

Assignment 15.1

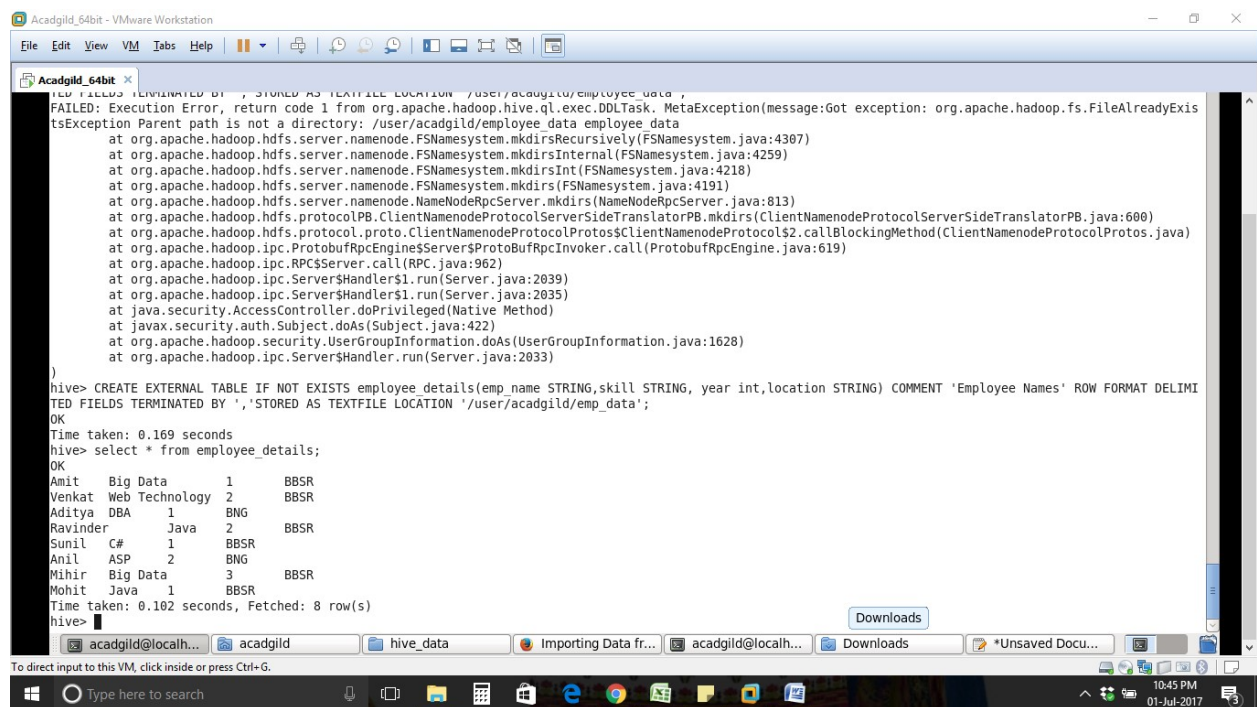
Problem Statement

Calculate the number of employees corresponding to each skill from the table 'employee' which is loaded in the Demo.

Table created

Create external table if not exists employee_details(emp_name String, skill string,year int, location string) comment 'Employee Names ' row format delimited fields terminated by ',' stored as textfile location '/user/acadgild/emp_data';

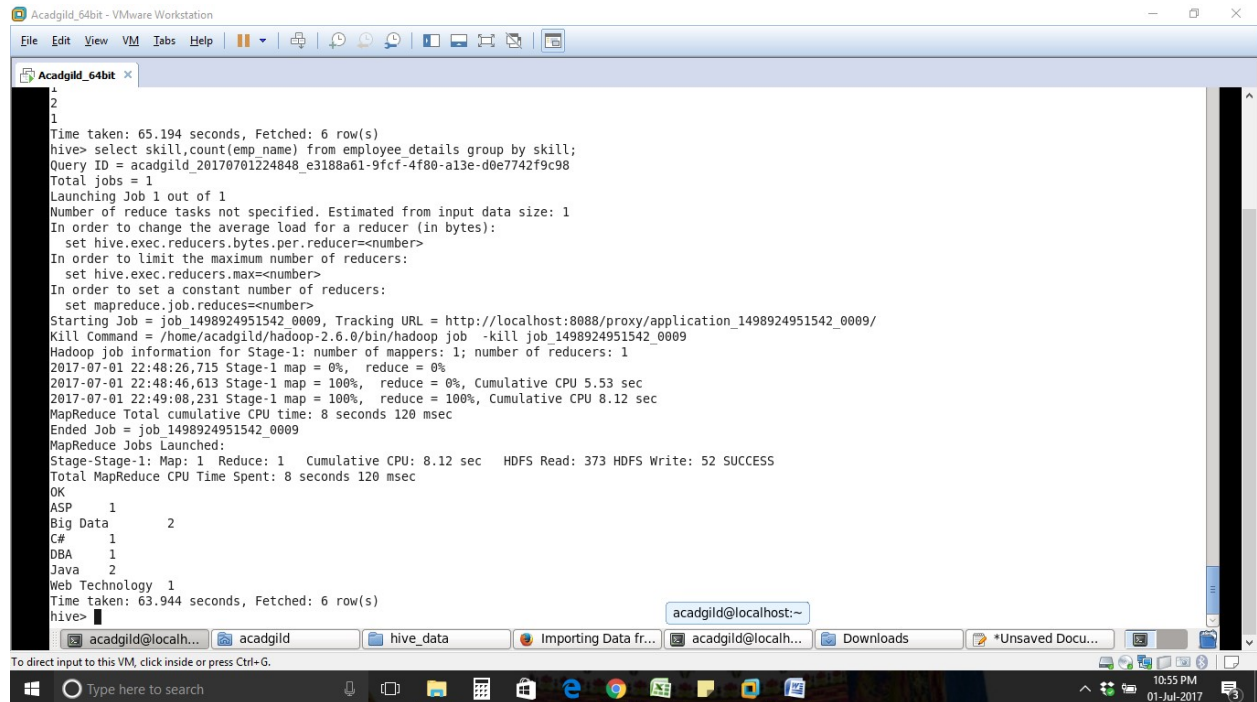
Select * from employee_details;



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
TERMINATED BY ', STORED AS TEXTFILE LOCATION '/user/acadgild/emp_data';
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. MetaException(message:Got exception: org.apache.hadoop.fs.FileAlreadyExistsException Parent path is not a directory: /user/acadgild/employee_data employee_data
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirsRecursively(FSNamesystem.java:4307)
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirsInternal(FSNamesystem.java:4259)
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirsInt(FSNamesystem.java:4218)
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirs(FSNamesystem.java:4191)
    at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.mkdirs(NameNodeRpcServer.java:813)
    at org.apache.hadoop.hdfs.protocolPB.ClientNameNodeProtocolServerSideTranslatorPB.mkdirs(ClientNameNodeProtocolServerSideTranslatorPB.java:600)
    at org.apache.hadoop.hdfs.protocol.proto.ClientNameNodeProtocolProtos$ClientNameNodeProtocols$2.callBlockingMethod(ClientNameNodeProtocolProtos.java)
    at org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine.java:619)
    at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:962)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2039)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2035)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1628)
    at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2033)
)
hive> CREATE EXTERNAL TABLE IF NOT EXISTS employee_details(emp_name STRING,skill STRING, year int,location STRING) COMMENT 'Employee Names ' ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/acadgild/emp_data';
OK
Time taken: 0.169 seconds
hive> select * from employee_details;
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: 0.102 seconds, Fetched: 8 row(s)
hive>
```

Calculate the no of employees in each skill

Select skill, count(emp_name) from employee_details group by skill;



```
Acadgild_64bit x
1
2
1
Time taken: 65.194 seconds, Fetched: 6 row(s)
hive> select skill,count(emp_name) from employee_details group by skill;
Query ID = acadgild_20170701224848_e3188a61-9fcf-4f80-a13e-d0e7742f9c98
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.reduces=<number>
Starting Job = job_1498924951542_0009, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0009/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-07-01 22:48:26,715 Stage-1 map = 0%, reduce = 0%
2017-07-01 22:48:46,613 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.53 sec
2017-07-01 22:49:08,231 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.12 sec
MapReduce Total cumulative CPU time: 8 seconds 120 msec
Ended Job = job_1498924951542_0009
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.12 sec HDFS Read: 373 HDFS Write: 52 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 120 msec
OK
ASP      1
Big Data    2
C#         1
DBA        1
Java       2
Web Technology  1
Time taken: 63.944 seconds, Fetched: 6 row(s)
hive>
```