Pig, MapReduce, Spark and any SQL like MySQL, Oracle, SparkSQL, Solr SQL, Phoenix and more.

Dashboards to dynamically interact and visualize data with Solr or SQL.

Scheduler of jobs and workflows.

Browsers for jobs, HDFS, S3 files, SEL Tables, Indexes, Git files, Sentry permissions, Sqoop and more.

The screenshot shows the Hadoop Hue Web Interface. On the left, there's a sidebar with various database and system tables listed under 'Tables'. The main area has a 'Query' tab selected, displaying a SQL query:

```

1 Some high risks were detected.
2 -- the objective is to find the JIRAs in Hue where there are multiple SFDC tickets linked
3 -- it reveals the soft spots in the product
4
5 SELECT sfdc.jira_c.name,
6       sfdc.jira_c.jira_summary_c,
7       COUNT(sfdc.jira_c.name) AS tickets
8   FROM sfdc.jira_c
9     JOIN sfdc.ticket ON sfdc.jira_c.id = sfdc.ticket.jira_c_id
10  WHERE sfdc.cases.component_c IN ('Hue')
11    AND sfdc.jira_c.case_c = sfdc.cases.id
12    AND sfdc.jira_c.summary_c = sfdc.jira_c.name
13    AND sfdc.ticket.statusname NOT IN ('Resolved', 'Closed')
14    AND sfdc.jira_c.name NOT LIKE 'CLX'
15 GROUP BY sfdc.jira_c.name, sfdc.jira_c.jira_summary_c
16 HAVING COUNT(sfdc.jira_c.name) > 1
17 ORDER BY count(jira_c.name) DESC

```

Below the query editor, there are filters for 'component: Hue', 'type: Escalation', and 'date: 2016-10-01'. To the right, a table titled 'Results (19)' shows a list of tickets with columns 'name', 'jira\_summary\_c', and 'tickets'.

**Figure 1. Hadoop Hue Web Interface**

## 2. Precautions :

1. Internet should be active
2. Copy the Query and its output before executing next query

## 3. Installation Steps / Performance Steps -

1. Open [www.gethue.com](http://www.gethue.com), login with : Username – demo, password: demo
2. Select SQL Menu and it will display the databases

This screenshot shows the Hue interface with the 'SQL' menu selected. The left sidebar lists available databases: 'default', 'canada\_facts', 'employee', 's2', 'sample\_07', 'sample\_08', 'simple', and 'web\_logs'. The main area has a 'Hive' tab selected, showing a query history entry: 'Example: SELECT \* FROM tablename, or press CTRL + E'.

3. Write a Query on available tables :

The screenshot shows the Hue Editor interface. In the top navigation bar, there are tabs for 'Hue - Editor', 'Big SQL Sandbox | IBM', and 'IBM BigInsights'. The main area is titled 'HUE' with a 'Query' dropdown and a search bar. On the left, a sidebar shows a tree view of tables under the 'default' database, including 'canada\_facts', 'employee', 's2', 'sample\_07', 'sample\_08', 'simple', and 'web\_logs'. The main pane displays a query editor with the following code:

```
1 select * from sample_08 where sample_08.salary > 100000
```

Below the code, it says '51m, 20s default text ?'. At the bottom of the editor, there are buttons for 'Query History', 'Saved Queries', and 'Query Builder'.

4. Execute the Query:

## 4. Observations

List observations here (if any)

## 5. Results:

1 CREATE DATABASE PROFESSIONALK;

```
/usr/bin/hive -e "CREATE DATABASE PROFESSIONALK"  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=hive_20200115061824_b8f077d8-843e-4208-b685-ef85fb60697c); Time taken: 0.073 seconds  
INFO : OK
```

✓ Success.

```
1| USE PROFESSIONALJK;
```

```
0/1: USE PROFESSIONALJK
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20200115062212_48ef7538-dfde-4d4a-945d-ce6b91a8dcba); Time taken: 0.005 seconds
INFO : OK
```

✓ Success.

```
1| CREATE TABLE PROFESSIONALJK.TEST(ID INT, NAME VARCHAR(20));
```

```
0/1: CREATE TABLE PROFESSIONALJK.TEST(ID INT, NAME VARCHAR(20))
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20200115062436_66ae2800-1732-43f4-9142-3e6a27e4d307); Time taken: 0.054 seconds
INFO : OK
```

✓ Success.

```
1| SELECT * FROM PROFESSIONALJK.TEST;
```

Query History	Saved Queries	Query Builder	Results (1)
test.id		test.name	
1	1	JAWWAD	

## References :

- [1] Apache Licence : <https://github.com/cloudera/hue#license>
- [2] Hue Live Interface available at : <http://demo.gethue.com/hue/accounts/login/?next=/>



o Learning Outcomes Achieved:

1. Students handled the hue interface for Hadoop.
2. Hive SQL and Pig Scripting functionality was demonstrated by students.
3. One hive SQL query and one pig latin script was tested.
4. The output were analyzed for relevance.

o Conclusion:

1. Applications of the studied technique in industry.
  - a. hue is a interface written in Python, it gives better accessibility to Hadoop clusters.
  - b. Hive and pig are the database related components of hadoop.

2. Engineering Relevance.

- a. Data manipulation functionality is provided through Hive SQL.

3. Skills Developed

- a. Use of hue interface for Hadoop management and component access.
- b. Running Hive SQL queries and Pig script.