Lab 08 – Apache Impala

Full name: Chloe Tee Rouyi

Student ID: 0354731

Impala practices

1. Open terminal

2. Check running services: **sudo jps**

```
[cloudera@quickstart ~]$ sudo jps
6906 RunJar
5794 JobHistoryServer
7439 HRegionServer
5287 DataNode
5741 Bootstrap
8233
6291 HMaster
7939 Bootstrap
9389 Jps
7281 Bootstrap
5213 QuorumPeerMain
8209 Bootstrap
5873 NodeManager
5389 JournalNode
5602 SecondaryNameNode
8273
6726 ThriftServer
5476 NameNode
6823 RunJar
7306 HistoryServer
6430 RESTServer
6125 ResourceManager
```

3. Sign in as superuser using the following command (password is cloudera):

    a. [cloudera@quickstart ~]$ **su**

    Password: **cloudera**

```
[cloudera@quickstart ~]$ su
Password:
[root@quickstart cloudera]#
```

4. Start Impala Shell **impala-shell**

   after a while impala shell will be appeared. The prompt will be like "[quickstart.cloudera:21000] >"

```
[root@quickstart cloudera]# impala-shell
Starting Impala Shell without Kerberos authentication
Connected to quickstart.cloudera:21000
Server version: impalad version 2.7.0-cdh5.10.0 RELEASE (build 785a073cd07e2540d
521ecebb8b38161ccbd2aa2)
***********************************************************************************
***
Welcome to the Impala shell.
(Impala Shell v2.7.0-cdh5.10.0 (785a073) built on Fri Jan 20 12:03:56 PST 2017)

When you set a query option it lasts for the duration of the Impala shell sessio
n.
***********************************************************************************
***
[quickstart.cloudera:21000] >
```

5. Execute the following commands and write your understanding about what each command does:

   a. **help;**

```
[quickstart.cloudera:21000] > help;

Documented commands (type help <topic>):
========================================
compute  describe  explain  profile  select  shell  tip    use     version
connect  exit      history  quit     set     show   unset  values  with

Undocumented commands:
======================
alter   delete  drop  insert  source  summary  upsert
create  desc    help  load    src     update
```

   b. **version;**

```
[quickstart.cloudera:21000] > version;
Shell version: Impala Shell v2.7.0-cdh5.10.0 (785a073) built on Fri Jan 20 12:03:56 PST 2017
Server version: impalad version 2.7.0-cdh5.10.0 RELEASE (build 785a073cd07e2540d521ecebb8b38161ccbd2aa2)
[quickstart.cloudera:21000] >
```

c. **history;**

The “history” command of Impala displays the last 10 commands executed in the shell.

```
[quickstart.cloudera:21000] > history;
[1]: help;
[2]: version;
[3]: history;
```

d. **exit; or quit;**

```
[quickstart.cloudera:21000] > exit;
Goodbye cloudera
```

6. connect to a given instance of Impala:

connect;

```
[quickstart.cloudera:21000] > connect;
Connected to quickstart.cloudera:21000
Server version: impalad version 2.7.0-cdh5.10.0 RELEASE (build 785a073cd07e2540d521ecebb8b38161ccbd2aa2)
```

Impala Query Practices

7. Create a new database in Impala:

   a. Syntax: **CREATE DATABASE IF NOT EXISTS database_name;**

      Example: **CREATE DATABASE IF NOT EXISTS retaildb;**

```
[quickstart.cloudera:21000] > CREATE DATABASE IF NOT EXISTS retaildb;
Query: create DATABASE IF NOT EXISTS retaildb

Fetched 0 row(s) in 1.75s
```

   b. **Create a few more databases**

```
[quickstart.cloudera:21000] > CREATE DATABASE IF NOT EXISTS users;
Query: create DATABASE IF NOT EXISTS users

Fetched 0 row(s) in 0.17s
[quickstart.cloudera:21000] > CREATE DATABASE IF NOT EXISTS chloeLab08;
Query: create DATABASE IF NOT EXISTS chloeLab08

Fetched 0 row(s) in 0.17s
```

8. To see the list of the existing databases and verify the new databases added to the list:

   a. **show databases;**

```
[quickstart.cloudera:21000] > show databases;
Query: show databases
+------------------+-----------------------------------------------+
| name             | comment                                       |
+------------------+-----------------------------------------------+
| _impala_builtins | System database for Impala builtin functions  |
| chloelab08       |                                               |
| default          | Default Hive database                         |
| retaildb         |                                               |
| users            |                                               |
+------------------+-----------------------------------------------+
Fetched 5 row(s) in 0.12s
```

9. To drop database:

   a. Syntax: **DROP (DATABASE|SCHEMA) [IF EXISTS] database_name [RESTRICT | CASCADE] [LOCATION hdfs_path];**

```
[quickstart.cloudera:21000] > DROP DATABASE IF EXISTS users;
Query: drop DATABASE IF EXISTS users
[quickstart.cloudera:21000] > show databases;
Query: show databases
+------------------+-----------------------------------------------+
| name             | comment                                       |
+------------------+-----------------------------------------------+
| _impala_builtins | System database for Impala builtin functions  |
| chloelab08       |                                               |
| default          | Default Hive database                         |
| retaildb         |                                               |
+------------------+-----------------------------------------------+
Fetched 4 row(s) in 0.11s
```

10. To start using a database that is existed in Impala, use the following command:

    Syntax: **USE db_name;**

    USE retaildb;

```
[quickstart.cloudera:21000] > USE retaildb;
Query: use retaildb
```

11. Create tables inside a database:

    CREATE TABLE IF NOT EXISTS retaildb.retailers
    (firstname STRING, lastname STRING, age INT, phone STRING);

    show tables;

```
[quickstart.cloudera:21000] > CREATE TABLE IF NOT EXISTS retaildb.retailers (fir
stname STRING, lastname STRING, age INT, phone STRING);
Query: create TABLE IF NOT EXISTS retaildb.retailers (firstname STRING, lastname
 STRING, age INT, phone STRING)

Fetched 0 row(s) in 0.49s
[quickstart.cloudera:21000] > show tables;
Query: show tables
+-----------+
| name      |
+-----------+
| retailers |
+-----------+
Fetched 1 row(s) in 0.11s
```

12. Insert data:

a. Syntax 1: **insert into table_name (column1, column2, column3,...columnN) values (value1, value2, value3,...valueN);**

Syntax 2: **Insert into table_name values (value1, value2, value2);**

Example: **insert into retailers (firstname, lastname, age, phone) VALUES (‘David’, 'Smith', 32, '+60126753578');**

```
[quickstart.cloudera:21000] > insert into retailers (firstname, lastname, age, phone) VALUES ('Chloe','Tee',19,'+601248066
65');
Query: insert into retailers (firstname, lastname, age, phone) VALUES ('Chloe','Tee',19,'+60124806665')
Query submitted at: 2024-11-26 18:37:13 (Coordinator: http://quickstart.cloudera:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?query_id=ea4ab696b0af6f88:1f2883d100000000
Modified 1 row(s) in 1.04s
```

Additional Practice:

```
[quickstart.cloudera:21000] > insert into retailers (firstname, lastname, age, phone) VALUES ('Harry','Potter',50,'+601234
56789');
Query: insert into retailers (firstname, lastname, age, phone) VALUES ('Harry','Potter',50,'+60123456789')
Query submitted at: 2024-11-26 18:38:32 (Coordinator: http://quickstart.cloudera:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?query_id=18499695f0d1d3e9:2e5f0fc900000000
Modified 1 row(s) in 0.21s
[quickstart.cloudera:21000] > insert into retailers (firstname, lastname, age, phone) VALUES ('Ginny','Weasley',45,'+60123
456789');
Query: insert into retailers (firstname, lastname, age, phone) VALUES ('Ginny','Weasley',45,'+60123456789')
Query submitted at: 2024-11-26 18:38:55 (Coordinator: http://quickstart.cloudera:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?query_id=1545272970209047:83caec4200000000
Modified 1 row(s) in 0.11s
```

13. Do the following practices:

a. Create “employee” table:

create table employee (Id INT, name STRING, age INT,address STRING, salary BIGINT);

```
[quickstart.cloudera:21000] > create table employee (Id INT, name STRING, age IN
T, address STRING, salary BIGINT);
Query: create table employee (Id INT, name STRING, age INT, address STRING, sala
ry BIGINT)

Fetched 0 row(s) in 0.19s
```

b. Add some records by specifying attribute names:

insert into employee (ID,NAME,AGE,ADDRESS,SALARY) VALUES (1, 'Ramesh', 32, 'Ahmedabad', 20000); insert into employee (ID,NAME,AGE,ADDRESS,SALARY) VALUES (2, 'Khilan', 25, 'Delhi', 15000);

```
[quickstart.cloudera:21000] > insert into employee (Id, name, age, address, sala
ry) VALUES (1, 'Ramesh',32,'Ahmedabad',20000);
Query: insert into employee (Id, name, age, address, salary) VALUES (1, 'Ramesh'
,32,'Ahmedabad',20000)
Query submitted at: 2024-11-26 18:46:43 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=b9414219e7701ac8:6b77bd4b00000000
Modified 1 row(s) in 4.39s
[quickstart.cloudera:21000] > insert into employee (Id, name, age, address, sala
ry) VALUES (2, 'Khilan',25,'Delhi',15000);
Query: insert into employee (Id, name, age, address, salary) VALUES (2, 'Khilan'
,25,'Delhi',15000)
Query submitted at: 2024-11-26 18:47:36 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=fb447fd7e464fa9e:98c864af00000000
Modified 1 row(s) in 0.11s
```

c. Add some records without specifying attribute names but following the exact order of the attributes in the table:

Insert into employee values (3, 'kaushik', 23, 'Kota', 30000);

Insert into employee values (4, 'Chaitali', 25, 'Mumbai', 35000);

Insert into employee values (5, 'Hardik', 27, 'Bhopal', 40000); Insert into employee values (6, 'Komal', 22, 'MP', 32000);

```
[quickstart.cloudera:21000] > insert into employee values (3,'Kaushik',23,'Kota'
,30000);
Query: insert into employee values (3,'Kaushik',23,'Kota',30000)
Query submitted at: 2024-11-26 18:48:25 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=1d431124593010de:9984e6cc00000000
Modified 1 row(s) in 0.22s
[quickstart.cloudera:21000] > insert into employee values (3,'Kaushik',23,'Kota'
,30000);
Query: insert into employee values (3,'Kaushik',23,'Kota',30000)
Query submitted at: 2024-11-26 18:48:25 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=1d431124593010de:9984e6cc00000000
Modified 1 row(s) in 0.22s
[quickstart.cloudera:21000] > insert into employee values (4,'Chaitali',25,'Mumb
ai',35000); insert into employee values (5,'Hardik',27,'Bhopal',40000); insert i
nto employee values (6,'Komal',22,'MP',32000);
Query: insert into employee values (4,'Chaitali',25,'Mumbai',35000)
Query submitted at: 2024-11-26 18:50:53 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=5c48b97bcdb570e2:9a1bae5d00000000
Modified 1 row(s) in 0.12s
Query: insert into employee values (5,'Hardik',27,'Bhopal',40000)
Query submitted at: 2024-11-26 18:50:53 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=7a4dbeebc202351d:87c4db6600000000
Modified 1 row(s) in 0.11s
Query: insert into employee values (6,'Komal',22,'MP',32000)
Query submitted at: 2024-11-26 18:50:53 (Coordinator: http://quickstart.cloudera
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=8042e2a28beface3:d682971700000000
Modified 1 row(s) in 0.11s
```

14. Overwriting a record in a table using overwrite clause: syntax: **Insert overwrite table_name values (value1, value2, value2);** Example: **Insert overwrite employee values (1, 'Ram', 26, 'Vishakhapatnam', 37000);**

```
[quickstart.cloudera:21000] > insert overwrite employee values (1,'Ram',26,'Vish
akhapatnam',37000);
Query: insert overwrite employee values (1,'Ram',26,'Vishakhapatnam',37000)
Query submitted at: 2024-11-26 18:54:02 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=124c8138b0f60dd9:b4603bc900000000
Modified 1 row(s) in 0.21s
```

Other queries in Impala

15. Select some records:

    Select with specifying attribute names:

    Syntax: **SELECT column1, column2, columnN from table_name;**

    Example: **SELECT ID, Name, AGE FROM employee;**

```
[quickstart.cloudera:21000] > SELECT Id, name, age FROM employee;
Query: select Id, name, age FROM employee
Query submitted at: 2024-11-26 23:24:42 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=b143f5110a328fc1:3d0a5cc900000000
+----+------+-----+
| id | name | age |
+----+------+-----+
| 1  | Ram  | 26  |
+----+------+-----+
Fetched 1 row(s) in 0.65s
```

Select without specifying attribute names:

Syntax: **SELECT * FROM table_name;**

Example: **SELECT * FROM employee;**

```
[quickstart.cloudera:21000] > SELECT * FROM employee;
Query: select * FROM employee
Query submitted at: 2024-11-26 23:25:46 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=124930c56561d272:f8017cc300000000
+----+------+-----+----------------+--------+
| id | name | age | address        | salary |
+----+------+-----+----------------+--------+
| 1  | Ram  | 26  | Vishakhapatnam | 37000  |
+----+------+-----+----------------+--------+
Fetched 1 row(s) in 0.28s
```

Check and see if the syntaxes are case sensitive or not and add your findings in your report.

```
[quickstart.cloudera:21000] > select * from employee;
Query: select * from employee
Query submitted at: 2024-11-26 23:26:38 (Coordinator: http://quickstart.cloudera
:25000)
Query progress can be monitored at: http://quickstart.cloudera:25000/query_plan?
query_id=a144505f098e8e53:5ad0d75200000000
+----+------+-----+----------------+--------+
| id | name | age | address        | salary |
+----+------+-----+----------------+--------+
| 1  | Ram  | 26  | Vishakhapatnam | 37000  |
+----+------+-----+----------------+--------+
Fetched 1 row(s) in 0.26s
```

It is not case sensitive as the syntax 'select * from employee;' still works and produces the same output as 'SELECT * FROM employee;' regardless of it being lowercase.

16. The describe statement in Impala gives some information about the table such as the column names and their data types.

    Syntax: **describe table_name;** Example:

    describe employee;

```
[quickstart.cloudera:21000] > describe employee;
Query: describe employee
+---------+--------+---------+
| name    | type   | comment |
+---------+--------+---------+
| id      | int    |         |
| name    | string |         |
| age     | int    |         |
| address | string |         |
| salary  | bigint |         |
+---------+--------+---------+
Fetched 5 row(s) in 0.19s
```

17. You can alter a table:

    a. Syntax: **ALTER TABLE [old_db_name.]old_table_name RENAME TO [new_db_name.]new_table_name;**

       Example: **ALTER TABLE my_ retailerdb.employee RENAME TO retailerdb.users;**

       Verify changes: **Show tables;**

```
[quickstart.cloudera:21000] > ALTER TABLE retaildb.employee RENAME TO retaildb.u
sers;
Query: alter TABLE retaildb.employee RENAME TO retaildb.users

Fetched 0 row(s) in 0.29s
[quickstart.cloudera:21000] > show tables;
Query: show tables
+-----------+
| name      |
+-----------+
| retailers |
| users     |
+-----------+
Fetched 2 row(s) in 0.11s
```

b. Add new columns:

Syntax: **ALTER TABLE name ADD COLUMNS (col_spec[, col_spec ...])**

Example: **ALTER TABLE users ADD COLUMNS (account_no BIGINT, phone_no BIGINT);**

```
[quickstart.cloudera:21000] > ALTER TABLE users ADD COLUMNS(account_no BIGINT, p
hone_no BIGINT);
Query: alter TABLE users ADD COLUMNS(account_no BIGINT, phone_no BIGINT)

Fetched 0 row(s) in 5.32s
[quickstart.cloudera:21000] > describe users;
Query: describe users
+------------+--------+---------+
| name       | type   | comment |
+------------+--------+---------+
| id         | int    |         |
| name       | string |         |
| age        | int    |         |
| address    | string |         |
| salary     | bigint |         |
| account_no | bigint |         |
| phone_no   | bigint |         |
+------------+--------+---------+
Fetched 7 row(s) in 0.01s
```

c. Dop a column:

Syntax: **ALTER TABLE name DROP [COLUMN] column_name**

Example: **ALTER TABLE users DROP account_no;**

Verify the changes: **describe users;**

```
[quickstart.cloudera:21000] > ALTER TABLE users DROP account_no;
Query: alter TABLE users DROP account_no

Fetched 0 row(s) in 0.43s
[quickstart.cloudera:21000] > describe users;
Query: describe users
+----------+--------+---------+
| name     | type   | comment |
+----------+--------+---------+
| id       | int    |         |
| name     | string |         |
| age      | int    |         |
| address  | string |         |
| salary   | bigint |         |
| phone_no | bigint |         |
+----------+--------+---------+
Fetched 6 row(s) in 0.11s
```

d. Changing the name and type of a column:

Syntax: **ALTER TABLE name CHANGE column_name new_name new_type;**

Example: **ALTER TABLE users CHANGE phone_no e_mail string;**

```
[quickstart.cloudera:21000] > ALTER TABLE users CHANGE phone_no e_mail STRING;
Query: alter TABLE users CHANGE phone_no e_mail STRING

Fetched 0 row(s) in 0.38s
[quickstart.cloudera:21000] > describe users;
Query: describe users
+---------+--------+---------+
| name    | type   | comment |
+---------+--------+---------+
| id      | int    |         |
| name    | string |         |
| age     | int    |         |
| address | string |         |
| salary  | bigint |         |
| e_mail  | string |         |
+---------+--------+---------+
Fetched 6 row(s) in 0.11s
```

18. You can drop a table using the command:

Syntax: **DROP table database_name.table_name;**

Example: **drop table if exists retaildb.retailers;** Verify the changes:

use retaildb; show tables;

```
[quickstart.cloudera:21000] > DROP TABLE IF EXISTS retaildb.retailers;
Query: drop TABLE IF EXISTS retaildb.retailers
[quickstart.cloudera:21000] > use retaildb;
Query: use retaildb
[quickstart.cloudera:21000] > show tables;
Query: show tables
+-------+
| name  |
+-------+
| users |
+-------+
Fetched 1 row(s) in 0.11s
```