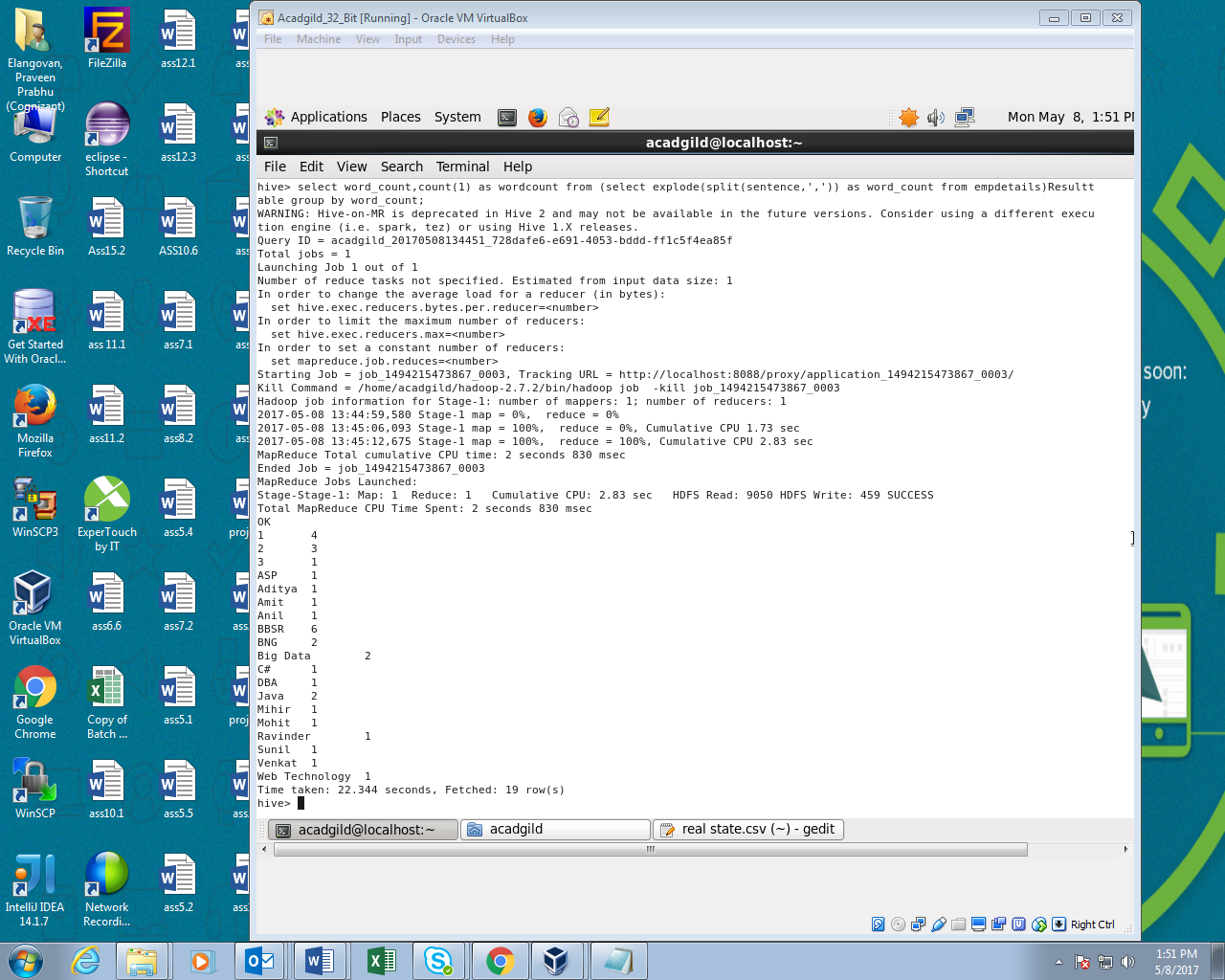
Perform word count in Hive for above given dataset.

First I created the table using the given data set. Then using split function I splitted the words using comma separator.explode function used to shift the words into subsequent rows. And finaly using count and group by function we can perform word count for unique words.



2. Explain the working of Partitioning in brief.

Table partitioning means dividing table data into some parts based on the values of particular columns like date or country,segregate the input records into different files/directories based on date or country.

Partitioning can be done based on more than column which will impose multi\_dimensional structure on directory storage.For example,in addition to partitioning log records by date column , we can also sup divide the single day records into country wise separate files b including country column into partitioning.

There are two types,

1)Static partitioning

2)Dynamic partitioning

Static partitioning:

In this mode input data should contain the columns listed only in table definition but not the columns defined in partitioned by clause.

If our input column layout is according to the expected layout and we already have separate input files for each partitioned key value pairs,like one separate file for each combination of country and state values and these files can be easily loaded into partitioned table

Dynamic partitioning:

Instead of loading each partition with single SQL statements ,which will result in writing lot of sql statements for huge no of partition,HIVE supports dynamic partitioning with which we can add any number of partitons with single SQL execution.

Hive will automatically splits our 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**

Insert input data files individually into a partition table is Static Partition Usually when loading files (big files) into Hive tables static partitions are preferred

Static Partition saves your time in loading data compared to dynamic partition You “statically” add a partition in table and move the file into the partition of the table.

We can alter the partition in static partition

You can get the partition column value form the filename, day of date etc without reading the whole big file. If you want to use Static partition in hive you should set property

set hive.mapred.mode = strict  
This property set by default in hive-site.xml Static partition is in Strict Mode You should use where clause to use limit in static partition You can perform Static partition on Hive Manage table or external table.

**Dynamic Partition in Hive**

single insert to partition table is known as dynamic partition

Usually dynamic partition load the data from non partitioned table

Dynamic Partition takes more time in loading data compared to static partition

When you have large data stored in a table then Dynamic partition is suitable.

If you want to partition number of column but you don’t know how many columns then also dynamic partition is suitable

Dynamic partition there is no required where clause to use limit. we can’t perform alter on Dynamic partition

You can perform dynamic partition on hive external table and managed table If you want to use Dynamic partition in hive then mode is in nonstrict mode Here is hive dynamic partition properties you should allow