Hands-On Exercise: Running Queries from Shells, Scripts, and Hue

Files and Data Used in This Exercise

Exercise directory	$ADIR/exercises/queries
Hive/Impala tables	customers, products

In this exercise, you will practice using the Hue query editor and the Impala and Beeline shells to execute simple queries. These exercises use the tables that have been populated with data you imported to HDFS using Sqoop in an earlier exercise.

Exploring the customers Table Using Hue

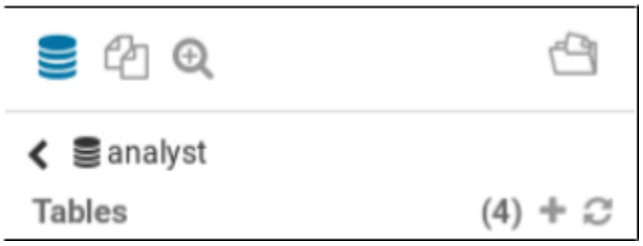
One way to run Hive and Impala queries is through your web browser using Hue's Query Editors. This is especially convenient if you use more than one computer—or if you use a device (such as a tablet) that isn't capable of running the Impala or Beeline shells itself—because it does not require any software other than a web browser.

1. Start the Firefox web browser if it isn't running, then click on the Hue bookmark in the Firefox bookmark toolbar (or type http://localhost:8889/home into the address bar and press the Enter key).

2. After a few seconds, you should see Hue's home screen. If you don't currently have an active session, you will first be prompted to log in. Enter in the username: **cloudera** and password: **cloudera** fields, and then click the **Sign In** button.

3. Click the **Query** button. Note that there are query editors for both Hive and Impala (as well as other tools), accessible using the drop-down menu attached to this button. The interface is very similar for both Hive and Impala. For these exercises, you will use the Impala query editor.

4. The left panel shows available databases. Navigate to the analyst database in Hive or Impala, if it's not already selected. You might need to click the back arrow.

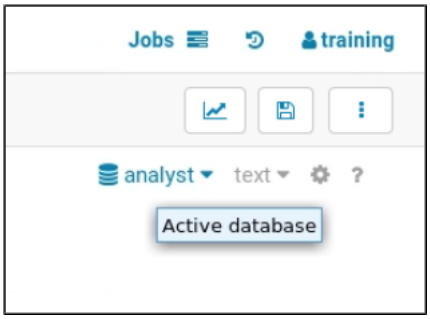


5. Below the selected database is a list of the tables in that database. Click the customers table to view the columns in the table. It may take a few seconds to display the columns.

6. Hover over the customers row and click the information **i** that appears next to the table name. Notice the sample data from the table. This is a quick way in Hue to get some idea what the data looks like.

7. Click the **X** to close the information box.

8. In the query editor panel, find the active database. This (not the left panel) specifies which database you are using with your queries. Make sure the analyst database is selected as the active database in the query editor panel. If necessary, select it yourself:



Running a Query Using Hue

Dualcore ran a contest in which customers posted videos of interesting ways to use their new tablets. A $5,000 prize will be awarded to the customer whose video received the highest rating.

However, the registration data was lost due to an RDBMS crash, and the only information we have is from the videos. The winning customer introduced herself only as "Bridget from Kansas City" in her video.

You will need to run a query that identifies the winner's record in our customer database so that we can send her the $5,000 prize.

9. All you know about the winner is that her name is Bridget and she lives in Kansas City. In the Impala Query Editor, enter a query in the text area to find the winning customer. Run the following command search for Kansas City customers with names such as Bridget, Bridgette, or Bridgitte.

```
SELECT * FROM customers
WHERE fname LIKE 'Bridg%' AND city = 'Kansas City';
```

10. After entering the query, click the **Execute** button (▶) to the left of the text area.

While the query is executing, the **Query History** tab displays the status of the query. When the query is complete, the **Results** tab opens, displaying the results of the query.

Question: Which customer did your query identify as the winner of the $5,000 prize?

Running a Query from the Impala Shell

Do the following to run a top-N query to identify the three most expensive products that Dualcore currently offers.

11. Start a terminal window if you don't currently have one running.

12. On the Linux command line in the terminal window, start the Impala shell:

```
$ impala-shell
```

Impala displays the URL of the Impala server in the shell command prompt, for example:

```
[localhost.localdomain:21000]>
```

13. At the prompt, switch to the analyst database by entering > USE analyst;

Remember that SQL commands in the shell must be terminated by a semicolon (;), unlike in the Hue query editor.

14. Next, review the schema of the products table by entering

```
> DESCRIBE products;
```

15. Show a sample of 10 records from the products table:

```
> SELECT * FROM products LIMIT 10;
```

16. Write and execute a query that displays the three most expensive products. (**Hint:** Use ORDER BY.)

17. When you are done, exit the Impala shell:

```
> quit;
```

Running a Script in the Impala Shell

The rules for the contest described earlier require that the winner bought the advertised tablet from Dualcore between May 1, 2013 and May 31, 2013. Before we can authorize our accounting department to pay the $5,000 prize, you must ensure that Bridget is eligible. Since this query involves joining data from several tables, and we have not yet covered joins, you've been provided with a script in the exercise directory.

18. Change to the directory for this hands-on exercise:

```
$ cd $ADIR/exercises/queries
```

19. Review the code for the query:

```
$ cat verify_tablet_order.sql
```

20. Execute the script using the shell's -f option:

```
$ impala-shell -f verify_tablet_order.sql
```

Question: Did Bridget order the advertised tablet in May?

Running a Query Using Beeline

21. At the command line in a terminal window, start Beeline, using username **cloudera** and password **cloudera**:

```
$ beeline -u jdbc:hive2://localhost:10000 \ -n cloudera -p cloudera
```

For the exercise environment, you should always use these options.

Beeline displays the URL of the Hive server in the shell command prompt, such as:
0: jdbc:hive2://localhost:10000>

Warning Messages

When you start Beeline in the exercise environment, you will see several lines starting with WARNING: and SLF4J:. These can safely be ignored.

22. Switch to the analyst database.

23. Write and execute a query to find all the Gigabux brand products whose price is less than 1000 (less than $10).

Hint: Use WHERE price < 1000 AND ….

24. Exit the Beeline shell:

> !exit