

Hive Data Definations:

Hive data definations basically refers to a set of commands which are used to define a data set i.e. defining the type of a data base or a table, the structure of its contents, etc. Eventually it entails the commands which are used to create or alter a database or a table.

DDL Commands

create database

drop database

create table

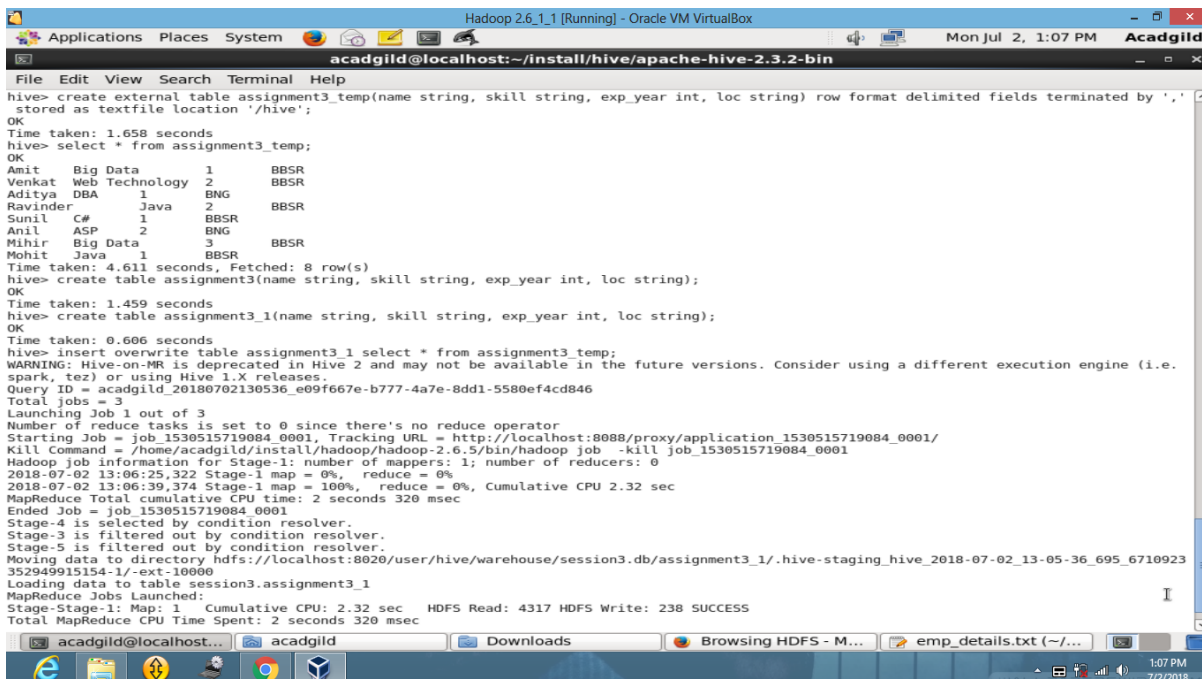
drop table

alter table

create index

create views

DDL Example:



```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
acadgild@localhost:~/install/hive/apache-hive-2.3.2-bin
File Edit View Search Terminal Help
hive> create external table assignment3_temp(name string, skill string, exp_year int, loc string) row format delimited fields terminated by ','
stored as textfile location '/hive';
OK
Time taken: 1.658 seconds
hive> select * from assignment3_temp;
OK
Amit      Big Data      1      BBSR
Venkat    Web Technology 2      BBSR
Aditya    DBA            1      BNG
Ravinder  Java           2      BBSR
Sunil     C#             1      BBSR
Anil      ASP            2      BNG
Mihir     Big Data      3      BBSR
Mohit     Java           1      BBSR
Time taken: 4.611 seconds, Fetched: 8 row(s)
hive> create table assignment3(name string, skill string, exp_year int, loc string);
OK
Time taken: 1.459 seconds
hive> create table assignment3_1(name string, skill string, exp_year int, loc string);
OK
Time taken: 0.606 seconds
hive> insert overwrite table assignment3_1 select * from assignment3_temp;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e.
spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180702130536_e09f667e-b777-4a7e-8dd1-5580ef4cd846
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1530515719084_0001, Tracking URL = http://localhost:8088/proxy/application_1530515719084_0001/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1530515719084_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2018-07-02 13:06:25.322 Stage-1 map = 0%, reduce = 0%
2018-07-02 13:06:39.374 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.32 sec
MapReduce Total cumulative CPU time: 2 seconds 320 msec
Ended Job = job_1530515719084_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://localhost:8020/user/hive/warehouse/session3.db/assignment3_1/.hive-staging_hive_2018-07-02_13-05-36_695_6710923
352949915154-1/-ext-10000
Loading data to table session3.assignment3_1
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 2.32 sec HDFS Read: 4317 HDFS Write: 238 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 320 msec
```

Hive Data Manipulations:

Hive data manipulation basically refers to a set of commands which are used to manipulate a data set i.e. manipulating the data in a data base or a table as required according to the task. Eventually it entails the commands which are used to fetch and alter data from a table or a set of tables(in case of joins).

Hive DML Commands

Select

Where

Group By

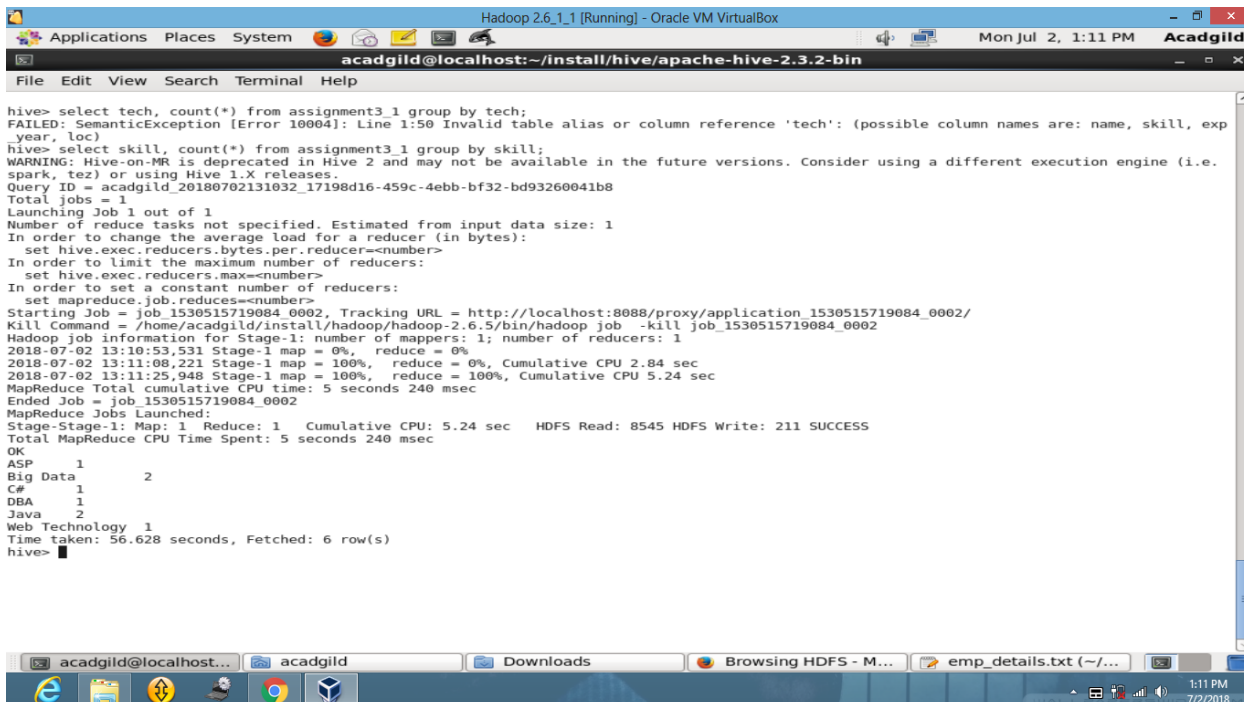
Order By

Load Data

Join:

- Inner Join
- Left Outer Join
- Right Outer Join
- Full Outer Join

DML example:

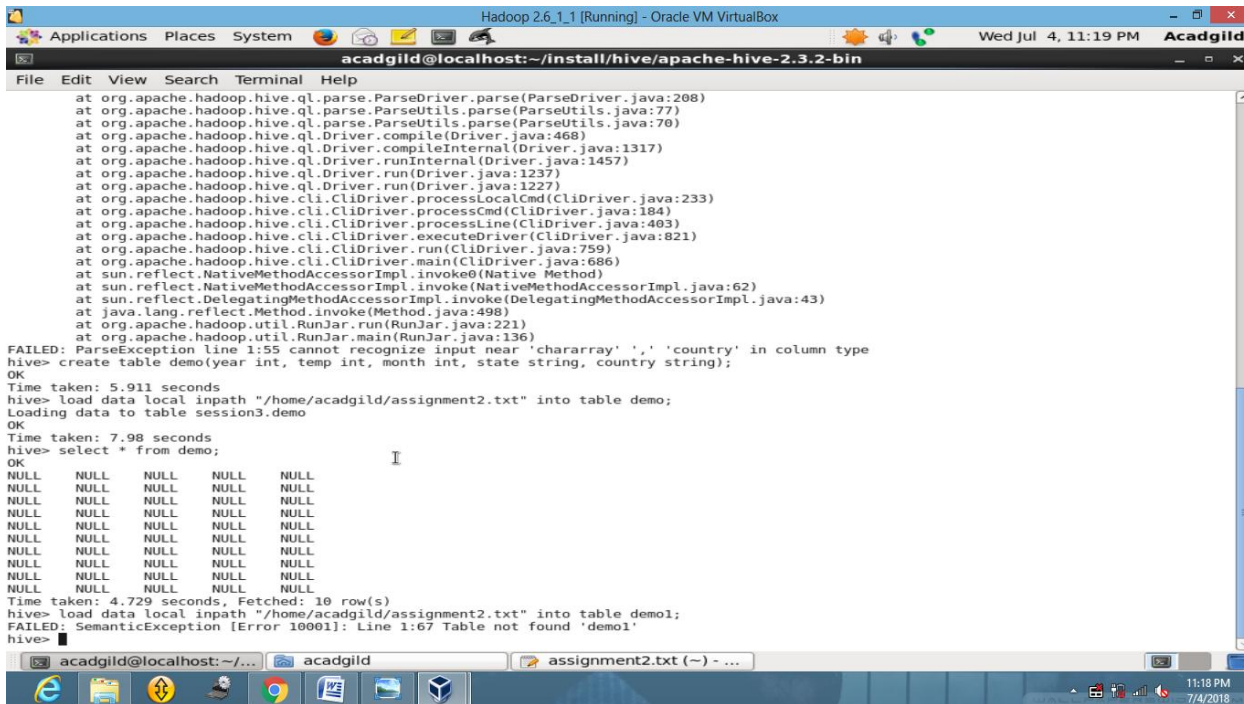


```
acadgild@localhost:~/install/hive/apache-hive-2.3.2-bin
hive> select tech, count(*) from assignment3_1 group by tech;
FAILED: SemanticException [Error 10004]: Line 1:50 Invalid table alias or column reference 'tech': (possible column names are: name, skill, exp
year, loc)
hive> select skill, count(*) from assignment3_1 group by skill;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e.
spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180702131032_17198d16-459c-4ebb-bf32-bd93260041b8
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reducers=<number>
Starting Job = job_1530515719084_0002, Tracking URL = http://localhost:8080/proxy/application_1530515719084_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1530515719084_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-07-02 13:10:53,531 Stage-1 map = 0%, reduce = 0%
2018-07-02 13:11:08,221 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.84 sec
2018-07-02 13:11:25,948 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.24 sec
MapReduce Total cumulative CPU time: 5 seconds 240 msec
Ended Job = job_1530515719084_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.24 sec HDFS Read: 8545 HDFS Write: 211 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 240 msec
OK
ASP          1
Big Data     2
C#           1
DBA          1
Java         2
Web Technology 1
Time taken: 56.628 seconds, Fetched: 6 row(s)
hive>
```

Hive QL Manipulations:

Hive QL or Hive query language data manipulation basically refers to a code which is used to manipulate a data set i.e. manipulating the data in a data base or a table as required according to the task. Eventually it entails the code which is used to load the data in a hive table straight from an externally stored file which may either be stored in hdfs or the local file system.

Hive QL Example:



```
acadmild@localhost:~/install/hive/apache-hive-2.3.2-bin
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:208)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:77)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:70)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:468)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1317)
at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1457)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1237)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1227)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:233)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:184)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:403)
at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:821)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:759)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:686)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:55 cannot recognize input near 'chararray' ',', 'country' in column type
hive> create table demo(year int, temp int, month int, state string, country string);
OK
Time taken: 5.911 seconds
hive> load data local inpath "/home/acadgild/assignment2.txt" into table demo;
Loading data to table session3.demo
OK
Time taken: 7.98 seconds
hive> select * from demo;
OK
NULL      NULL      NULL      NULL      NULL
NULL      NULL      NULL      NULL      NULL
NULL      NULL      NULL      NULL      NULL
NULL      NULL      NULL      NULL      NULL
NULL      NULL      NULL      NULL      NULL
NULL      NULL      NULL      NULL      NULL
NULL      NULL      NULL      NULL      NULL
NULL      NULL      NULL      NULL      NULL
NULL      NULL      NULL      NULL      NULL
NULL      NULL      NULL      NULL      NULL
Time taken: 4.729 seconds, Fetched: 10 row(s)
hive> load data local inpath "/home/acadgild/assignment2.txt" into table demo1;
FAILED: SemanticException [Error 10001]: Line 1:67 Table not found 'demo1'
hive>
```