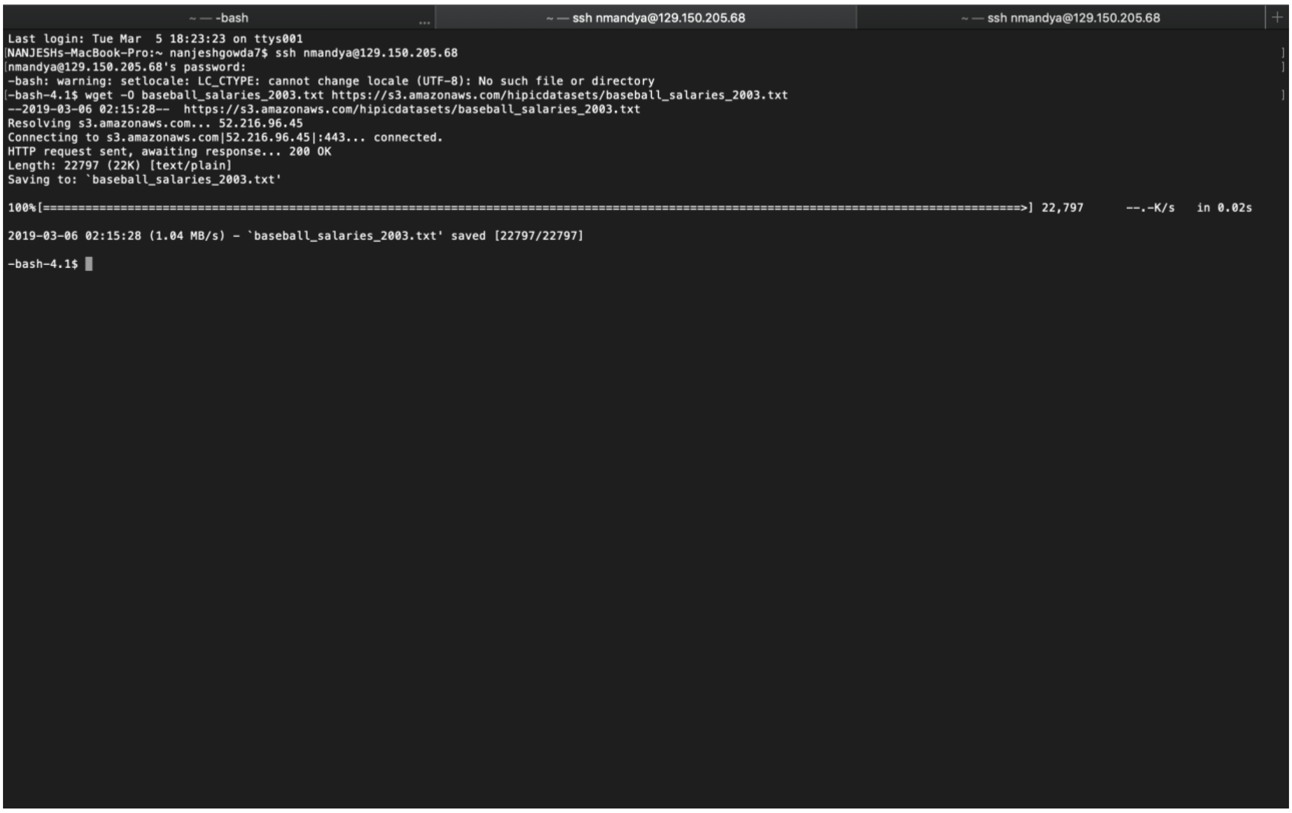
CIS-5200 MID-TERM EXAM-1
Prof. Jongwook Woo

Submitted by: Nanjesh Ramesh
CIN: 307402152

Question 1:

Downloading the file to my Oracle Cloud server using wget command.

wget -O baseball_salaries_2003.txt https://s3.amazonaws.com/hipicdatasets/baseball_salaries_2003.txt


~ — -bash
~ — ssh nmandya@129.150.205.68
~ — ssh nmandya@129.150.205.68

```
Last login: Tue Mar  5 18:23:23 on ttys001
NANJESHs-MacBook-Pro:~ nanjeshgowda7$ ssh nmandya@129.150.205.68
nmandya@129.150.205.68's password:
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
-bash-4.1$ wget -O baseball_salaries_2003.txt https://s3.amazonaws.com/hipicdatasets/baseball_salaries_2003.txt
--2019-03-06 02:15:28--  https://s3.amazonaws.com/hipicdatasets/baseball_salaries_2003.txt
Resolving s3.amazonaws.com... 52.216.96.45
Connecting to s3.amazonaws.com|52.216.96.45|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 22797 (22K) [text/plain]
Saving to: `baseball_salaries_2003.txt'

100%[========================================================================================================================>] 22,797      --.-K/s   in 0.02s

2019-03-06 02:15:28 (1.04 MB/s) - `baseball_salaries_2003.txt' saved [22797/22797]

-bash-4.1$ 
```

Question 2:

Uploading the file using hdfs command to the HDFS directory “baseball_salaries” of Oracle Cloud.

Changing the permission of the file.

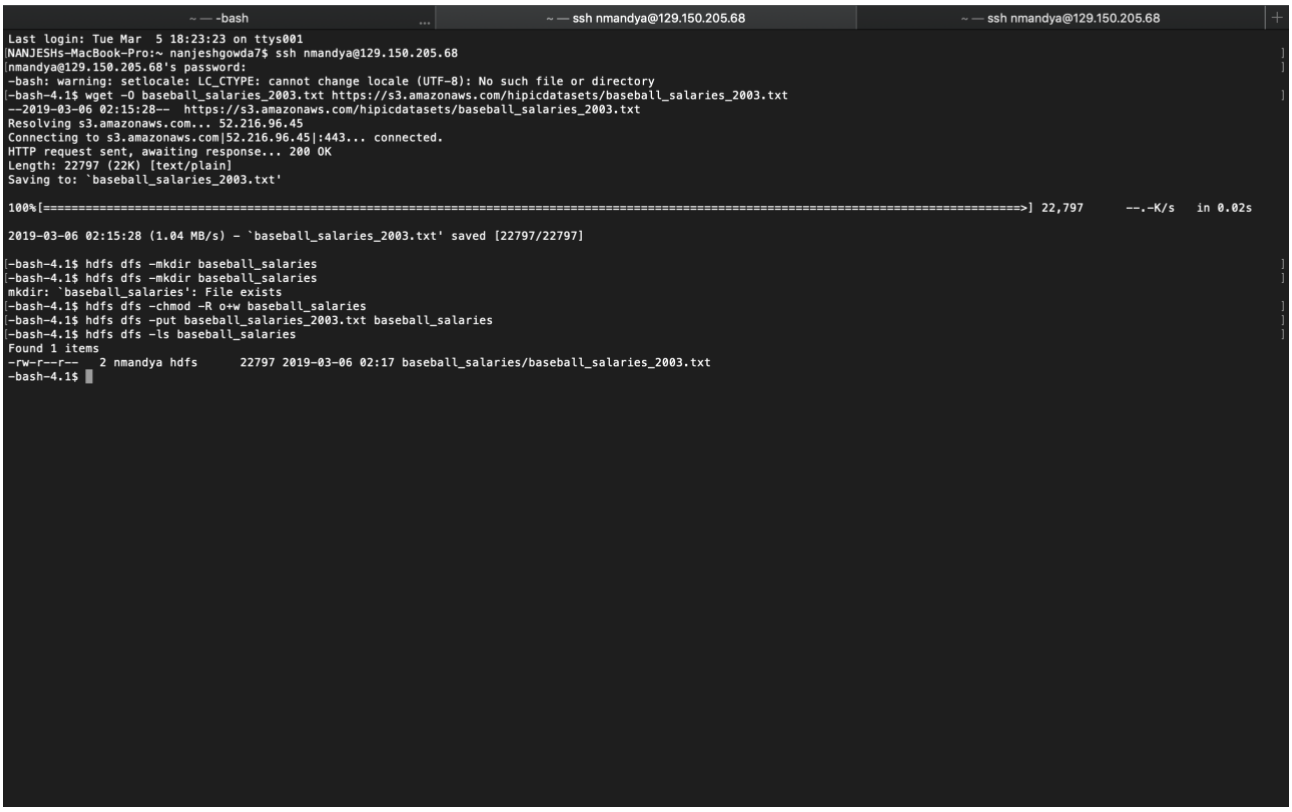
Displaying the result.

hdfs dfs -mkdir baseball_salaries

hdfs dfs -chmod -R o+w baseball_salaries

hdfs dfs -put baseball_salaries_2003.txt baseball_salaries

hdfs dfs -ls baseball_salaries

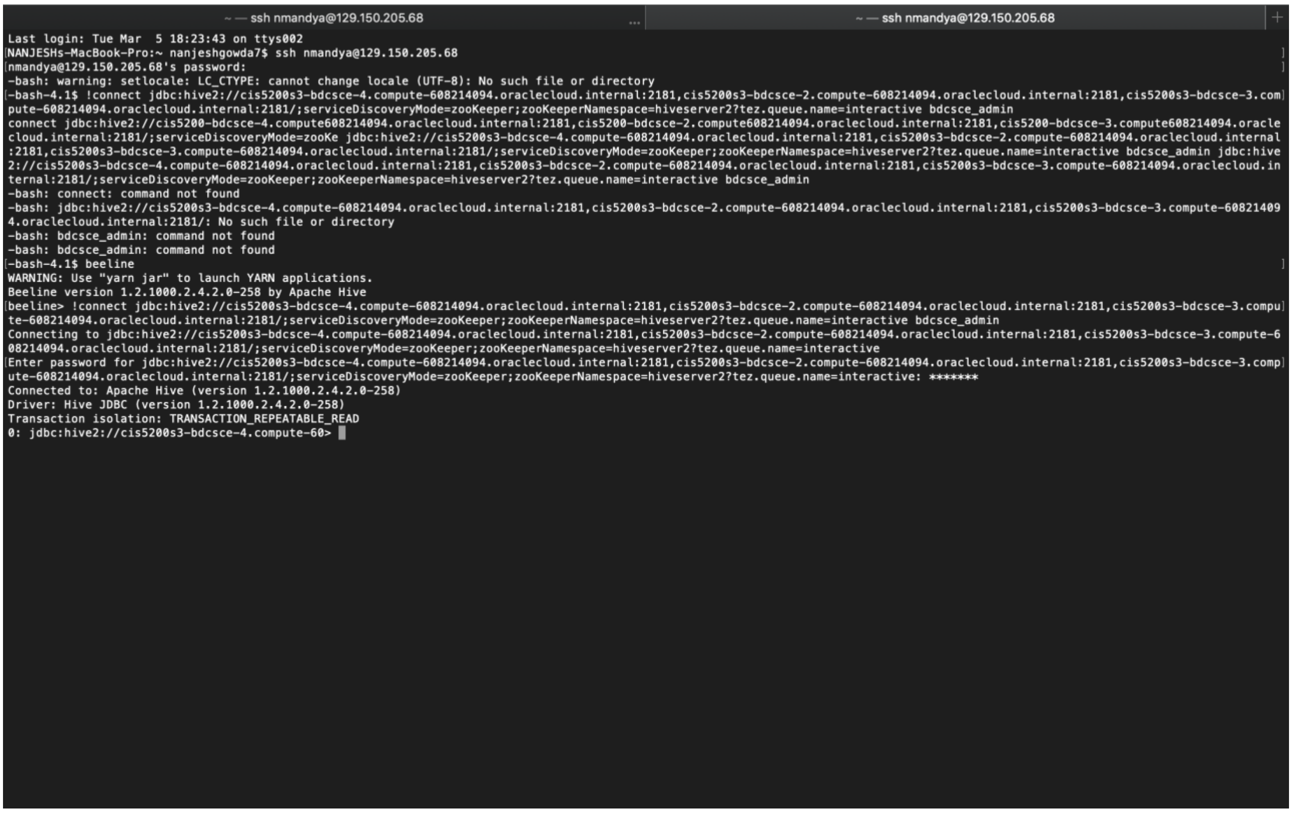

~ — -bash
~ — ssh nmandya@129.150.205.68
~ — ssh nmandya@129.150.205.68

```
Last login: Tue Mar  5 18:23:23 on ttys001
[NANJESHs-MacBook-Pro:~ nanjeshgowda7$ ssh nmandya@129.150.205.68
[nmandya@129.150.205.68's password:
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
[-bash-4.1$ wget -O baseball_salaries_2003.txt https://s3.amazonaws.com/hipicdatasets/baseball_salaries_2003.txt
--2019-03-06 02:15:28--  https://s3.amazonaws.com/hipicdatasets/baseball_salaries_2003.txt
Resolving s3.amazonaws.com... 52.216.96.45
Connecting to s3.amazonaws.com|52.216.96.45|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 22797 (22K) [text/plain]
Saving to: `baseball_salaries_2003.txt'

100%[=====================================================================================================================>] 22,797      --.-K/s   in 0.02s

2019-03-06 02:15:28 (1.04 MB/s) - `baseball_salaries_2003.txt' saved [22797/22797]

[-bash-4.1$ hdfs dfs -mkdir baseball_salaries
[-bash-4.1$ hdfs dfs -mkdir baseball_salaries
mkdir: `baseball_salaries': File exists
[-bash-4.1$ hdfs dfs -chmod -R o+w baseball_salaries
[-bash-4.1$ hdfs dfs -put baseball_salaries_2003.txt baseball_salaries
[-bash-4.1$ hdfs dfs -ls baseball_salaries
Found 1 items
-rw-r--r--   2 nmandya hdfs      22797 2019-03-06 02:17 baseball_salaries/baseball_salaries_2003.txt
-bash-4.1$ 
```

Connecting to beeline.

beeline

!connect jdbc:hive2://cis5200-bdcsce-4.compute-608214094.oraclecloud.internal:2181,cis5200-bdcsce-2.compute-608214094.oraclecloud.internal:2181,cis5200-bdcsce-3.compute-608214094.oraclecloud.internal:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive bdcsce_admin


~ — ssh nmandya@129.150.205.68
~ — ssh nmandya@129.150.205.68

```
Last login: Tue Mar  5 18:23:43 on ttys002
NANJESHs-MacBook-Pro:~ nanjeshgowda7$ ssh nmandya@129.150.205.68
nmandya@129.150.205.68's password:
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
-bash-4.1$ !connect jdbc:hive2://cis5200s3-bdcsce-4.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-2.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-3.com
pute-608214094.oraclecloud.internal:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive bdcsce_admin
connect jdbc:hive2://cis5200-bdcsce-4.compute608214094.oraclecloud.internal:2181,cis5200-bdcsce-2.compute608214094.oraclecloud.internal:2181,cis5200-bdcsce-3.compute608214094.oracle
cloud.internal:2181/;serviceDiscoveryMode=zooKe jdbc:hive2://cis5200s3-bdcsce-4.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-2.compute-608214094.oraclecloud.internal
:2181,cis5200s3-bdcsce-3.compute-608214094.oraclecloud.internal:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive bdcsce_admin jdbc:hive
2://cis5200s3-bdcsce-4.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-2.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-3.compute-608214094.oraclecloud.in
ternal:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive bdcsce_admin
-bash: connect: command not found
-bash: jdbc:hive2://cis5200s3-bdcsce-4.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-2.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-3.compute-60821409
4.oraclecloud.internal:2181/: No such file or directory
-bash: bdcsce_admin: command not found
-bash: bdcsce_admin: command not found
-bash-4.1$ beeline
WARNING: Use "yarn jar" to launch YARN applications.
Beeline version 1.2.1000.2.4.2.0-258 by Apache Hive
beeline> !connect jdbc:hive2://cis5200s3-bdcsce-4.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-2.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-3.compu
te-608214094.oraclecloud.internal:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive bdcsce_admin
Connecting to jdbc:hive2://cis5200s3-bdcsce-4.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-2.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-3.compute-6
08214094.oraclecloud.internal:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive
Enter password for jdbc:hive2://cis5200s3-bdcsce-4.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-2.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-3.comp
ute-608214094.oraclecloud.internal:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive: *******
Connected to: Apache Hive (version 1.2.1000.2.4.2.0-258)
Driver: Hive JDBC (version 1.2.1000.2.4.2.0-258)
Transaction isolation: TRANSACTION_REPEATABLE_READ
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60>
```

Question 3:

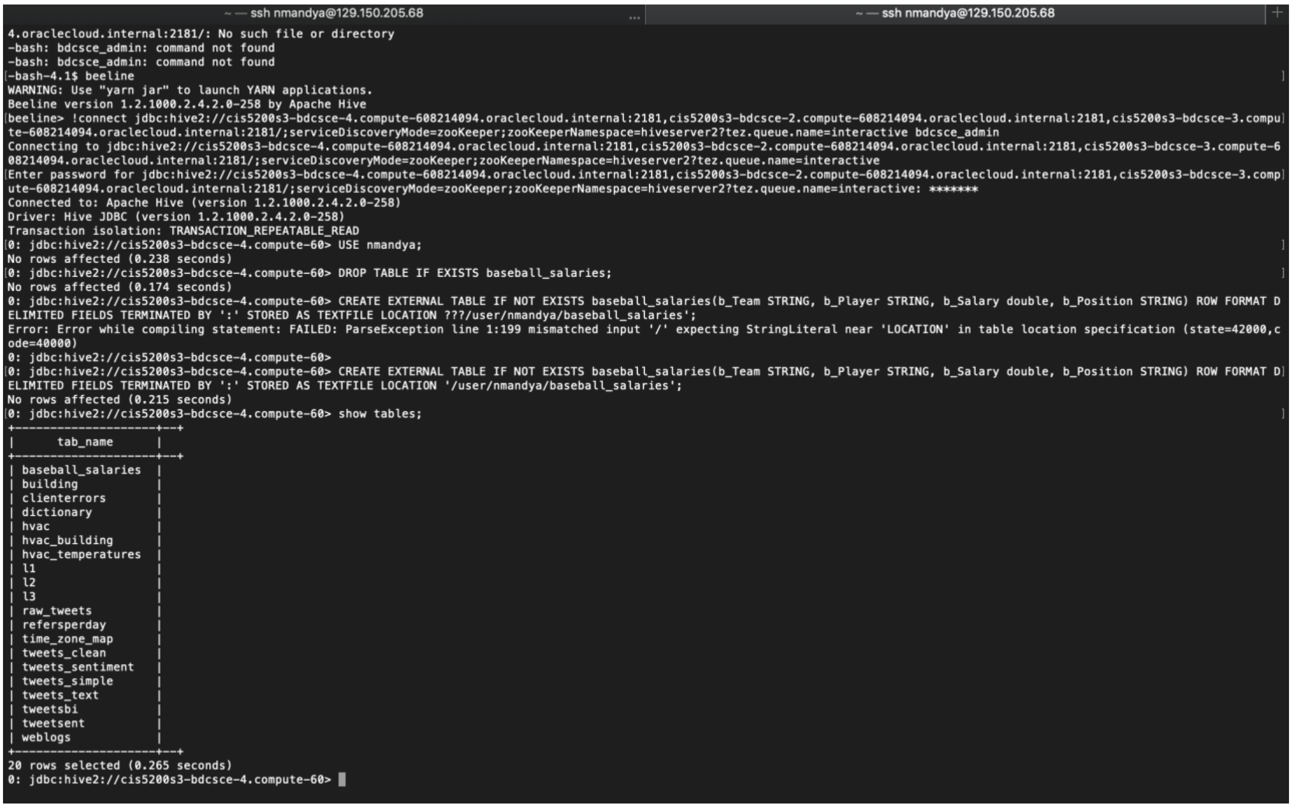
Creating table baseball_salaries by using my Database nmandya.

USE nmandya;

DROP TABLE IF EXISTS baseball_salaries;

CREATE EXTERNAL TABLE IF NOT EXISTS baseball_salaries(b_Team STRING, b_Player STRING, b_Salary double, b_Position STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ':' STORED AS TEXTFILE LOCATION ‘/user/nmandya/baseball_salaries';

show tables;

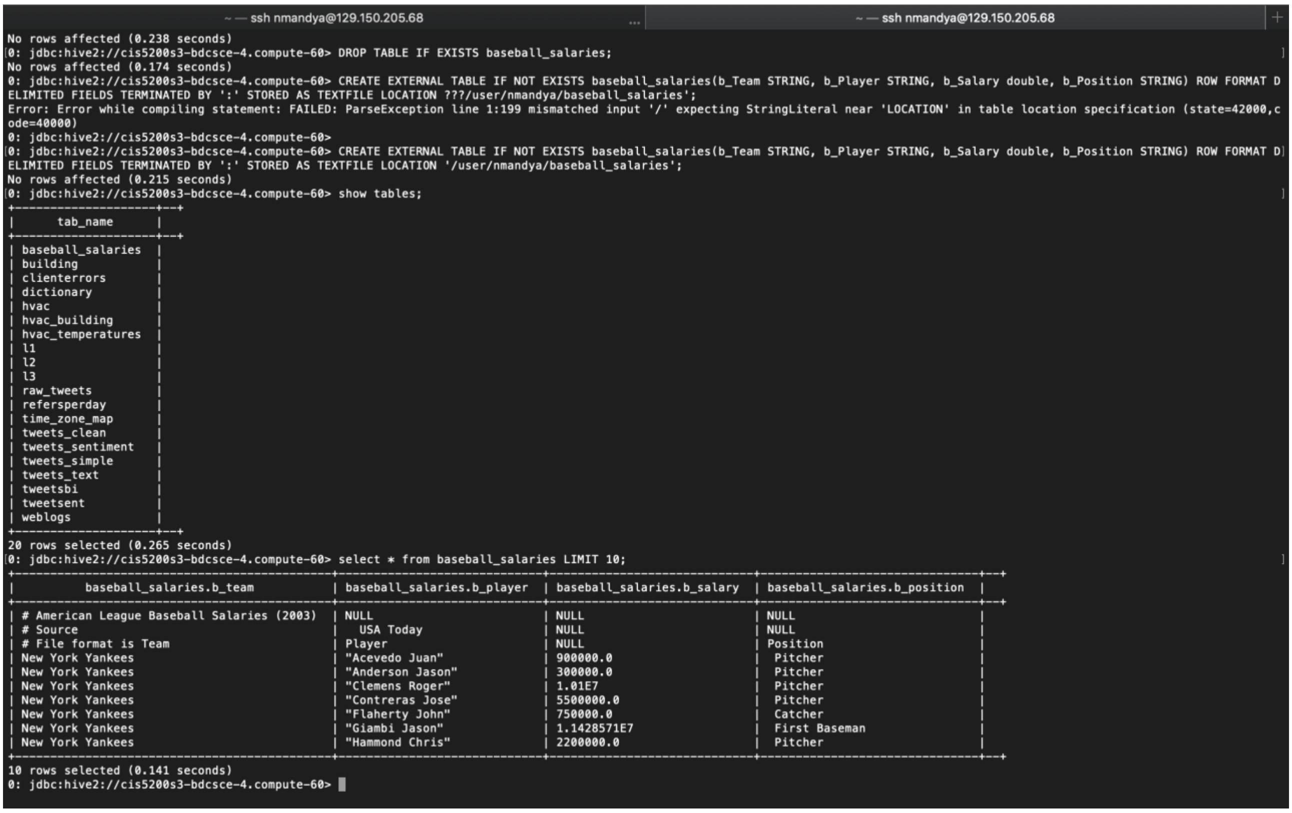

~ — ssh nmandya@129.150.205.68
~ — ssh nmandya@129.150.205.68

```
4.oraclecloud.internal:2181/: No such file or directory
-bash: bdcsce_admin: command not found
-bash: bdcsce_admin: command not found
[-bash-4.1$ beeline
WARNING: Use "yarn jar" to launch YARN applications.
Beeline version 1.2.1000.2.4.2.0-258 by Apache Hive
[beeline> !connect jdbc:hive2://cis5200s3-bdcsce-4.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-2.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-3.compu]
te-608214094.oraclecloud.internal:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive bdcsce_admin
Connecting to jdbc:hive2://cis5200s3-bdcsce-4.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-2.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-3.compute-6
08214094.oraclecloud.internal:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive
[Enter password for jdbc:hive2://cis5200s3-bdcsce-4.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-2.compute-608214094.oraclecloud.internal:2181,cis5200s3-bdcsce-3.comp]
ute-608214094.oraclecloud.internal:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive: *******
Connected to: Apache Hive (version 1.2.1000.2.4.2.0-258)
Driver: Hive JDBC (version 1.2.1000.2.4.2.0-258)
Transaction isolation: TRANSACTION_REPEATABLE_READ
[0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> USE nmandya;
No rows affected (0.238 seconds)
[0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> DROP TABLE IF EXISTS baseball_salaries;
No rows affected (0.174 seconds)
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> CREATE EXTERNAL TABLE IF NOT EXISTS baseball_salaries(b_Team STRING, b_Player STRING, b_Salary double, b_Position STRING) ROW FORMAT D
ELIMITED FIELDS TERMINATED BY ':' STORED AS TEXTFILE LOCATION ???/user/nmandya/baseball_salaries';
Error: Error while compiling statement: FAILED: ParseException line 1:199 mismatched input '/' expecting StringLiteral near 'LOCATION' in table location specification (state=42000,c
ode=40000)
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60>
[0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> CREATE EXTERNAL TABLE IF NOT EXISTS baseball_salaries(b_Team STRING, b_Player STRING, b_Salary double, b_Position STRING) ROW FORMAT D]
ELIMITED FIELDS TERMINATED BY ':' STORED AS TEXTFILE LOCATION '/user/nmandya/baseball_salaries';
No rows affected (0.215 seconds)
[0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> show tables;
+--------------------+--+
|      tab_name      |
+--------------------+--+
| baseball_salaries  |
| building           |
| clienterrors       |
| dictionary         |
| hvac               |
| hvac_building      |
| hvac_temperatures  |
| l1                 |
| l2                 |
| l3                 |
| raw_tweets         |
| refersperday       |
| time_zone_map      |
| tweets_clean       |
| tweets_sentiment   |
| tweets_simple      |
| tweets_text        |
| tweetsbi           |
| tweetsent          |
| weblogs            |
+--------------------+--+
20 rows selected (0.265 seconds)
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> ▌
```

Question 4:

HiveQL to query 10 data from table "baseball_salaries".

select * from baseball_salaries LIMIT 10;



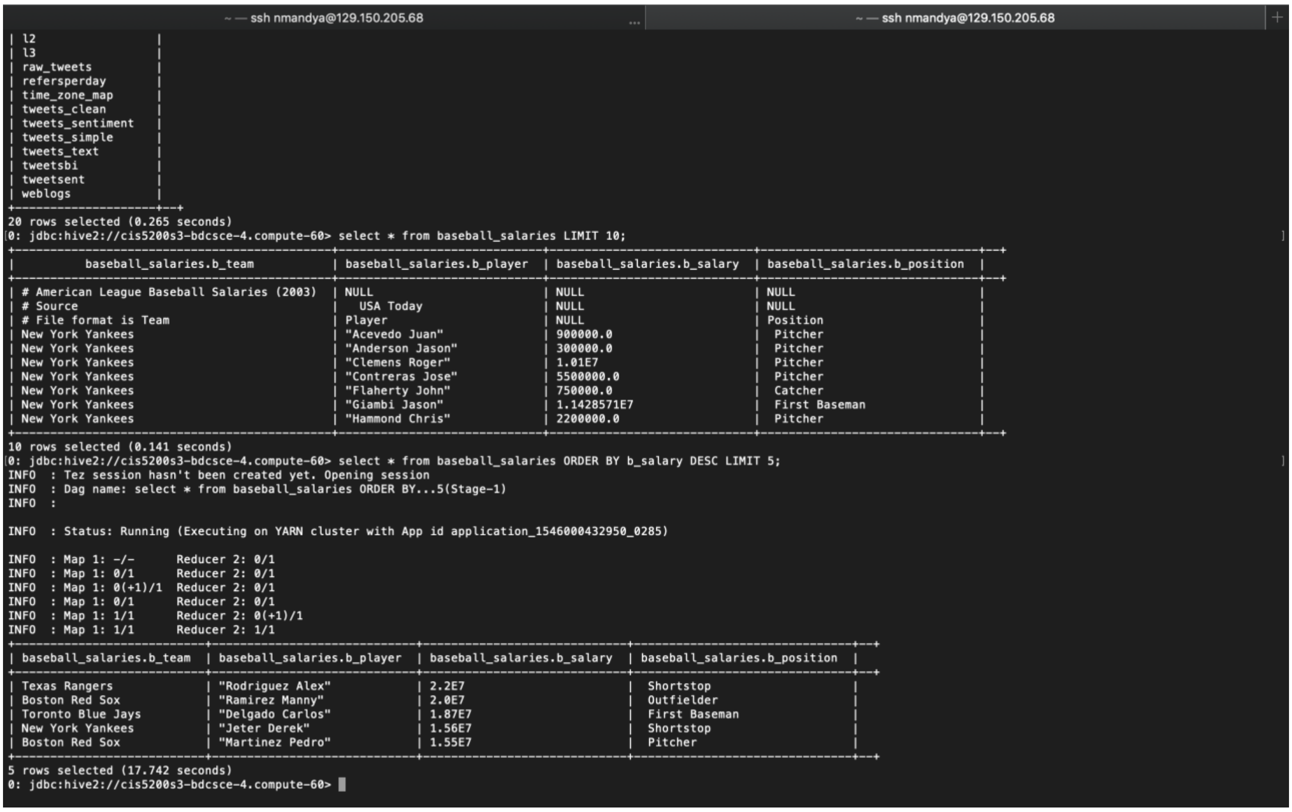
~ — ssh nmandya@129.150.205.68
~ — ssh nmandya@129.150.205.68

```
No rows affected (0.238 seconds)
[0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> DROP TABLE IF EXISTS baseball_salaries;                                                        ]
No rows affected (0.174 seconds)
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> CREATE EXTERNAL TABLE IF NOT EXISTS baseball_salaries(b_Team STRING, b_Player STRING, b_Salary double, b_Position STRING) ROW FORMAT D
ELIMITED FIELDS TERMINATED BY ':' STORED AS TEXTFILE LOCATION ???/user/nmandya/baseball_salaries';
Error: Error while compiling statement: FAILED: ParseException line 1:199 mismatched input '/' expecting StringLiteral near 'LOCATION' in table location specification (state=42000,c
ode=40000)
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60>
[0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> CREATE EXTERNAL TABLE IF NOT EXISTS baseball_salaries(b_Team STRING, b_Player STRING, b_Salary double, b_Position STRING) ROW FORMAT D]
ELIMITED FIELDS TERMINATED BY ':' STORED AS TEXTFILE LOCATION '/user/nmandya/baseball_salaries';
No rows affected (0.215 seconds)
[0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> show tables;                                                                                   ]
+--------------------+--+
|      tab_name      |
+--------------------+--+
| baseball_salaries  |
| building           |
| clienterrors       |
| dictionary         |
| hvac               |
| hvac_building      |
| hvac_temperatures  |
| l1                 |
| l2                 |
| l3                 |
| raw_tweets         |
| refersperday       |
| time_zone_map      |
| tweets_clean       |
| tweets_sentiment   |
| tweets_simple      |
| tweets_text        |
| tweetsbi           |
| tweetsent          |
| weblogs            |
+--------------------+--+
20 rows selected (0.265 seconds)
[0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> select * from baseball_salaries LIMIT 10;                                                      ]
+---------------------------------------------+-----------------------------+-----------------------------+-------------------------------+--+
|          baseball_salaries.b_team           | baseball_salaries.b_player  | baseball_salaries.b_salary  | baseball_salaries.b_position  |
+---------------------------------------------+-----------------------------+-----------------------------+-------------------------------+--+
| # American League Baseball Salaries (2003)  | NULL                        | NULL                        | NULL                          |
| # Source                                    |   USA Today                 | NULL                        | NULL                          |
| # File format is Team                       | Player                      | NULL                        | Position                      |
| New York Yankees                            | "Acevedo Juan"              | 900000.0                    |  Pitcher                      |
| New York Yankees                            | "Anderson Jason"            | 300000.0                    |  Pitcher                      |
| New York Yankees                            | "Clemens Roger"             | 1.01E7                      |  Pitcher                      |
| New York Yankees                            | "Contreras Jose"            | 5500000.0                   |  Pitcher                      |
| New York Yankees                            | "Flaherty John"             | 750000.0                    |  Catcher                      |
| New York Yankees                            | "Giambi Jason"              | 1.1428571E7                 |  First Baseman                |
| New York Yankees                            | "Hammond Chris"             | 2200000.0                   |  Pitcher                      |
+---------------------------------------------+-----------------------------+-----------------------------+-------------------------------+--+
10 rows selected (0.141 seconds)
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> █
```

Question 5:

HiveQL to query top 5 data ordered by the highest “Salary” column from the table “baseball_salaries”

select * from baseball_salaries ORDER BY b_salary DESC LIMIT 5;

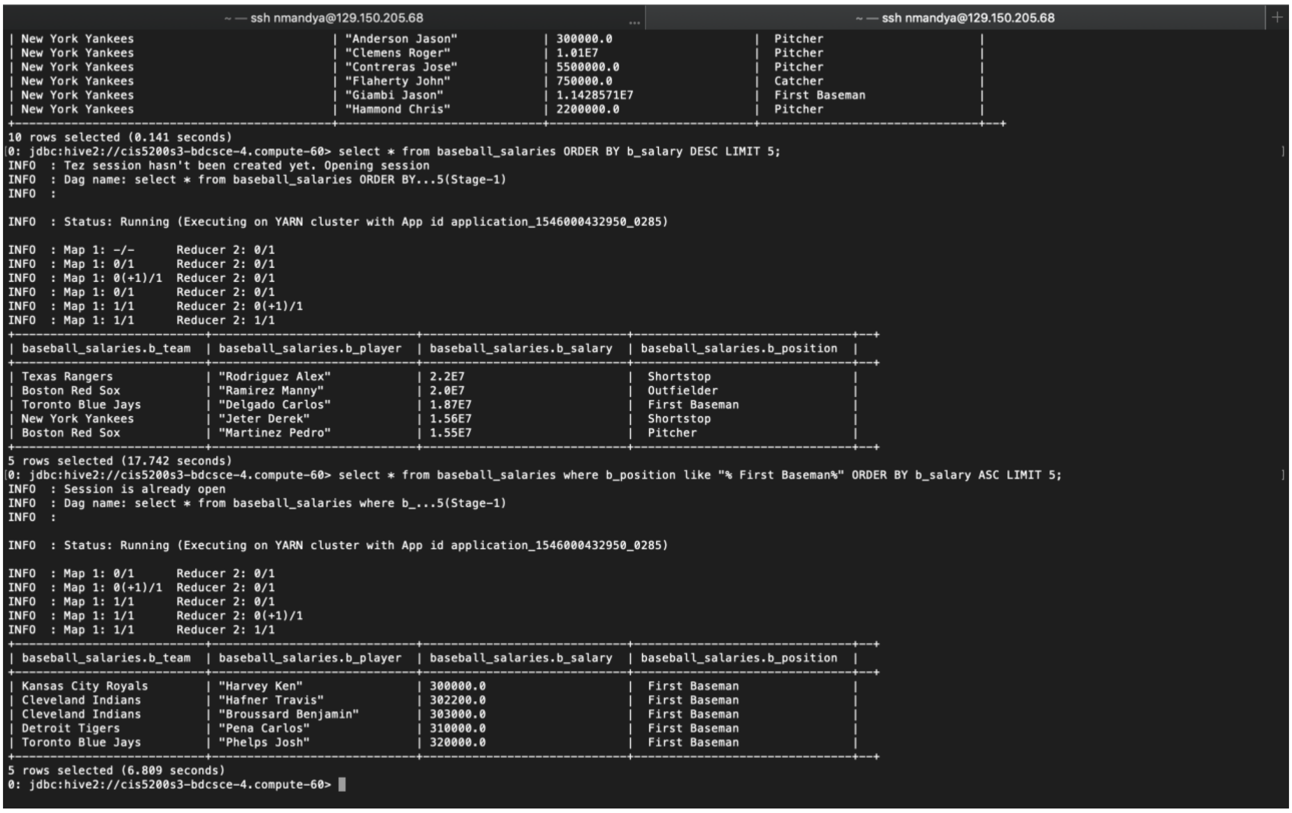

~ — ssh nmandya@129.150.205.68
~ — ssh nmandya@129.150.205.68

```
| l2                  |
| l3                  |
| raw_tweets          |
| refersperday        |
| time_zone_map       |
| tweets_clean        |
| tweets_sentiment    |
| tweets_simple       |
| tweets_text         |
| tweetsbi            |
| tweetsent           |
| weblogs             |
+---------------------+--+
20 rows selected (0.265 seconds)
[0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> select * from baseball_salaries LIMIT 10;
+---------------------------------------------+-----------------------------+-----------------------------+-------------------------------+--+
|         baseball_salaries.b_team            | baseball_salaries.b_player  | baseball_salaries.b_salary  | baseball_salaries.b_position  |
+---------------------------------------------+-----------------------------+-----------------------------+-------------------------------+--+
| # American League Baseball Salaries (2003)  | NULL                        | NULL                        | NULL                          |
| # Source                                    |   USA Today                 | NULL                        | NULL                          |
| # File format is Team                       | Player                      | NULL                        | Position                      |
| New York Yankees                            | "Acevedo Juan"              | 900000.0                    |  Pitcher                      |
| New York Yankees                            | "Anderson Jason"            | 300000.0                    |  Pitcher                      |
| New York Yankees                            | "Clemens Roger"             | 1.01E7                      |  Pitcher                      |
| New York Yankees                            | "Contreras Jose"            | 5500000.0                   |  Pitcher                      |
| New York Yankees                            | "Flaherty John"             | 750000.0                    |  Catcher                      |
| New York Yankees                            | "Giambi Jason"              | 1.1428571E7                 |  First Baseman                |
| New York Yankees                            | "Hammond Chris"             | 2200000.0                   |  Pitcher                      |
+---------------------------------------------+-----------------------------+-----------------------------+-------------------------------+--+
10 rows selected (0.141 seconds)
[0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> select * from baseball_salaries ORDER BY b_salary DESC LIMIT 5;
INFO  : Tez session hasn't been created yet. Opening session
INFO  : Dag name: select * from baseball_salaries ORDER BY...5(Stage-1)
INFO  :

INFO  : Status: Running (Executing on YARN cluster with App id application_1546000432950_0285)

INFO  : Map 1: -/-      Reducer 2: 0/1
INFO  : Map 1: 0/1      Reducer 2: 0/1
INFO  : Map 1: 0(+1)/1  Reducer 2: 0/1
INFO  : Map 1: 0/1      Reducer 2: 0/1
INFO  : Map 1: 1/1      Reducer 2: 0(+1)/1
INFO  : Map 1: 1/1      Reducer 2: 1/1
+---------------------------+-----------------------------+-----------------------------+-------------------------------+--+
| baseball_salaries.b_team  | baseball_salaries.b_player  | baseball_salaries.b_salary  | baseball_salaries.b_position  |
+---------------------------+-----------------------------+-----------------------------+-------------------------------+--+
| Texas Rangers             | "Rodriguez Alex"            | 2.2E7                       |  Shortstop                    |
| Boston Red Sox            | "Ramirez Manny"             | 2.0E7                       |  Outfielder                   |
| Toronto Blue Jays         | "Delgado Carlos"            | 1.87E7                      |  First Baseman                |
| New York Yankees          | "Jeter Derek"               | 1.56E7                      |  Shortstop                    |
| Boston Red Sox            | "Martinez Pedro"            | 1.55E7                      |  Pitcher                      |
+---------------------------+-----------------------------+-----------------------------+-------------------------------+--+
5 rows selected (17.742 seconds)
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> 
```

Question 6:

HiveQL of the lowest 5 data ordered by the “Salary” column from the table at the position of “ First Baseman”.

select * from baseball_salaries where b_position like "% First Baseman%" ORDER BY b_salary ASC LIMIT 5;

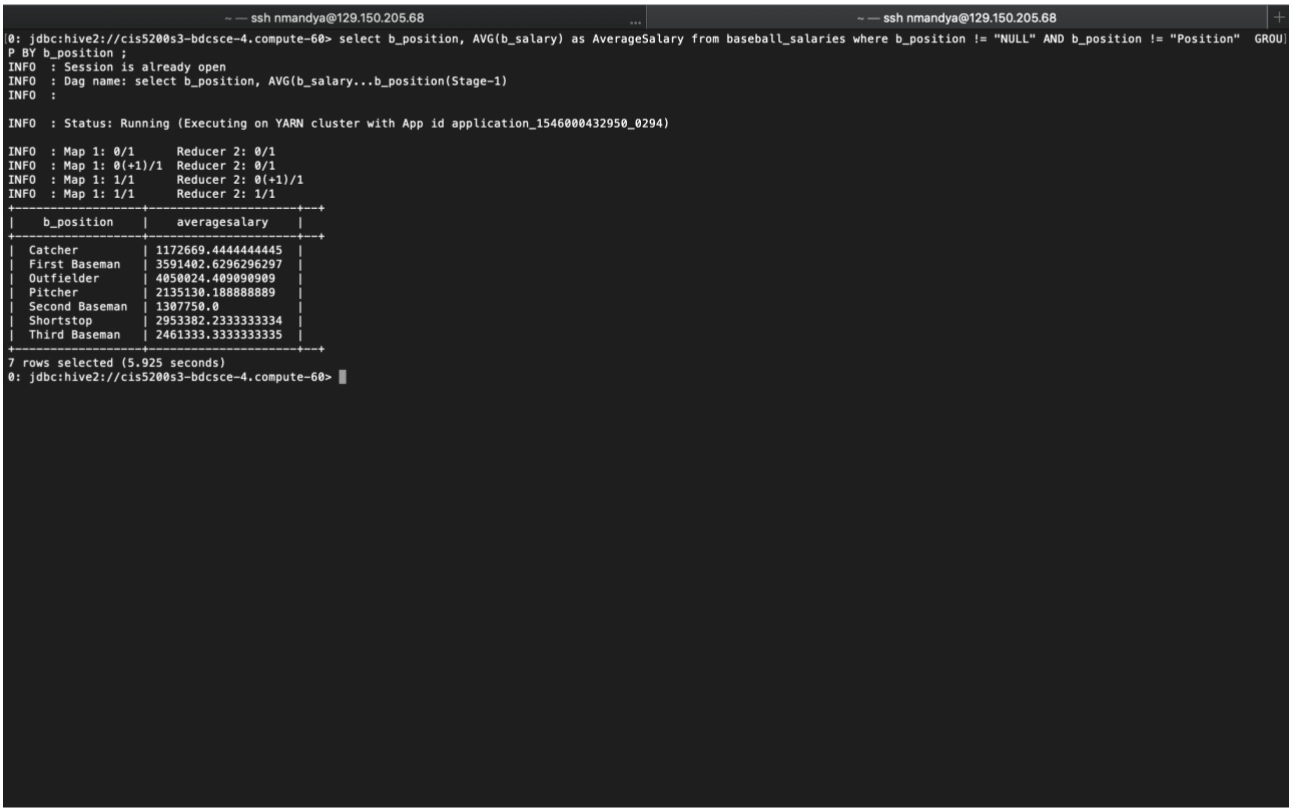

~ — ssh nmandya@129.150.205.68
~ — ssh nmandya@129.150.205.68

```
| New York Yankees                          | "Anderson Jason"          | 300000.0                      |  Pitcher                      |
| New York Yankees                          | "Clemens Roger"           | 1.01E7                        |  Pitcher                      |
| New York Yankees                          | "Contreras Jose"          | 5500000.0                     |  Pitcher                      |
| New York Yankees                          | "Flaherty John"           | 750000.0                      |  Catcher                      |
| New York Yankees                          | "Giambi Jason"            | 1.1428571E7                   |  First Baseman                |
| New York Yankees                          | "Hammond Chris"           | 2200000.0                     |  Pitcher                      |
+-------------------------------------------+---------------------------+-------------------------------+-------------------------------+--+
10 rows selected (0.141 seconds)
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> select * from baseball_salaries ORDER BY b_salary DESC LIMIT 5;
INFO  : Tez session hasn't been created yet. Opening session
INFO  : Dag name: select * from baseball_salaries ORDER BY...5(Stage-1)
INFO  :

INFO  : Status: Running (Executing on YARN cluster with App id application_1546000432950_0285)

INFO  : Map 1: -/-      Reducer 2: 0/1
INFO  : Map 1: 0/1      Reducer 2: 0/1
INFO  : Map 1: 0(+1)/1  Reducer 2: 0/1
INFO  : Map 1: 0/1      Reducer 2: 0/1
INFO  : Map 1: 1/1      Reducer 2: 0(+1)/1
INFO  : Map 1: 1/1      Reducer 2: 1/1
+---------------------------+-----------------------------+-----------------------------+-------------------------------+--+
| baseball_salaries.b_team  | baseball_salaries.b_player  | baseball_salaries.b_salary  | baseball_salaries.b_position  |
+---------------------------+-----------------------------+-----------------------------+-------------------------------+--+
| Texas Rangers             | "Rodriguez Alex"            | 2.2E7                       |  Shortstop                    |
| Boston Red Sox            | "Ramirez Manny"             | 2.0E7                       |  Outfielder                   |
| Toronto Blue Jays         | "Delgado Carlos"            | 1.87E7                      |  First Baseman                |
| New York Yankees          | "Jeter Derek"               | 1.56E7                      |  Shortstop                    |
| Boston Red Sox            | "Martinez Pedro"            | 1.55E7                      |  Pitcher                      |
+---------------------------+-----------------------------+-----------------------------+-------------------------------+--+
5 rows selected (17.742 seconds)
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> select * from baseball_salaries where b_position like "% First Baseman%" ORDER BY b_salary ASC LIMIT 5;
INFO  : Session is already open
INFO  : Dag name: select * from baseball_salaries where b_...5(Stage-1)
INFO  :

INFO  : Status: Running (Executing on YARN cluster with App id application_1546000432950_0285)

INFO  : Map 1: 0/1      Reducer 2: 0/1
INFO  : Map 1: 0(+1)/1  Reducer 2: 0/1
INFO  : Map 1: 1/1      Reducer 2: 0/1
INFO  : Map 1: 1/1      Reducer 2: 0(+1)/1
INFO  : Map 1: 1/1      Reducer 2: 1/1
+---------------------------+-----------------------------+-----------------------------+-------------------------------+--+
| baseball_salaries.b_team  | baseball_salaries.b_player  | baseball_salaries.b_salary  | baseball_salaries.b_position  |
+---------------------------+-----------------------------+-----------------------------+-------------------------------+--+
| Kansas City Royals        | "Harvey Ken"                | 300000.0                    |  First Baseman                |
| Cleveland Indians         | "Hafner Travis"             | 302200.0                    |  First Baseman                |
| Cleveland Indians         | "Broussard Benjamin"        | 303000.0                    |  First Baseman                |
| Detroit Tigers            | "Pena Carlos"               | 310000.0                    |  First Baseman                |
| Toronto Blue Jays         | "Phelps Josh"               | 320000.0                    |  First Baseman                |
+---------------------------+-----------------------------+-----------------------------+-------------------------------+--+
5 rows selected (6.809 seconds)
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60>
```

Question 7:

HiveQL to determine the average salary of the players in each position.

select b_position, AVG(b_salary) as AverageSalary from baseball_salaries where b_position != "NULL" AND b_position != "Position" GROUP BY b_position ;



~ — ssh nmandya@129.150.205.68
~ — ssh nmandya@129.150.205.68

```
[0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> select b_position, AVG(b_salary) as AverageSalary from baseball_salaries where b_position != "NULL" AND b_position != "Position"  GROU]
P BY b_position ;
INFO  : Session is already open
INFO  : Dag name: select b_position, AVG(b_salary...b_position(Stage-1)
INFO  :

INFO  : Status: Running (Executing on YARN cluster with App id application_1546000432950_0294)

INFO  : Map 1: 0/1      Reducer 2: 0/1
INFO  : Map 1: 0(+1)/1  Reducer 2: 0/1
INFO  : Map 1: 1/1      Reducer 2: 0(+1)/1
INFO  : Map 1: 1/1      Reducer 2: 1/1
+-----------------+---------------------+--+
|   b_position    |    averagesalary    |
+-----------------+---------------------+--+
|  Catcher        | 1172669.4444444445  |
|  First Baseman  | 3591402.6296296297  |
|  Outfielder     | 4050024.409090909   |
|  Pitcher        | 2135130.188888889   |
|  Second Baseman | 1307750.0           |
|  Shortstop      | 2953382.2333333334  |
|  Third Baseman  | 2461333.3333333335  |
+-----------------+---------------------+--+
7 rows selected (5.925 seconds)
0: jdbc:hive2://cis5200s3-bdcsce-4.compute-60> ▌
```