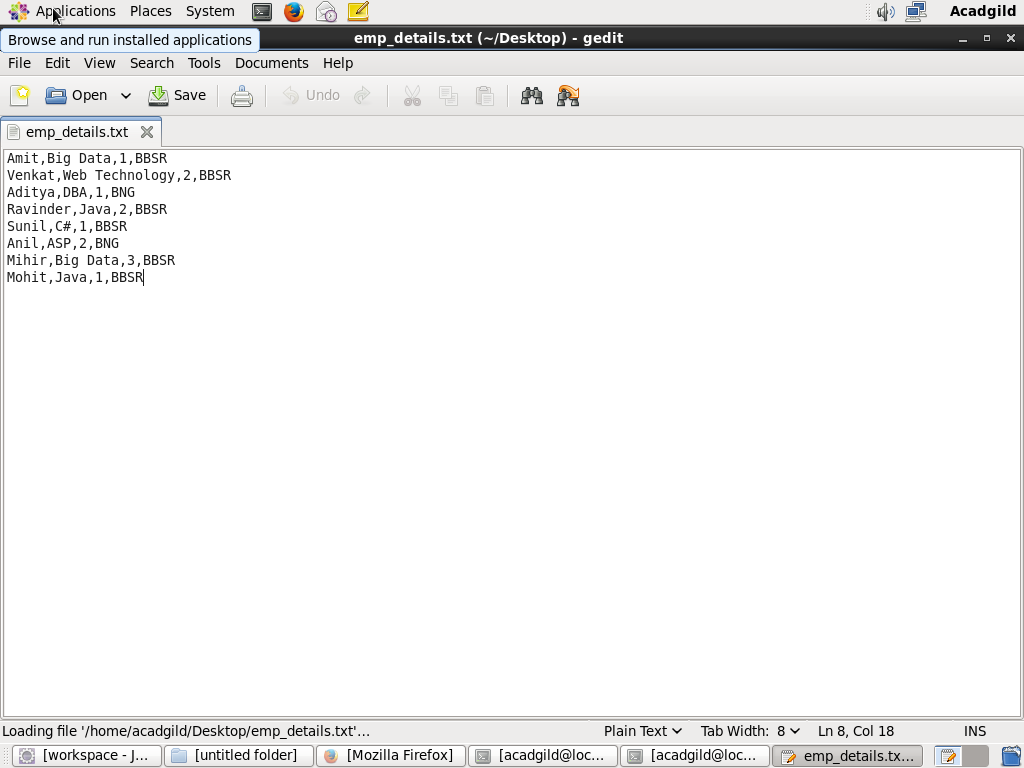
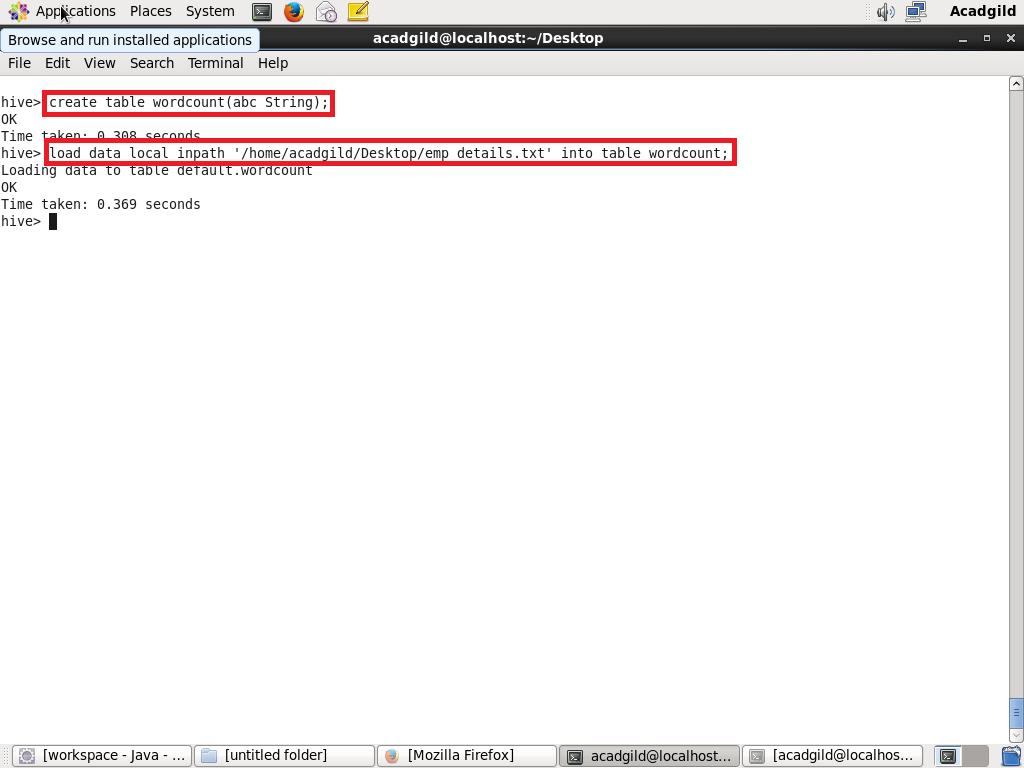
Problem Statement:

Perform word count in Hive for the given dataset.

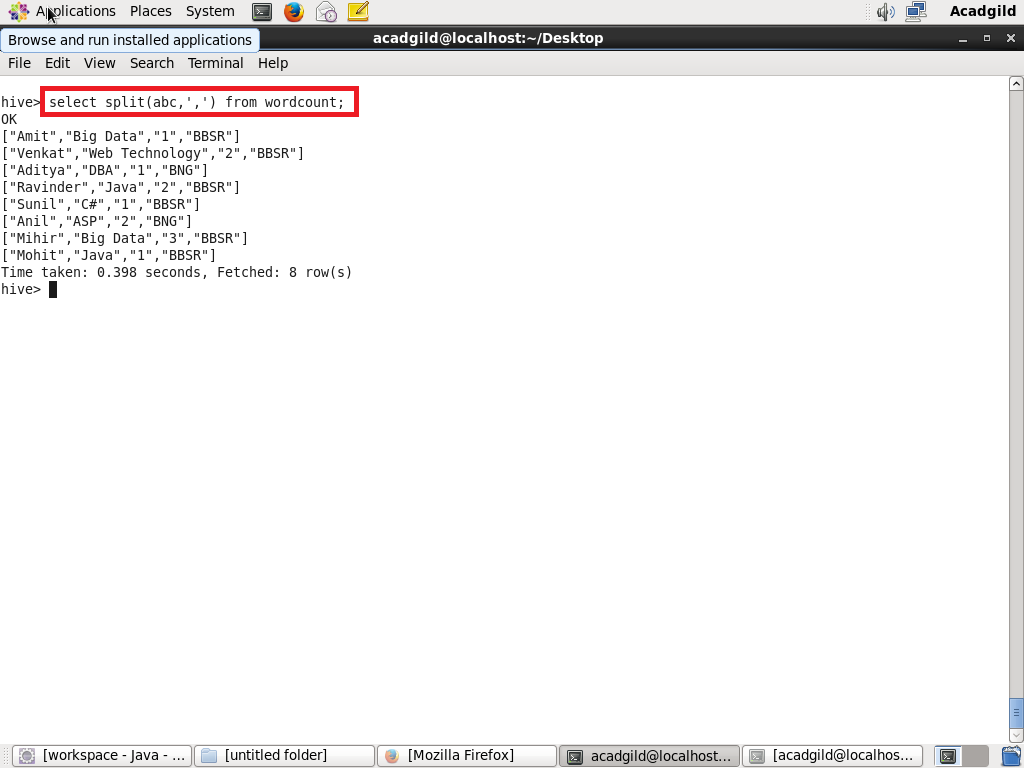
Dataset:



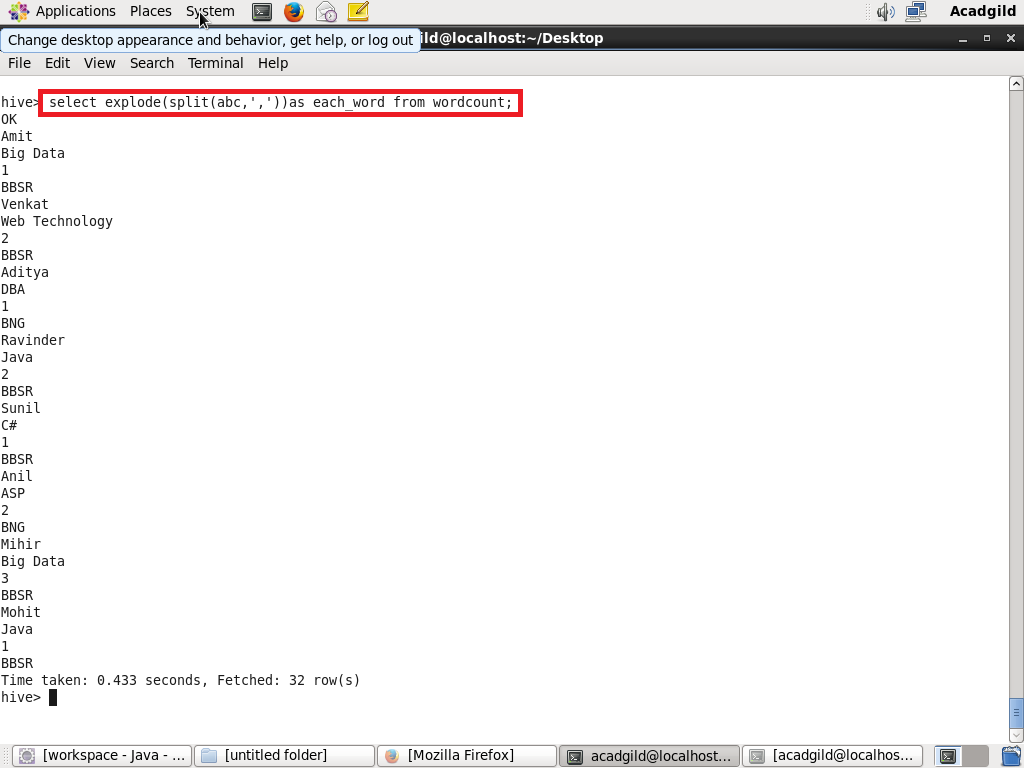
Create the table and load the data into the table:



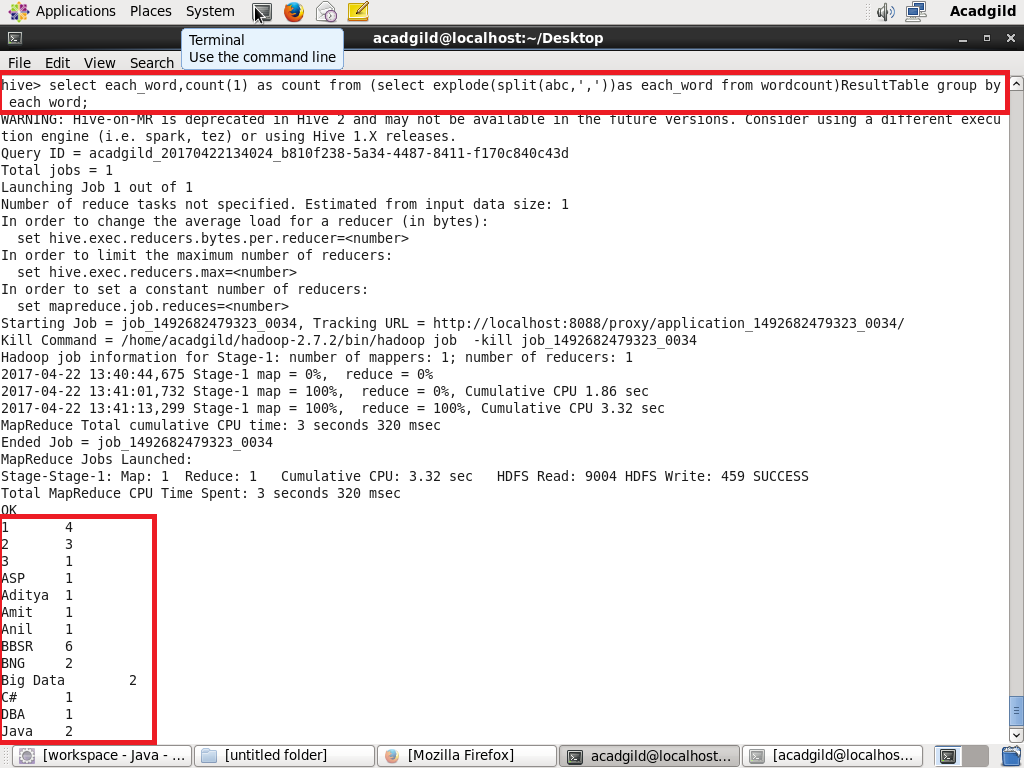
Split command:

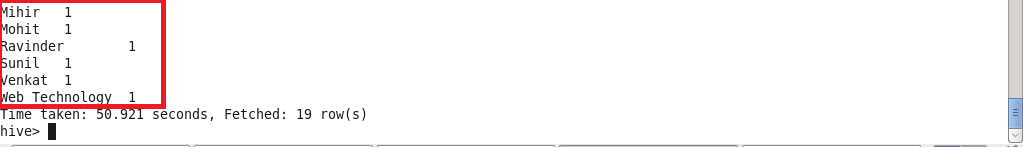


Explode command:



Count and group by command & output:





Problem Statement:

* Explain the working of Partitioning in brief.

Solution:

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

-Tables or partitions are sub-divided into buckets, to provide extra structure to the data that may be used for more efficient querying. Bucketing works based on the value of hash function of some column of a table.

-For example, a table named Table1 contains employee data such as id, name, dept, and yoj (i.e., year of joining). Suppose we need to retrieve the details of all employees who joined in 2012. A query searches the whole table for the required information. However, if we partition the employee data with the year and store it in a separate file, it reduces the query processing time. The following example shows how to partition a file and its data:

The following file contains employeedata table.

id, name, dept, yoj

1, gopal, TP, 2012

2, kiran, HR, 2012

3, kaleel,SC, 2013

4, Prasanth, SC, 2013

The above data is partitioned into two files using year.

1, gopal, TP, 2012

2, kiran, HR, 2012

3, kaleel,SC, 2013

4, Prasanth, SC, 2013

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

Solution:

Static partition in hive:

* Insert input data files individually into a partition table is Static Partition
* Usually when loading files (big files) into [Hive tables](http://www.hadooptpoint.com/hive-create-table-examples/) static partitions are preferred
* Static Partition saves your time in loading data compared to dynamic partition
* “Statically” add a partition in table and move the file into the partition of the table.
* We can alter the partition in static partition
* We can get the partition column value form the filename, day of date etc without reading the whole big file.
* If we want to use Static partition in hive we should set property set hive.mapred.mode = strict This property set by default in hive-site.xml
* Static partition is in Strict Mode
* We should use where clause to use limit in static partition
* We 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 we have large data stored in a table then Dynamic partition is suitable.
* If we 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
* We can perform dynamic partition on hive external table and managed table
* If we want to use Dynamic partition in hive then mode is in nonstrict mode
* Here is hive dynamic partition properties you should allow