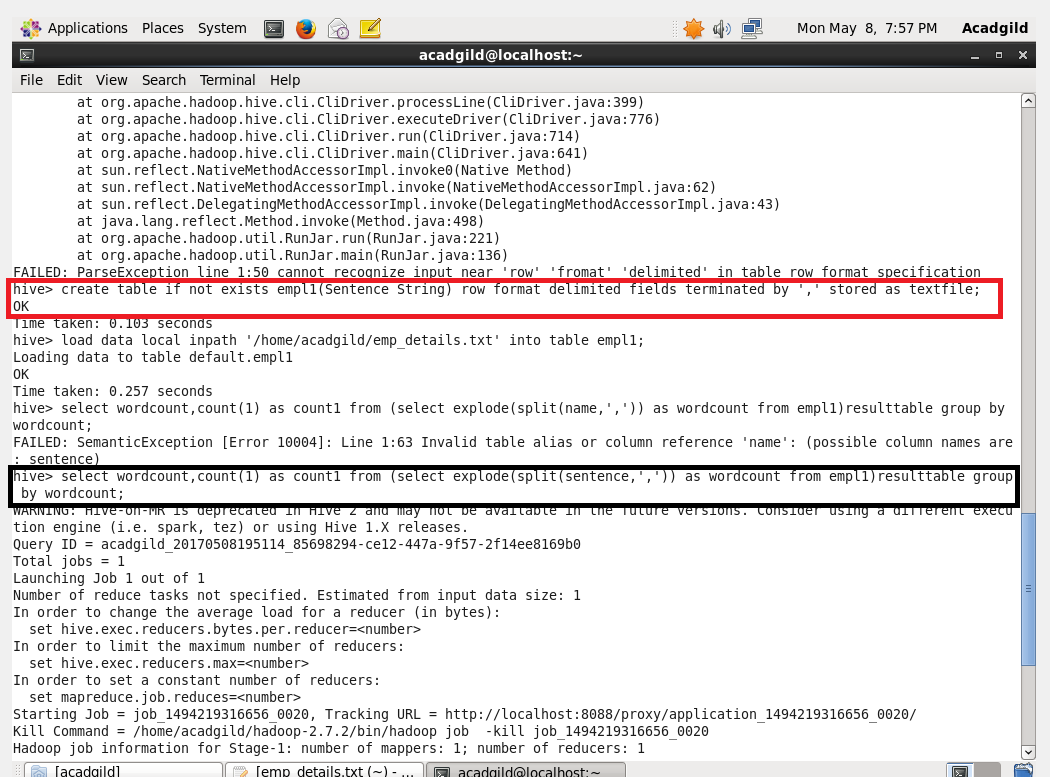
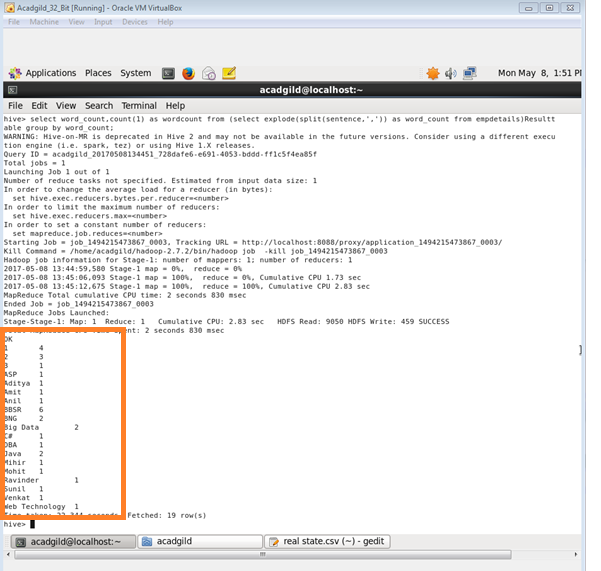
**Perform word count in Hive for above given dataset.**

****

****

**2. Explain the working of Partitioning in brief.**

The term ‘Table partitioning’ - dividing the data inside the table into number of parts which is based on column values.

(eg date or country, segregating the input records which is into the different types of files/directories which was based on date or country).

* Partitioning will be done on the basis of column where multi\_dimensional structure will be imposed on the directory for storage.
* For example- additional to partitioning log the records of the date column we can also sub divide to a single day record which will be country wise separate files
* And it will include country column into partitioning.

There are two types,

1)Static partitioning

2)Dynamic partitioning

**1.)Static partitioning*:***

* In static partitioning the input data will contain all the columns listed detail which will be done only in table definition
* but columns will not be defined in partitioned by clause.
* If input column layout = expected layout
* Then the input files will be separated for each partitioned key value pairs
* For example- there will be one separate file allotted for each combination of country and state values
* These files can be easily loaded into partitioned table

**2. Dynamic partitioning:**

* Partition will be loaded for for each partition.
* Which will be will be with the help of SQL statements.
* And it will result in a lot of SQL statements
* As well there will be a huge no of partition.
* HIVE will support dynamic partition
* So we can add any number of partitions using a single SQL execution.
* Using Hive automatic splitting of data into separate partition files based on the values of partition keys present in the input files.

**3. Explain the difference between Static and Dynamic Partitioning in Hive with an example.**

S**tatic Partition in Hive:**

* Individually input data files were inserted into the partition table
* Static Partition is usual while loading files (big files) into Hive tables
* Mostly static partitions are preferred.
* Using Static Partition will save the time of data loading when compared to dynamic partition
* “Statically” one can add data to the a partition table and move the file into the table.
* One can even alter the partition in static.
* Using file name you can get value of the partition column.
* Eg-day of date etc..
* Without reading the whole big file.
* Hive should be proper if one needs to use Static partition.

set hive.mapred.mode = strict

* It will be set default in hive-site.xml
* Strict Mode will be followed in Static partition.
* You should use where clause to use limit in static partition
* One can perform Static partition on Hive Manage table or external table.

**Dynamic Partition in Hive**

* When there is a Single insert to partition table is known as dynamic partition
* dynamic partition will usually load data from non partitioned table.
* But more time is needed for data loading when compared to static partition.
* There will be large data stored in a table so Dynamic partition is suitable.
* When there is a need to partition number of column but number of column’s were not known to know the column number dynamic partition is suitable
* No requirement of where clause to use limit.
* Alteration can’t be performed on Dynamic partition.
* You can perform dynamic partition on hive external table
* If one need to use Dynamic partition in hive then mode should be in nonstrict mode
* In hive dynamic partition properties you should allow.