

BIG DATA ASSIGNMENT-2

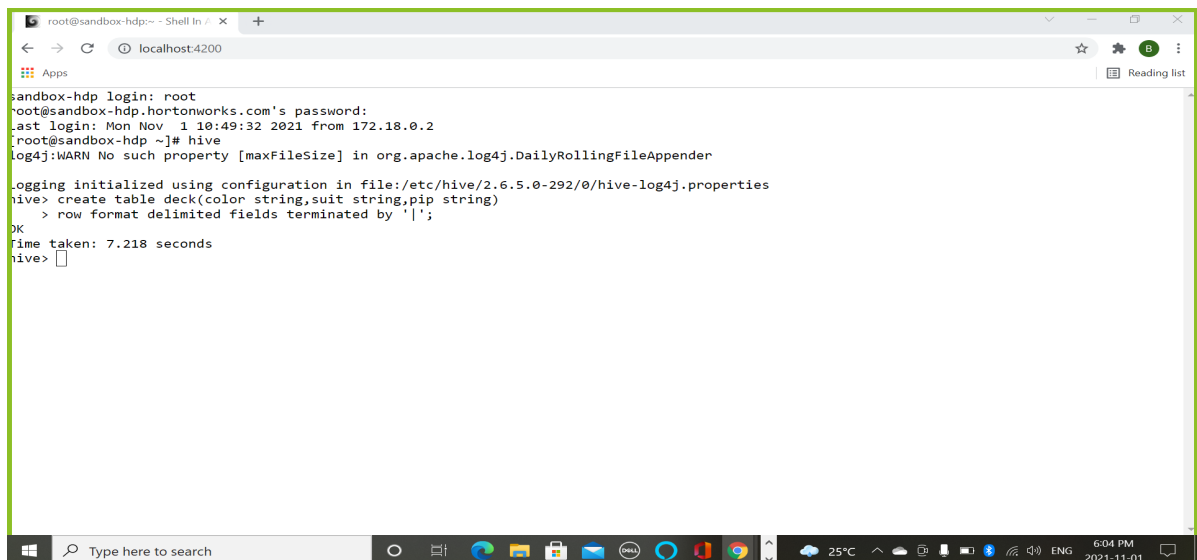
T.Divya Reddy
AP18110010118
CSE-B

Dataset:

BLACK|SPADE|2
BLACK|SPADE|3
BLACK|SPADE|4
BLACK|SPADE|5
BLACK|SPADE|6
BLACK|SPADE|7
BLACK|SPADE|8
.....

1.) Create a managed table and load the data from LFS

Step-1: Create a managed table deck in Hive having column names as color,suit and pip

A screenshot of a terminal window titled 'root@sandbox-hdp:~ - Shell in ~'. The terminal shows the user logging in as root, then running 'hive'. The Hive prompt shows the command 'create table deck(color string,suit string,pip string)' followed by a carriage return. The output indicates the table was created successfully, showing the row format and the time taken (7.218 seconds). The terminal window is running on a Windows 10 desktop environment, with the taskbar visible at the bottom showing various application icons and the system clock at 6:04 PM on 2021-11-01.

```
root@sandbox-hdp login: root
root@sandbox-hdp.hortonworks.com's password:
Last login: Mon Nov 1 10:49:32 2021 from 172.18.0.2
[root@sandbox-hdp ~]# hive
log4j:WARN No such property [maxFileSize] in org.apache.log4j.DailyRollingFileAppender

Logging initialized using configuration in file:/etc/hive/2.6.5.0-292/0/hive-log4j.properties
hive> create table deck(color string,suit string,pip string)
> row format delimited fields terminated by '|';
OK
Time taken: 7.218 seconds
hive>
```

Step-2: Describe the table to know location of it

```
root@sandbox-hdp:~ - Shell In / x +
localhost:4200
Apps
Reading list

hive> describe formatted deck;
OK
# col_name          data_type          comment
color               string
suit                string
pip                 string

# Detailed Table Information
Database:           default
Owner:              root
CreateTime:         Mon Nov 01 12:34:24 UTC 2021
LastAccessTime:     UNKNOWN
Protect Mode:       None
Retention:          0
Location:           hdfs://sandbox-hdp.hortonworks.com:8020/apps/hive/warehouse/deck
Table Type:         MANAGED_TABLE
Table Parameters:
  COLUMN_STATS_ACCURATE {\"BASIC_STATS\": \"true\"}
  numFiles              0
  numRows              0
  rawDataSize          0
  totalSize            0
  transient_lastDdlTime 1635770064

# Storage Information
SerDe Library:      org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat:        org.apache.hadoop.mapred.TextInputFormat
OutputFormat:       org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
Compressed:         No
```

Step-3: Load cards data into the table and display data using select command

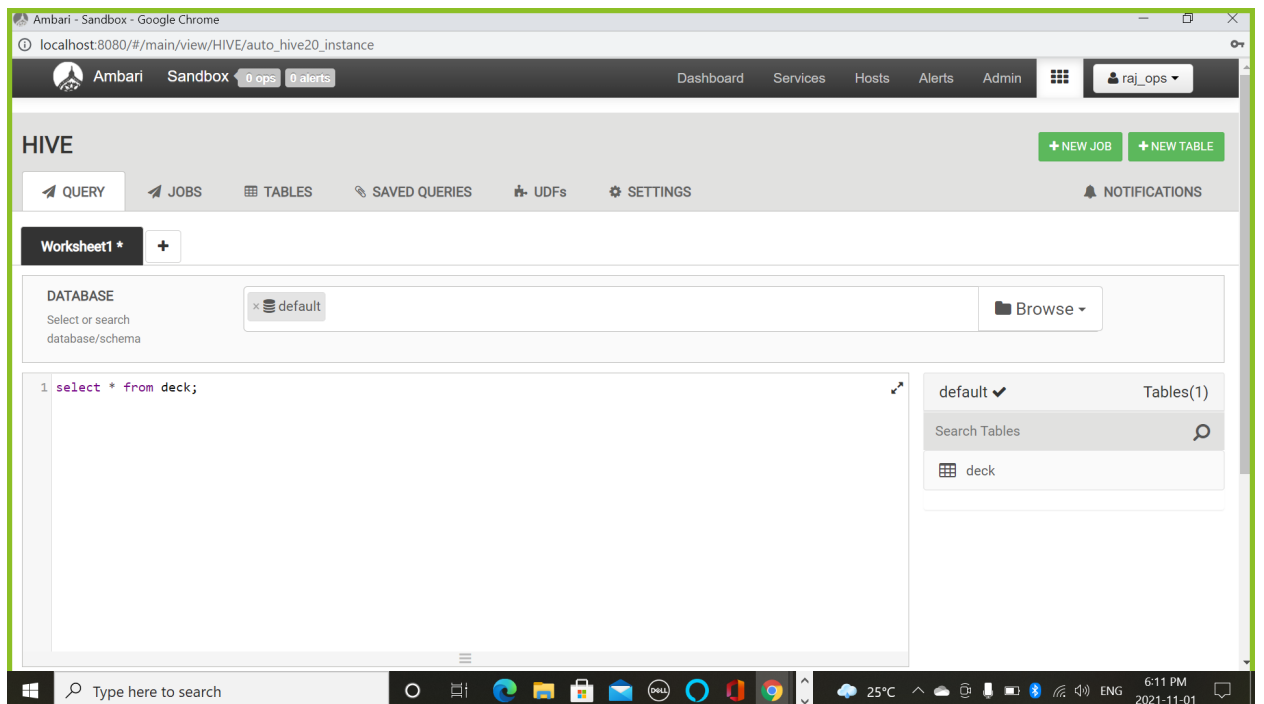
```
root@sandbox-hdp:~ - Shell In / x +
localhost:4200
Apps
Reading list

hive> load data local inpath '/hdfs://sandbox-hdp.hortonworks.com:8020/apps/hive/warehouse/deck/deckofcards.txt' into table deck
> select * from deck
> select * from deck;
FAILED: ParseException line 2:0 missing EOF at 'select' near 'deck'
hive> select * from deck;
OK
BLACK SPADE 2
BLACK SPADE 3
BLACK SPADE 4
BLACK SPADE 5
BLACK SPADE 6
BLACK SPADE 7
BLACK SPADE 8
BLACK SPADE 9
BLACK SPADE 10
BLACK SPADE J
BLACK SPADE Q
BLACK SPADE K
BLACK SPADE A
BLACK CLUB 2
BLACK CLUB 3
BLACK CLUB 4
BLACK CLUB 5
BLACK CLUB 6
BLACK CLUB 7
BLACK CLUB 8
BLACK CLUB 9
BLACK CLUB 10
BLACK CLUB J
BLACK CLUB Q
BLACK CLUB K
BLACK CLUB A
```

```
root@sandbox-hdp:~ - Shell In A x
localhost:4200
Apps | Reading list

BLACK CLUB K
BLACK CLUB A
RED DIAMOND 2
RED DIAMOND 3
RED DIAMOND 4
RED DIAMOND 5
RED DIAMOND 6
RED DIAMOND 7
RED DIAMOND 8
RED DIAMOND 9
RED DIAMOND 10
RED DIAMOND J
RED DIAMOND Q
RED DIAMOND K
RED DIAMOND A
RED HEART 2
RED HEART 3
RED HEART 4
RED HEART 5
RED HEART 6
RED HEART 7
RED HEART 8
RED HEART 9
RED HEART 10
RED HEART J
RED HEART Q
RED HEART K
RED HEART A
Time taken: 3.481 seconds, Fetched: 52 row(s)
hive>
```

Step-4: Using select command display the table in hive 2.0



Ambari - Sandbox - Google Chrome
localhost:8080/#/main/view/HIVE/auto_hive20_instance

Filter columns ✕

deck.color	deck.suit	deck.pip
BLACK	SPADE	2
BLACK	SPADE	3
BLACK	SPADE	4
BLACK	SPADE	5
BLACK	SPADE	6
BLACK	SPADE	7
BLACK	SPADE	8
BLACK	SPADE	9
BLACK	SPADE	10
BLACK	SPADE	J
BLACK	SPADE	Q
BLACK	SPADE	K
BLACK	SPADE	A
BLACK	CLUB	2
BLACK	CLUB	3

Type here to search

26°C 3:31 PM 2021-11-01

Ambari - Sandbox - Google Chrome
localhost:8080/#/main/view/HIVE/auto_hive20_instance

BLACK	SPADE	K
BLACK	SPADE	A
BLACK	CLUB	2
BLACK	CLUB	3
BLACK	CLUB	4
BLACK	CLUB	5
BLACK	CLUB	6
BLACK	CLUB	7
BLACK	CLUB	8
BLACK	CLUB	9
BLACK	CLUB	10
BLACK	CLUB	J
BLACK	CLUB	Q
BLACK	CLUB	K
BLACK	CLUB	A
RED	DIAMOND	2
RED	DIAMOND	3

Type here to search

26°C 3:31 PM 2021-11-01

2.) Create a managed table and load the data from HDFS

Step-1: Create managed table deck_of_cards in Hive having column names as color,suit and pip

```
root@sandbox-hdp:~ - Shell In A x +
localhost:4200
Apps
Reading list

sandbox-hdp login: root
root@sandbox-hdp.hortonworks.com's password:
Last login: Mon Nov  1 10:44:10 2021 from 172.18.0.2
[root@sandbox-hdp ~]# hive
log4j:WARN No such property [maxFileSize] in org.apache.log4j.DailyRollingFileAppender

Logging initialized using configuration in file:/etc/hive/2.6.5.0-292/0/hive-log4j.pro
perties
hive> create table deck_of_cards(color string,suit string,pip string)
> row format delimited fields terminated by '|'
> stored as textfile;
OK
Time taken: 4.148 seconds
```

Step-2: Describe table to know location of it

```
root@sandbox-hdp:~ - Shell In A x +
localhost:4200
Apps
Reading list

hive> describe formatted deck_of_cards;
OK
# col_name          data_type          comment
color               string
suit                string
pip                 string

# Detailed Table Information
Database:            default
Owner:               root
CreateTime:          Mon Nov 01 10:51:02 UTC 2021
LastAccessTime:      UNKNOWN
Protect Mode:        None
Retention:            0
Location:             hdfs://sandbox-hdp.hortonworks.com:8020/apps/hive/warehouse/de
ck_of_cards
Table Type:          MANAGED_TABLE
Table Parameters:
  COLUMN_STATS_ACCURATE {\"BASIC_STATS\": \"true\"}
  numFiles                0
  numRows                 0
  rawDataSize             0
  totalSize               0
  transient_lastDdlTime  1635763862

# Storage Information
SerDe Library:        org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat:          org.apache.hadoop.mapred.TextInputFormat
OutputFormat:          org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
```

Step-3: Load cards data into table and display it using select command

```
root@sandbox-hdp:~ - Shell In A x +
localhost:4200
Apps
Reading list

hive> load data inpath '/hdfs://sandbox-hdp.hortonworks.com:8020/apps/hive/warehouse/d
eck_of_cards/deckofcards.txt' into table deck_of_cards
> select * from deck_of_cards;
FAILED: ParseException line 2:0 missing EOF at 'select' near 'deck_of_cards'
hive> select * from deck_of_cards;
OK
BLACK SPADE 2
BLACK SPADE 3
BLACK SPADE 4
BLACK SPADE 5
BLACK SPADE 6
BLACK SPADE 7
BLACK SPADE 8
BLACK SPADE 9
BLACK SPADE 10
BLACK SPADE J
BLACK SPADE Q
BLACK SPADE K
BLACK SPADE A
BLACK CLUB 2
BLACK CLUB 3
BLACK CLUB 4
BLACK CLUB 5
BLACK CLUB 6
BLACK CLUB 7
BLACK CLUB 8
BLACK CLUB 9
BLACK CLUB 10
BLACK CLUB J
BLACK CLUB Q
```

```
root@sandbox-hdp:~ - Shell In A x +
localhost:4200
Apps
Reading list

BLACK CLUB K
BLACK CLUB A
RED DIAMOND 2
RED DIAMOND 3
RED DIAMOND 4
RED DIAMOND 5
RED DIAMOND 6
RED DIAMOND 7
RED DIAMOND 8
RED DIAMOND 9
RED DIAMOND 10
RED DIAMOND J
RED DIAMOND Q
RED DIAMOND K
RED DIAMOND A
RED HEART 2
RED HEART 3
RED HEART 4
RED HEART 5
RED HEART 6
RED HEART 7
RED HEART 8
RED HEART 9
RED HEART 10
RED HEART J
RED HEART Q
RED HEART K
RED HEART A
Time taken: 1.455 seconds, Fetched: 52 row(s)
hive>
```

Step-4: Using select command display contents of table in hive 2.0

Ambari - Sandbox - Google Chrome

localhost:8080/#/main/view/HIVE/auto_hive20_instance

Select or search database/schema

1 select * from deck_of_cards;

default Tables(1)

Search Tables

deck_of_cards

Execute Save As Insert UDF Visual Explain

RESULTS LOG VISUAL EXPLAIN TEZ UI

Filter columns

deck_of_cards.color	deck_of_cards.suit	deck_of_cards.pip
BLACK	SPADE	2
BLACK	SPADE	3

Ambari - Sandbox - Google Chrome

localhost:8080/#/main/view/HIVE/auto_hive20_instance

Execute Save As Insert UDF Visual Explain

RESULTS LOG VISUAL EXPLAIN TEZ UI

Filter columns

deck_of_cards.color	deck_of_cards.suit	deck_of_cards.pip
BLACK	SPADE	2
BLACK	SPADE	3
BLACK	SPADE	4
BLACK	SPADE	5
BLACK	SPADE	6
BLACK	SPADE	7
BLACK	SPADE	8
BLACK	SPADE	9
BLACK	SPADE	10
BLACK	SPADE	J
BLACK	SPADE	Q
BLACK	SPADE	K

Ambari - Sandbox - Google Chrome

localhost:8080/#/main/view/HIVE/auto_hive20_instance

BLACK	SPADE	8
BLACK	SPADE	9
BLACK	SPADE	10
BLACK	SPADE	J
BLACK	SPADE	Q
BLACK	SPADE	K
BLACK	SPADE	A
BLACK	CLUB	2
BLACK	CLUB	3
BLACK	CLUB	4
BLACK	CLUB	5
BLACK	CLUB	6
BLACK	CLUB	7
BLACK	CLUB	8
BLACK	CLUB	9
BLACK	CLUB	10
BLACK	CLUB	J
BLACK	CLUB	Q

Type here to search

25°C 4:58 PM 2021-11-01

Step-5: Checking managed tables in sandbox

Ambari - Sandbox - Google Chrome

localhost:8080/#/main/view/FILES/auto_files_instance

Ambari Sandbox 0 ops 0 alerts

Dashboard Services Hosts Alerts Admin

raj_ops

/ > apps > hive > warehouse

Total: 3 files or folders

+ Select All New Folder Upload

Search in current directory...

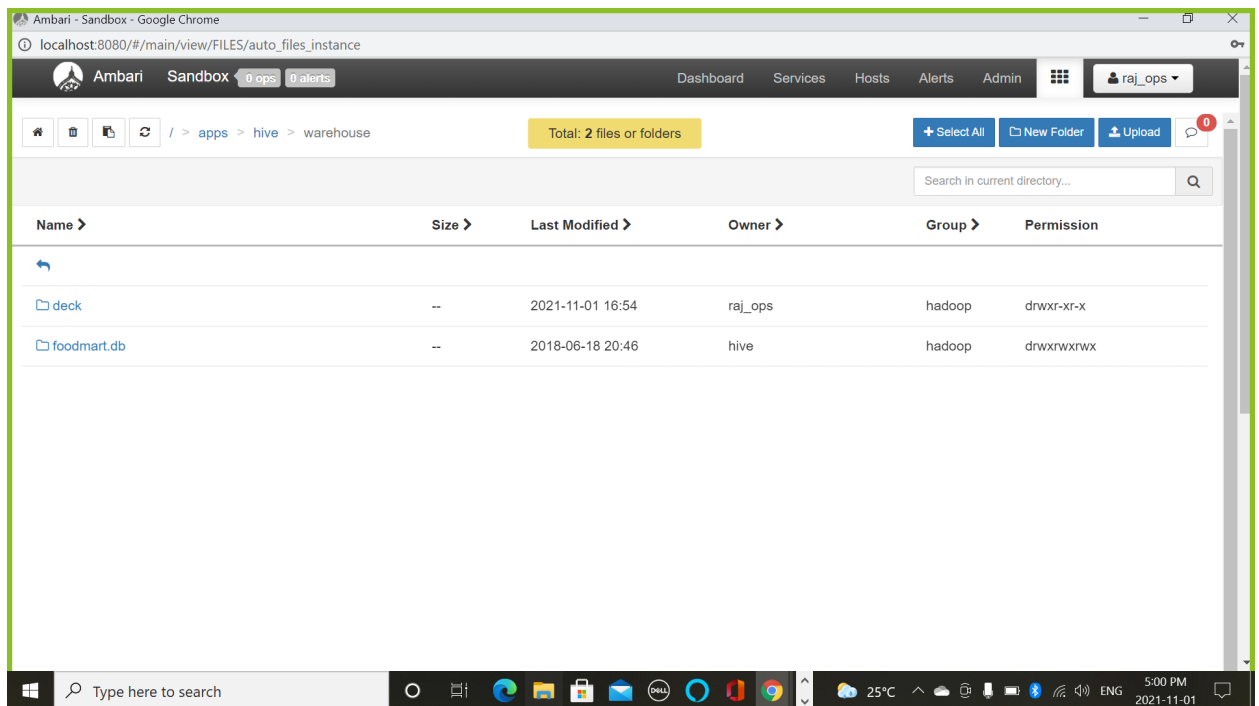
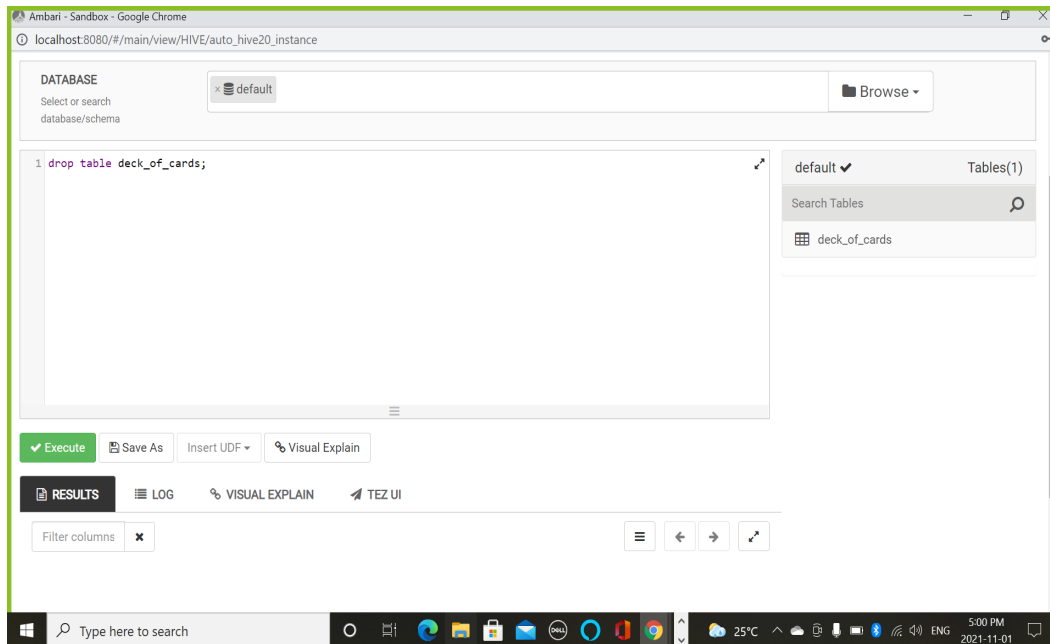
Name >	Size >	Last Modified >	Owner >	Group >	Permission
↶					
deck	--	2021-11-01 16:54	raj_ops	hadoop	drwxr-xr-x
deck_of_cards	--	2021-11-01 16:21	root	hadoop	drwxrwxrwx
foodmart.db	--	2018-06-18 20:46	hive	hadoop	drwxrwxrwx

Type here to search

25°C 4:59 PM 2021-11-01

3.) Drop a managed table and check the result in HDFS

Step-1: Drop managed table in hive 2.0



Step-2: Checking whether table exist or not

```
hive> select * from deck_of_cards;  
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'deck_of_cards'  
hive>
```



4.) Create an external table and load the data from LFS

Step-1: Create an external table deck_external and describe table to get location of it

```
hive> create external table deck_external(color string,suit string,pip string)
> row format delimited fields terminated by '|'
> stored as textfile;
OK
Time taken: 3.252 seconds
hive> describe formatted deck_external
> ;
OK
# col_name          data_type          comment
#-----
color               string
suit                string
pip                 string

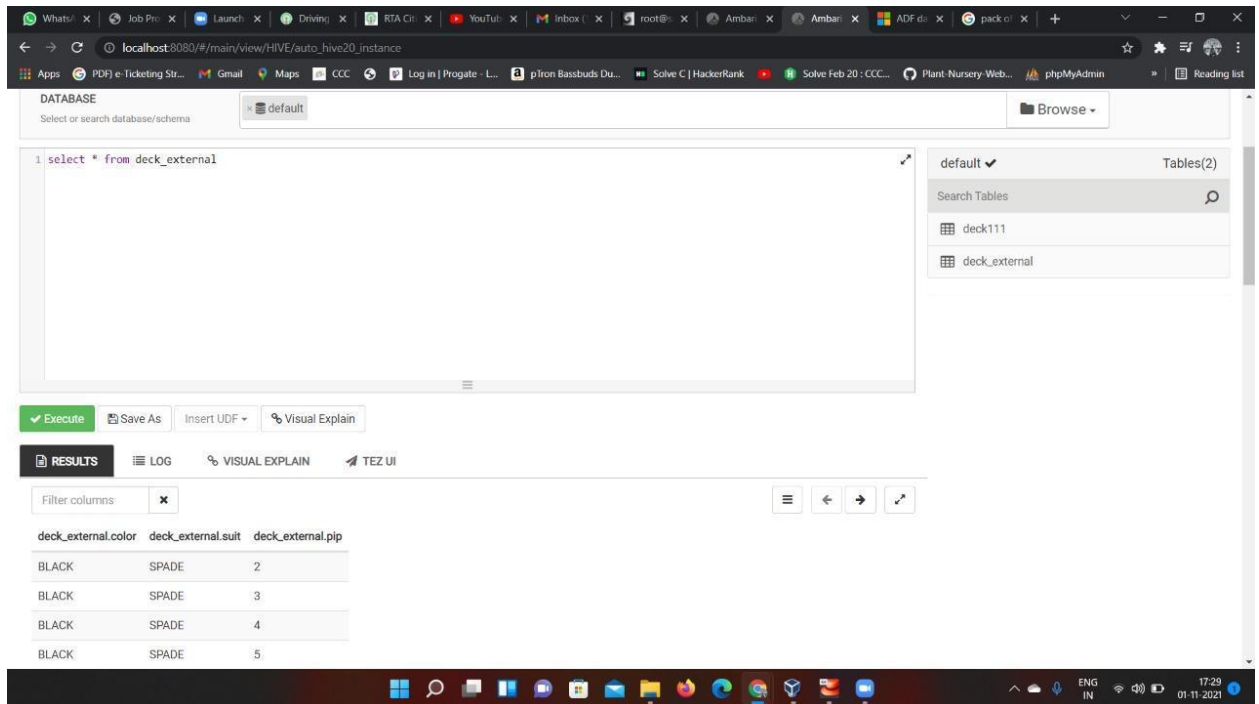
# Detailed Table Information
Database:            default
Owner:               root
CreateTime:          Sun Oct 17 19:25:21 UTC 2021
LastAccessTime:      UNKNOWN
Protect Mode:        None
Retention:           0
Location:            hdfs://sandbox-hdp.hortonworks.com:8020/apps/hive/warehouse/deck_external
Table Type:          EXTERNAL_TABLE
Table Parameters:
  COLUMN_STATS_ACCURATE  {\\"BASIC_STATS\\":\\"true\\"}
  EXTERNAL                TRUE
  numFiles                0
  numRows                 0
  rawDataSize             0
  totalSize               0
  transient_lastDdlTime  1634498721

# Storage Information
SerDe Library:       org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat:         org.apache.hadoop.mapred.TextInputFormat
OutputFormat:        org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
Compressed:          No
Num Buckets:         -1
Bucket Columns:      []
Sort Columns:        []
```

Step-2: Load cards data into table and display contents of it using select command

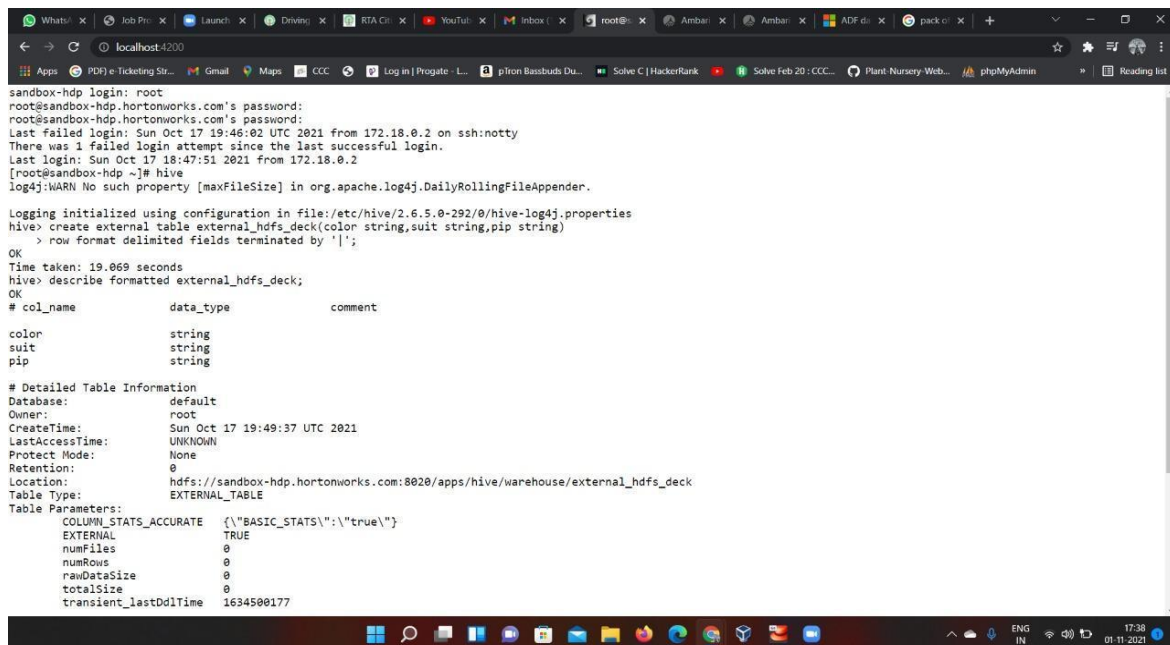
```
hive> load data local inpath '/hdfs://sandbox-hdp.hortonworks.com:8020/apps/hive/warehouse/deck_external/deckofcards.txt' into table deck_external
Time taken: 1.280 seconds, Fetched: 35 row(s)
hive> select * from deck_external;
FAILED: ParseException line 2:0 missing EOF at 'select' near 'deck_external'
hive> select * from deck_external;
OK
BLACK SPADE 2
BLACK SPADE 3
BLACK SPADE 4
BLACK SPADE 5
BLACK SPADE 6
BLACK SPADE 7
BLACK SPADE 8
BLACK SPADE 9
BLACK SPADE 10
BLACK SPADE J
BLACK SPADE Q
BLACK SPADE K
BLACK SPADE A
BLACK CLUB 2
BLACK CLUB 3
BLACK CLUB 4
BLACK CLUB 5
BLACK CLUB 6
BLACK CLUB 7
BLACK CLUB 8
BLACK CLUB 9
BLACK CLUB 10
BLACK CLUB J
BLACK CLUB Q
BLACK CLUB K
BLACK CLUB A
RED DIAMOND 2
```

Step-3: Display table in hive 2.0 using select command

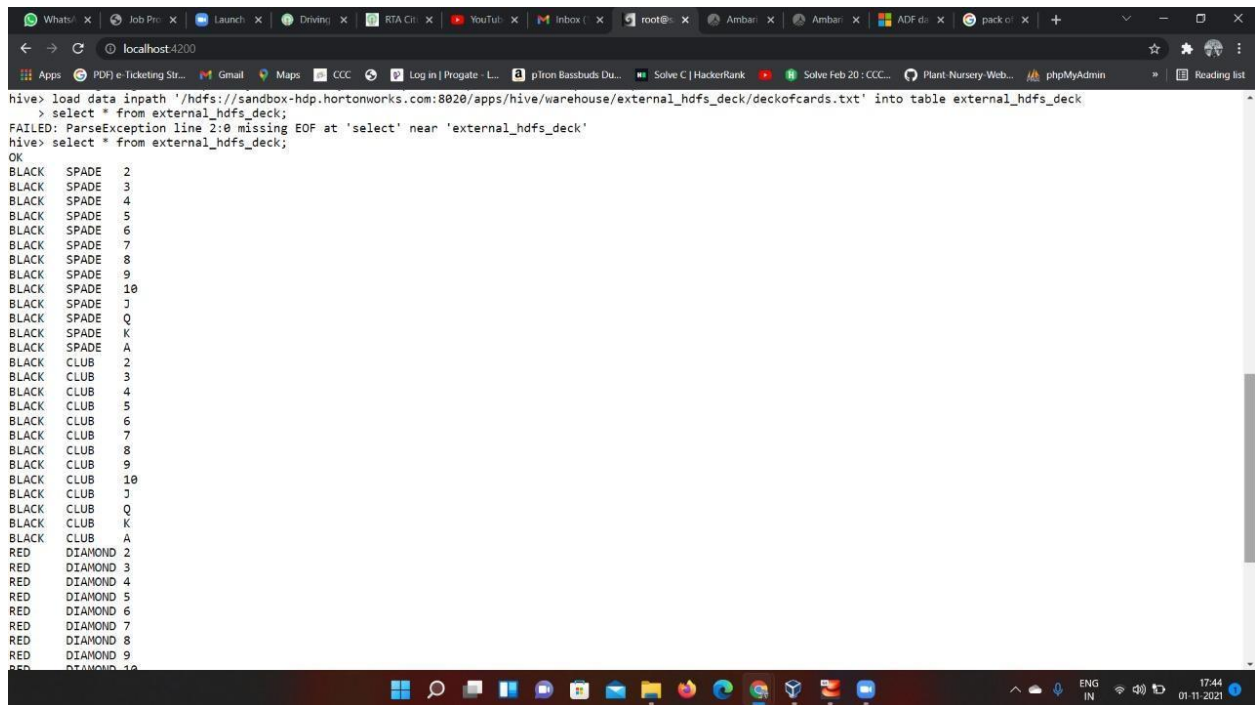


5.) Create an external table and load the data from HDFS

Step-1: Create an external table external_hdfs_deck and describe it to get location of it

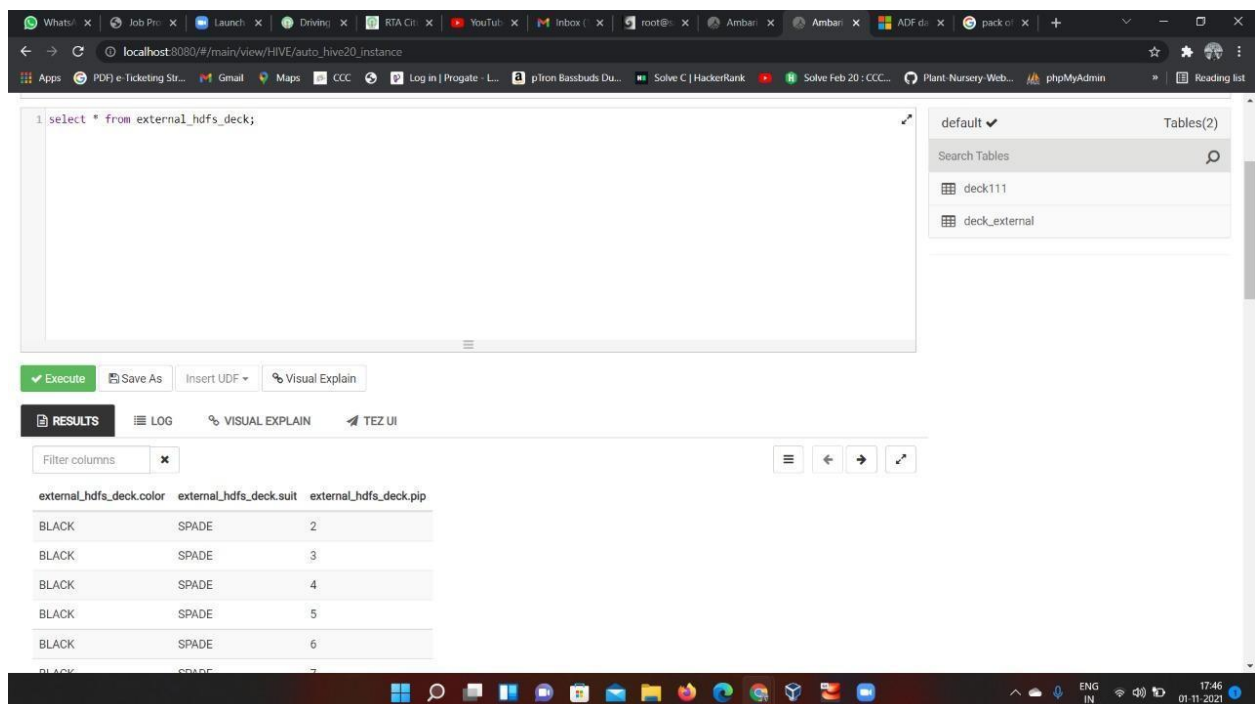


Step-2: Load cards data into table and display contents of it using select command



```
hive> load data inpath '/hdfs://sandbox-hdp.hortonworks.com:8020/apps/hive/warehouse/external_hdfs_deck/deckofcards.txt' into table external_hdfs_deck;
> select * from external_hdfs_deck;
FAILED: ParseException line 2:0 missing EOF at 'select' near 'external_hdfs_deck'
hive> select * from external_hdfs_deck;
OK
BLACK SPADE 2
BLACK SPADE 3
BLACK SPADE 4
BLACK SPADE 5
BLACK SPADE 6
BLACK SPADE 7
BLACK SPADE 8
BLACK SPADE 9
BLACK SPADE 10
BLACK SPADE J
BLACK SPADE Q
BLACK SPADE K
BLACK SPADE A
BLACK CLUB 2
BLACK CLUB 3
BLACK CLUB 4
BLACK CLUB 5
BLACK CLUB 6
BLACK CLUB 7
BLACK CLUB 8
BLACK CLUB 9
BLACK CLUB 10
BLACK CLUB J
BLACK CLUB Q
BLACK CLUB K
BLACK CLUB A
RED DIAMOND 2
RED DIAMOND 3
RED DIAMOND 4
RED DIAMOND 5
RED DIAMOND 6
RED DIAMOND 7
RED DIAMOND 8
RED DIAMOND 9
RED DIAMOND 10
```

Step-3: Using select command display table in hive 2.0



```
1 select * from external_hdfs_deck;
```

default Tables(2)

Search Tables

- deck111
- deck_external

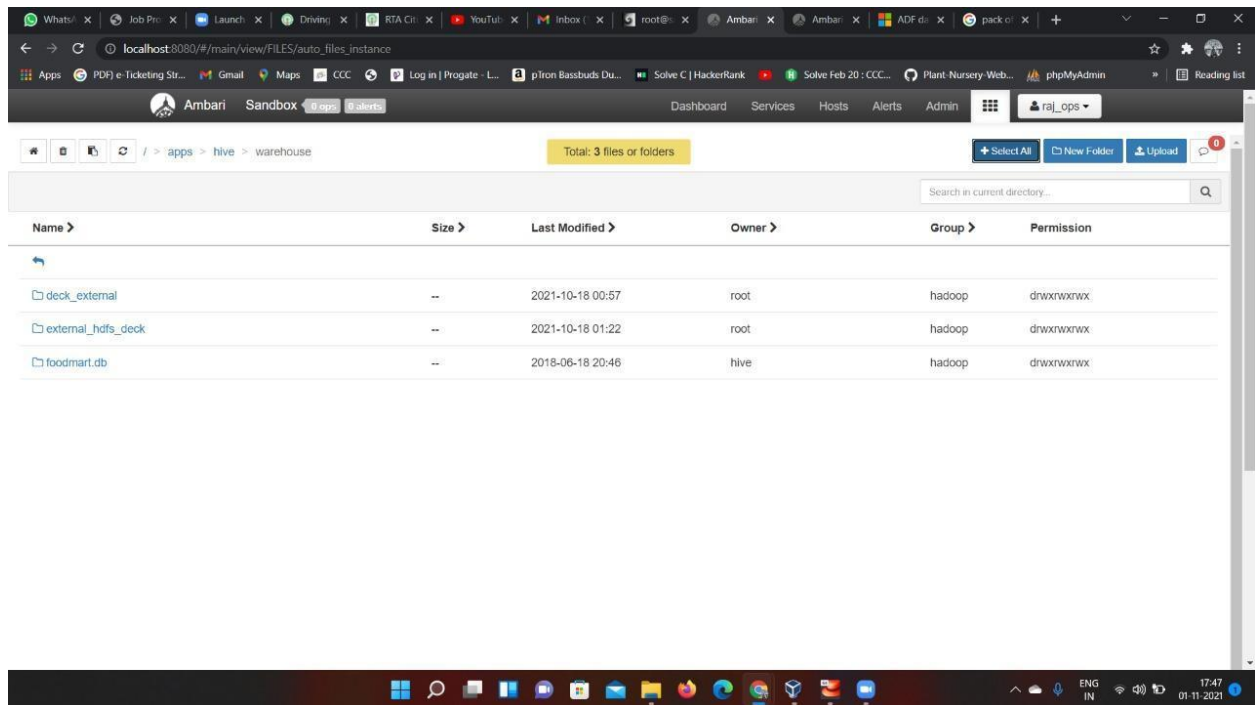
Execute Save As Insert UDF Visual Explain

RESULTS LOG VISUAL EXPLAIN TEZ UI

Filter columns

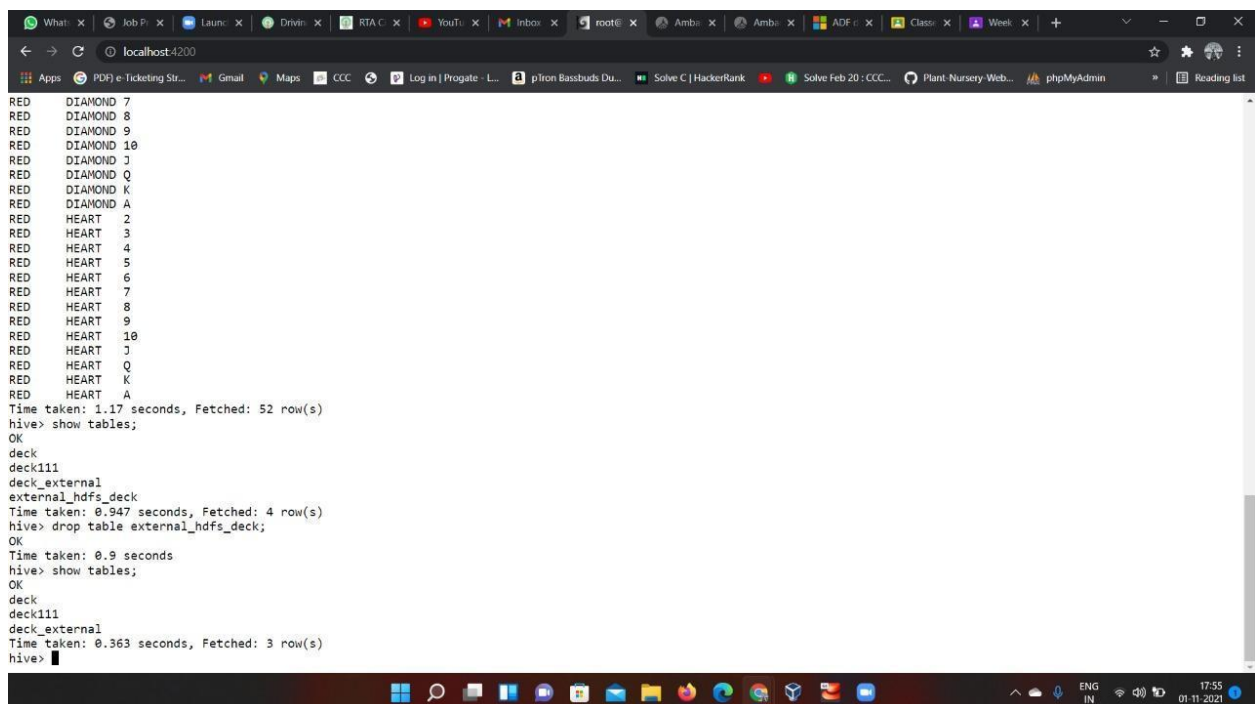
external_hdfs_deck.color	external_hdfs_deck.suit	external_hdfs_deck.pip
BLACK	SPADE	2
BLACK	SPADE	3
BLACK	SPADE	4
BLACK	SPADE	5
BLACK	SPADE	6
BLACK	SPADE	7

Step-4: Checking external tables in sandbox

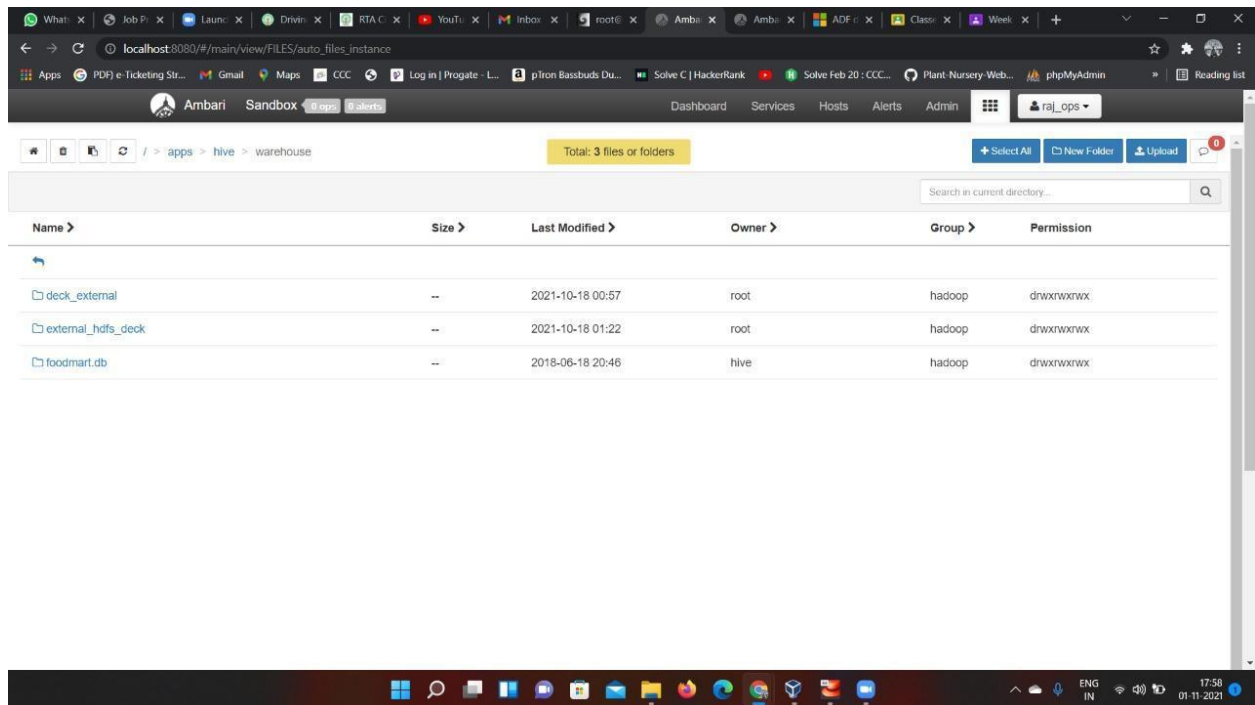


6.) Drop an external table and check the data from HDFS




Step-1: Drop external table using drop command and using show command checking table



Step-2: Checking status of table in sandbox



The screenshot shows the Ambari Sandbox interface. The breadcrumb navigation indicates the path: / > apps > hive > warehouse. A yellow box above the table states "Total: 3 files or folders". The table below lists the files in the directory:

Name >	Size >	Last Modified >	Owner >	Group >	Permission
 deck_external	--	2021-10-18 00:57	root	hadoop	drwxrwxrwx
 external_hdfs_deck	--	2021-10-18 01:22	root	hadoop	drwxrwxrwx
 foodmart.db	--	2018-06-18 20:46	hive	hadoop	drwxrwxrwx