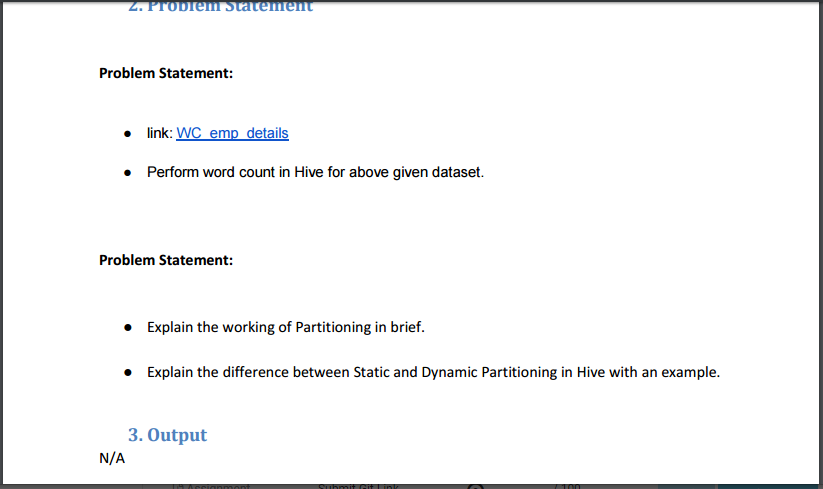
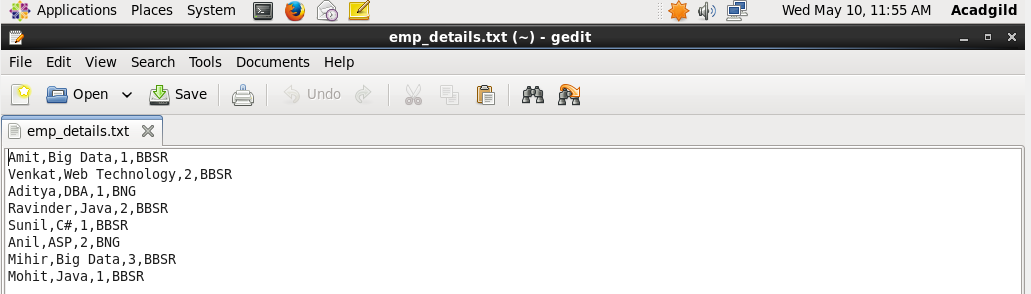
Asignment-26.1:

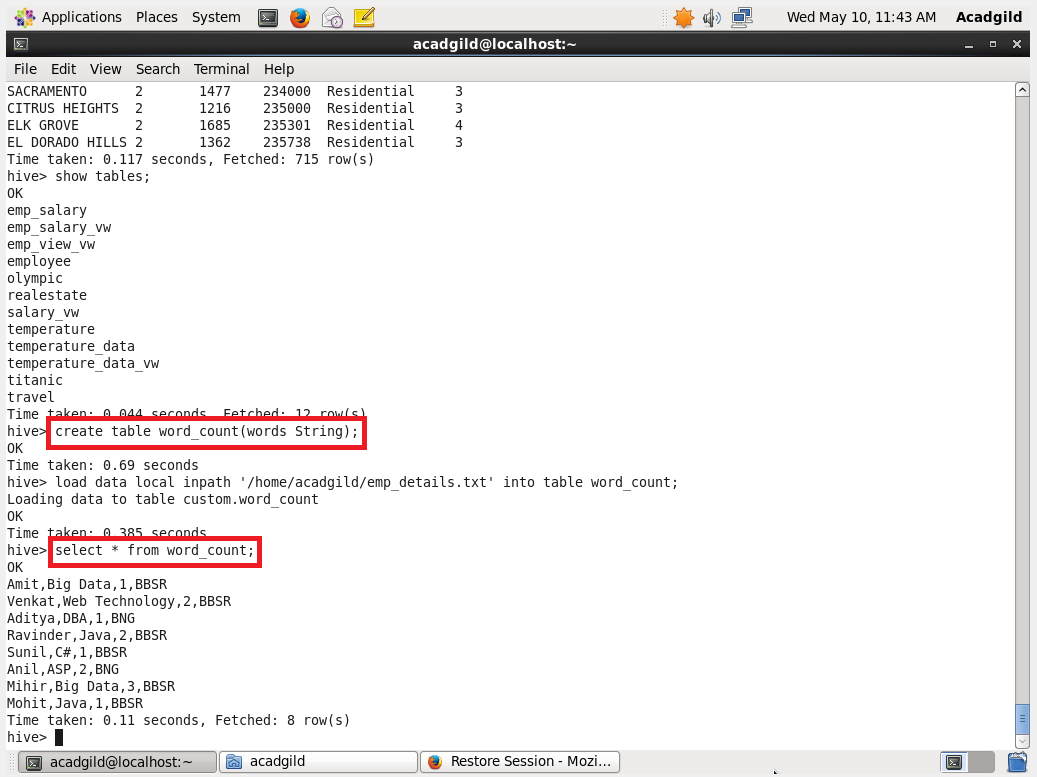


Perform word count in Hive for above given dataset.

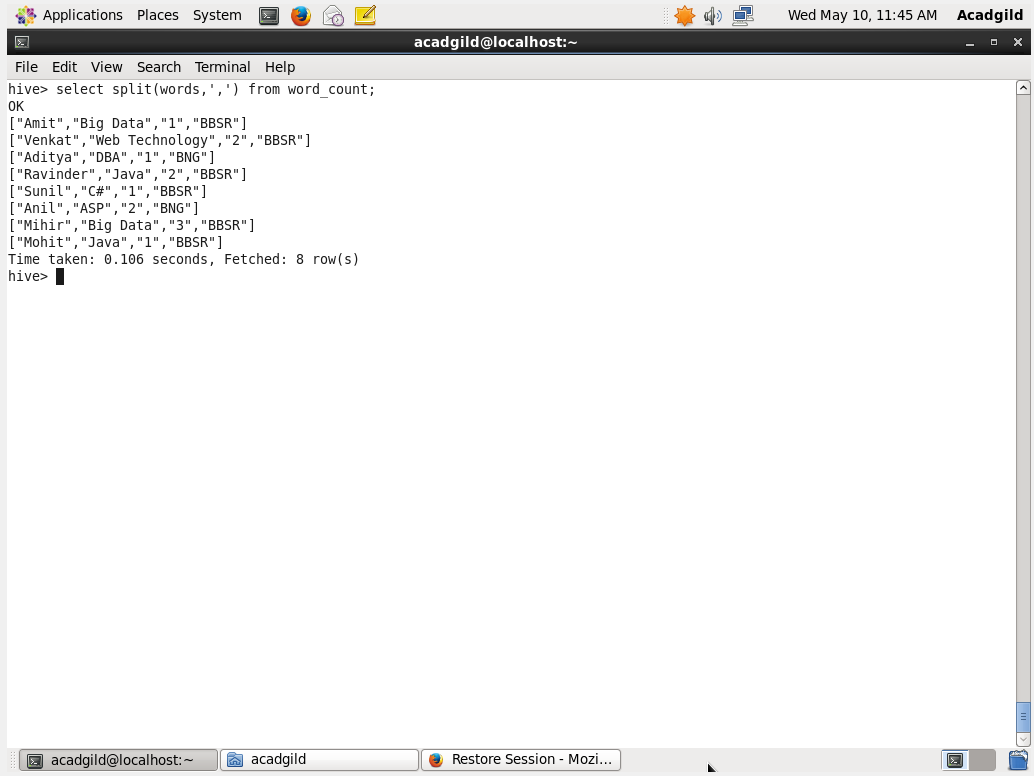
WC\_emp\_details.



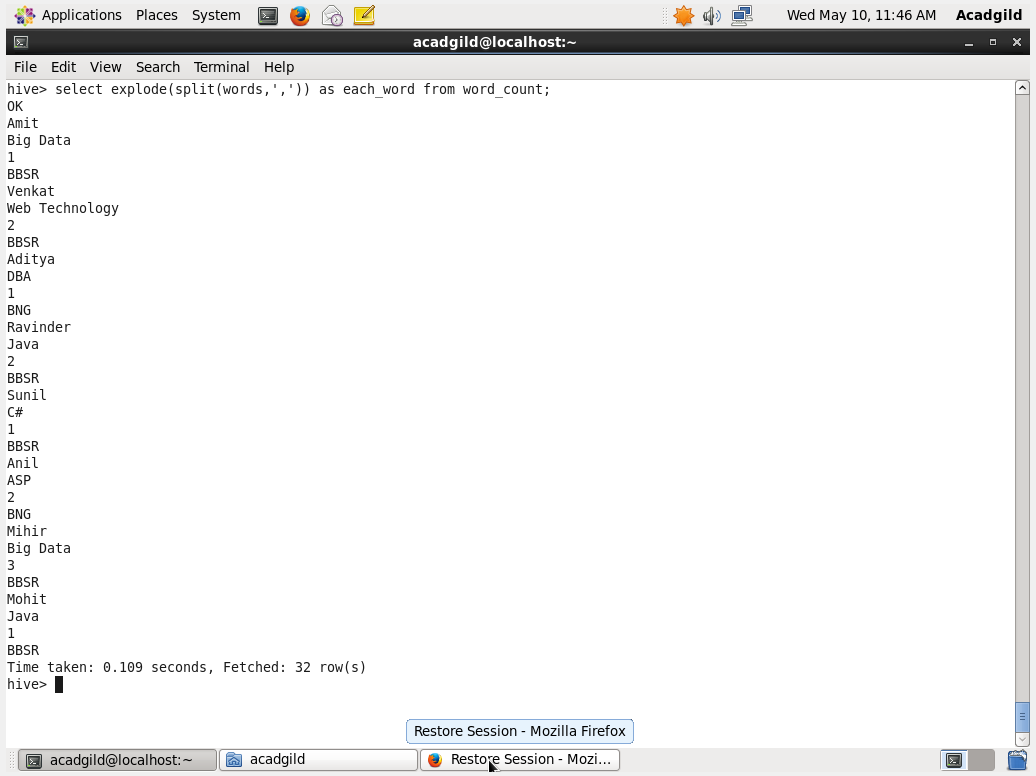
A table named word\_count is created with a column named words. The employee\_details data is loaded into the table. The Contents of the table is displayed.



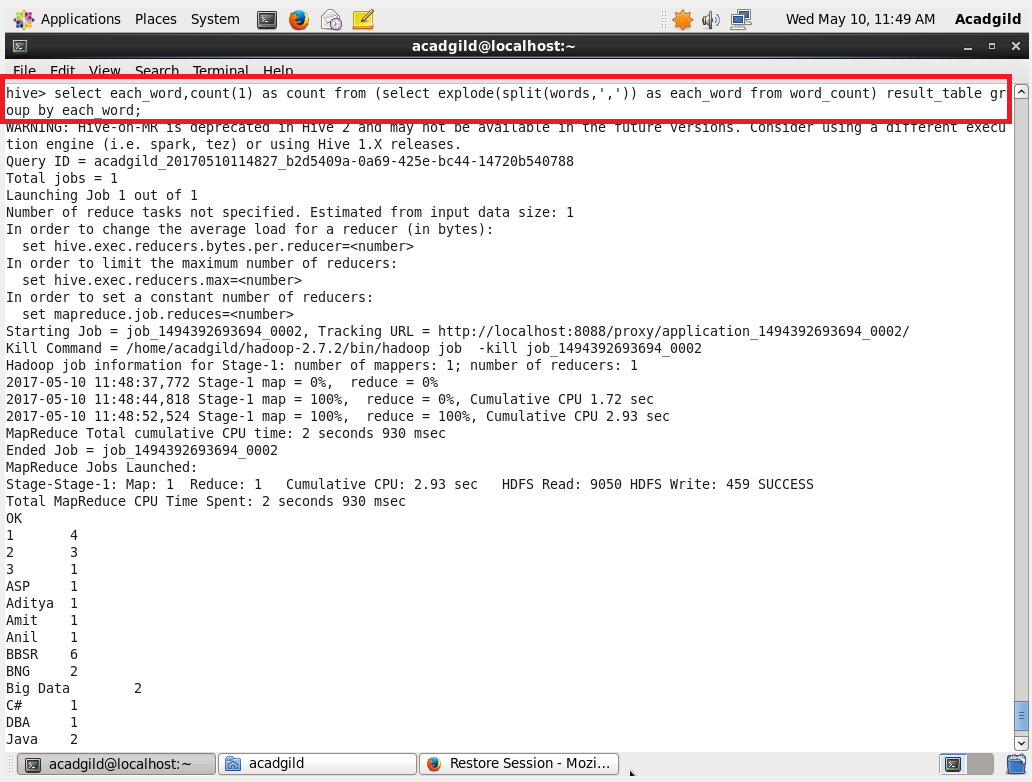
The data is splitted based on the comma.



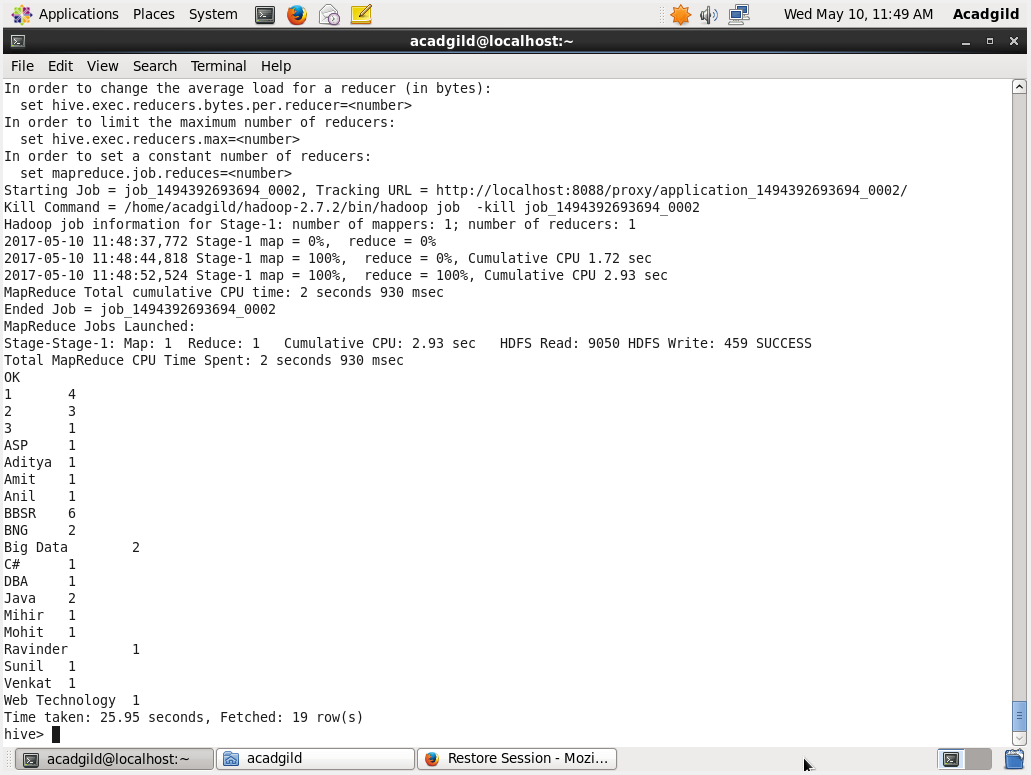
Then the words are displayed individually using the explode command.



The wordcount is performed by grouping each word.



WordCount Output:



Explain the working of Partitioning in brief.

1. Hive organizes tables into partitions. It is a way of dividing a table into related parts based on the values of partitioned columns such as date, city, and department.

2. Using partition, it is easy to query a portion of the data.

3. Tables or partitions are sub-divided into **buckets,** to provide extra structure to the data that may be used for more efficient querying.

4. Bucketing works based on the value of hash function of some column of a table.

Creating a partitioned table:

It is similar to ordinary table but just adding Partitioned By() where column based on which partitioning needs to be done

Example: A table named **employee\_data** contains employee data such as id, name, dept, and yoj (i.e., year of joining). Suppose you need to retrieve the details of all employees who joined in 2012. A query searches the whole table for the required information. However, if you partition the employee data with the year and store it in a separate file, it reduces the query processing time.

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

Static Partitioning:

Partitioning done with known fields is called as the static partitioning. We statically mention what partition has to be entered in to that partition. Static Partition saves your time in loading data compared to dynamic partition. “Statically” adding a partition in table and move the file into the partition of the table. This takes part of the database and stores into a file. Static partition is in Strict Mode.

Dynamic Partition:

Here no specific condition is given. Only the field is mentioned and the partition and the files are formed for each of the distinct entries.

For Dynamic partition we have to we have to set dynamic partition as nonstrict.