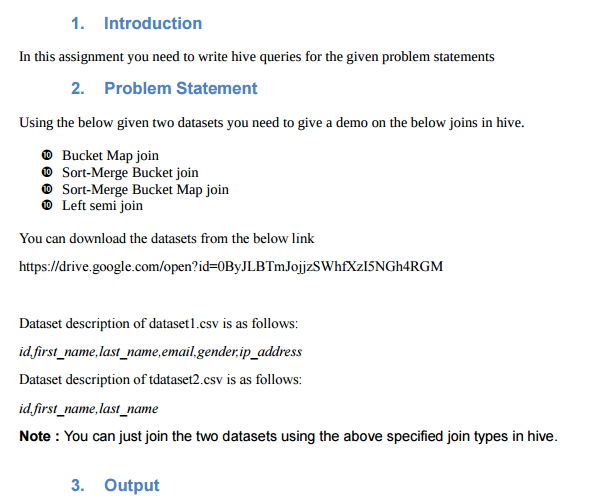
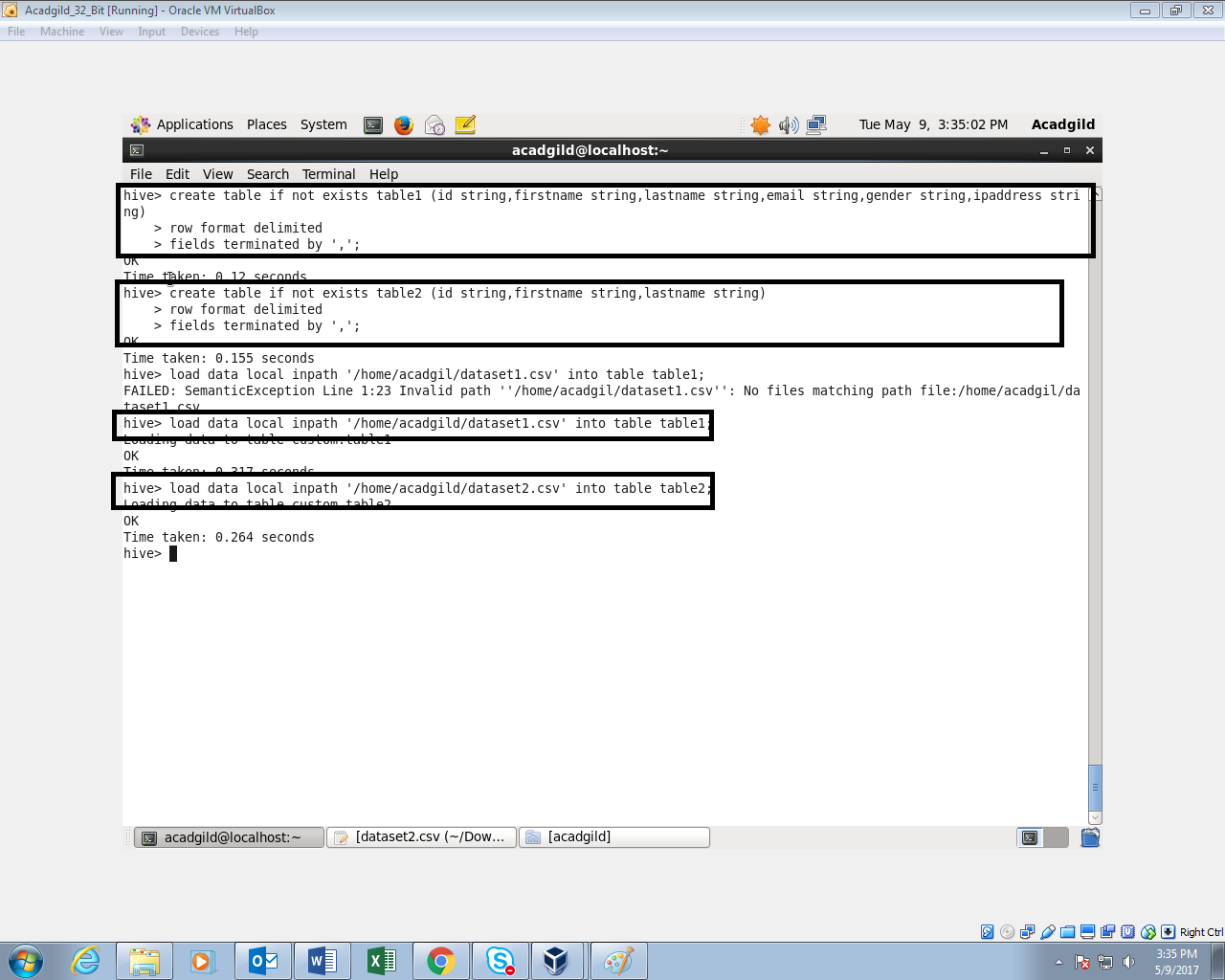
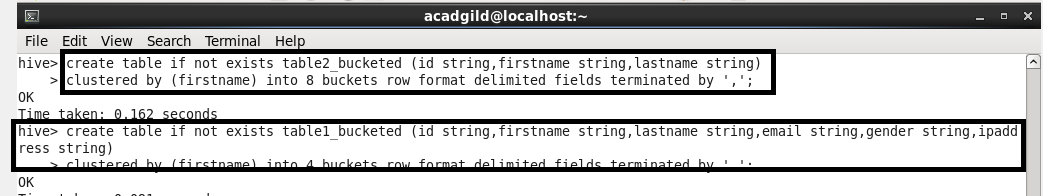
**ASSIGNMENT 27.6**



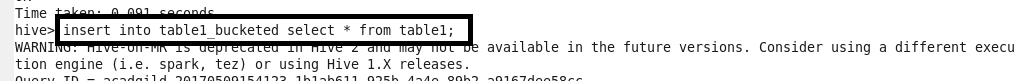
Two tables are created and data is loaded into tables



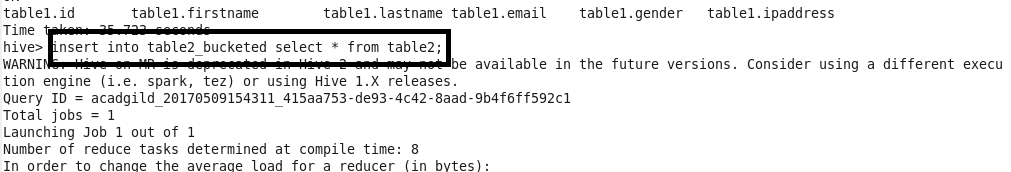
Two bucketed tables are created.one with four buckets and the other with eight buckets



Inserting the data in table1 into the bucketed table table1\_bucketed

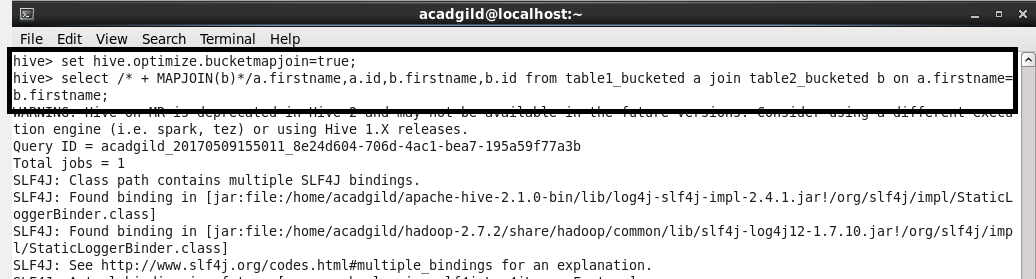


Inserting the data in table2 into the bucketed table table2\_bucketed

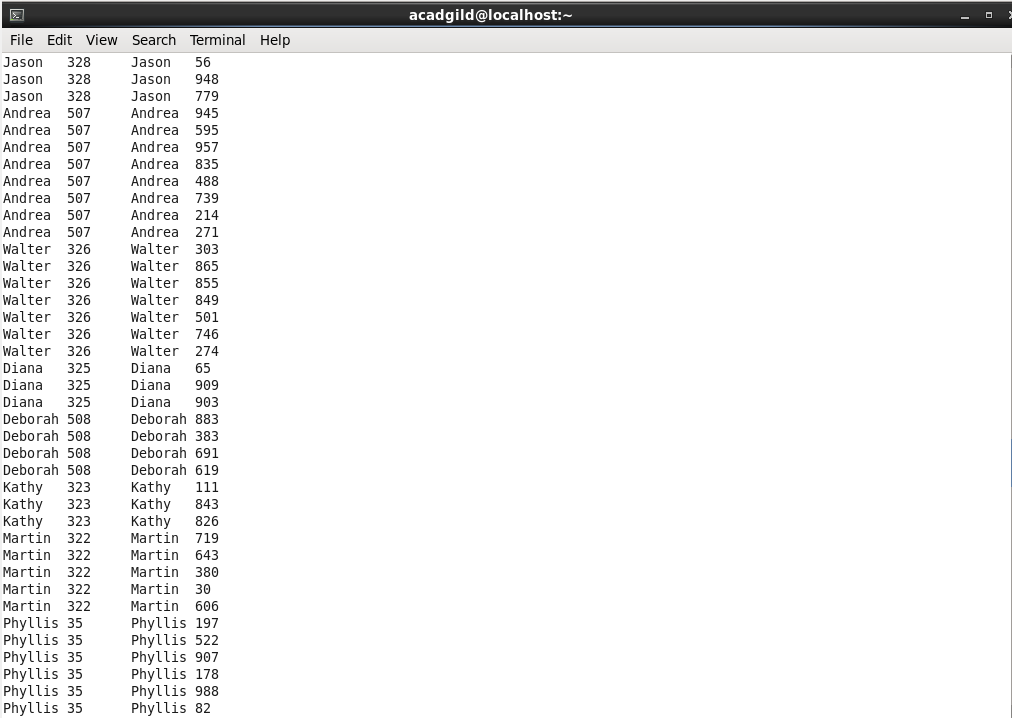


For performing Bucket-Map join, we need to set this property in the Hive shell.

**set hive.optimize.bucketmapjoin = true**

****

OUTPUT:

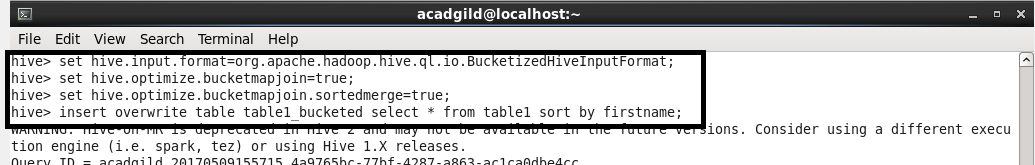


**SORT-MERGE BUCKET JOIN:**

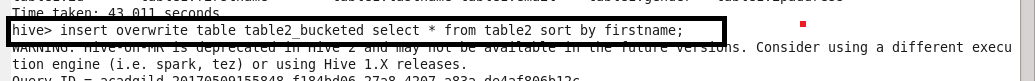
Sort-Merge bucket join is like reduce side join. Before joining we’ll have to sort the data and then only merge.

**Setting configuration:**

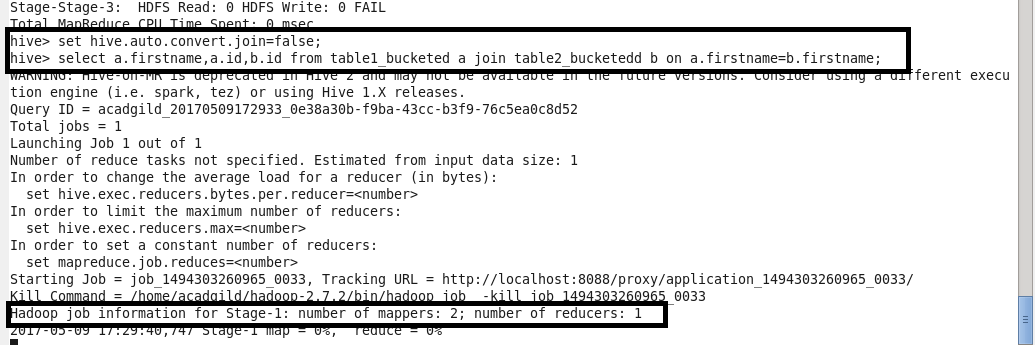
set hive.auto.convert.join= false;



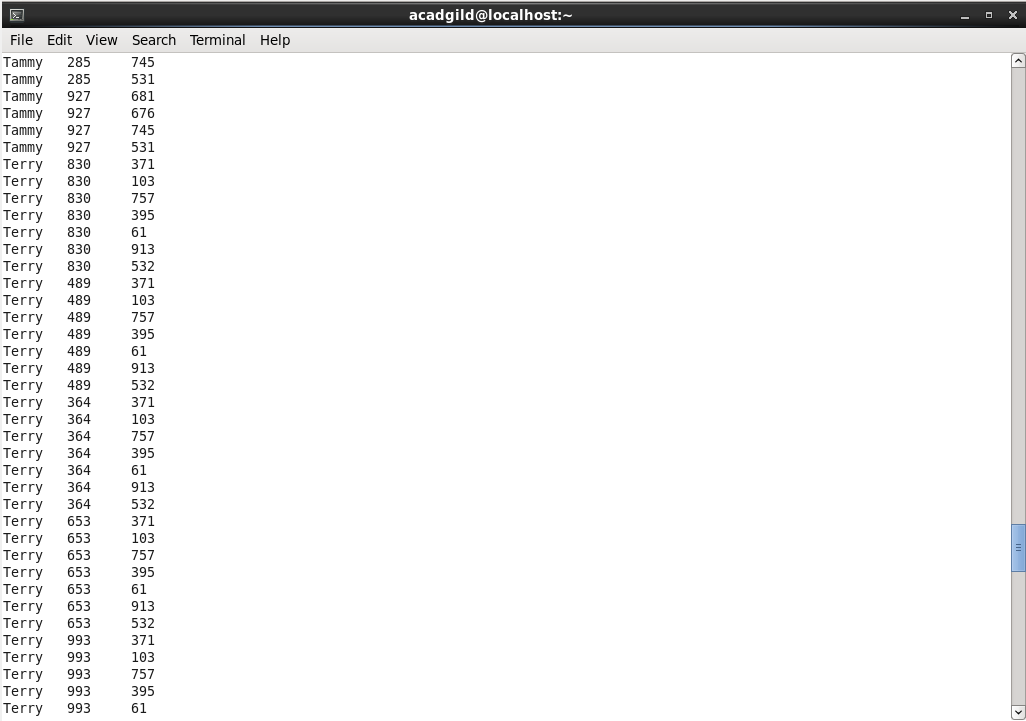
Sorting the contents of bucketed table 2



Writing Query:

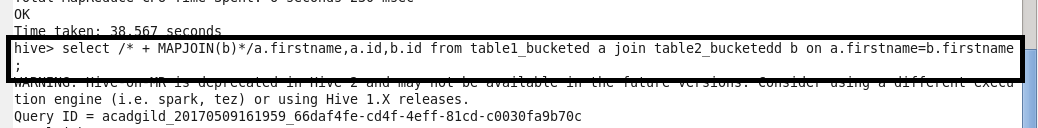


OUTPUT:

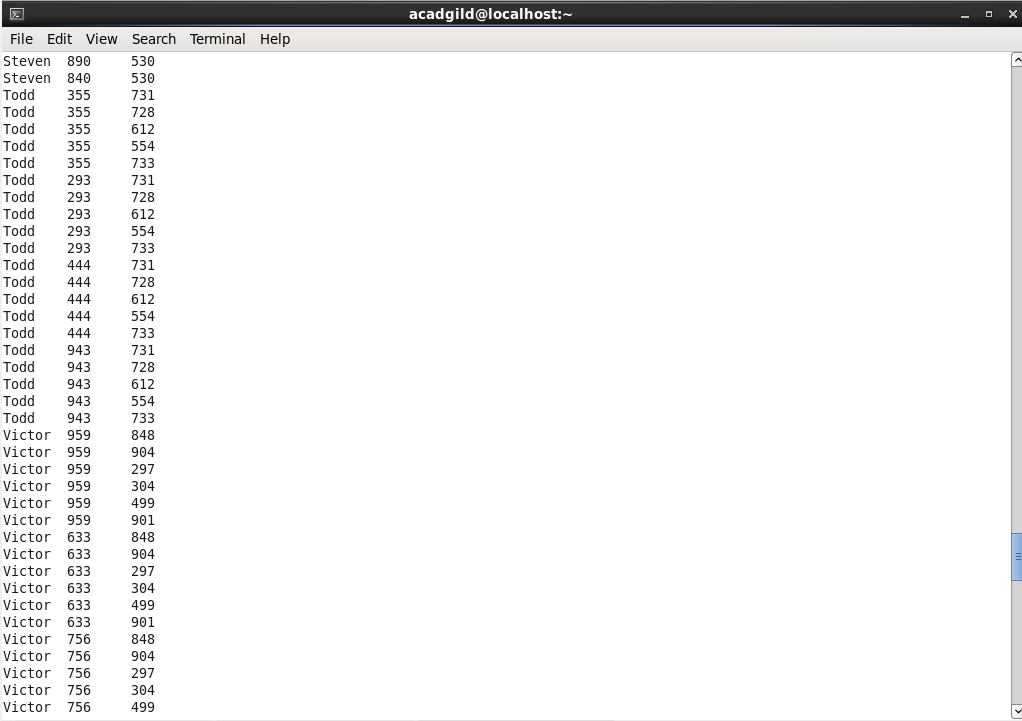


**SORT-MERGE BUCKET MAP JOIN:**

**set hive.auto.convert.join= true;**

****

**OUTPUT:**

****

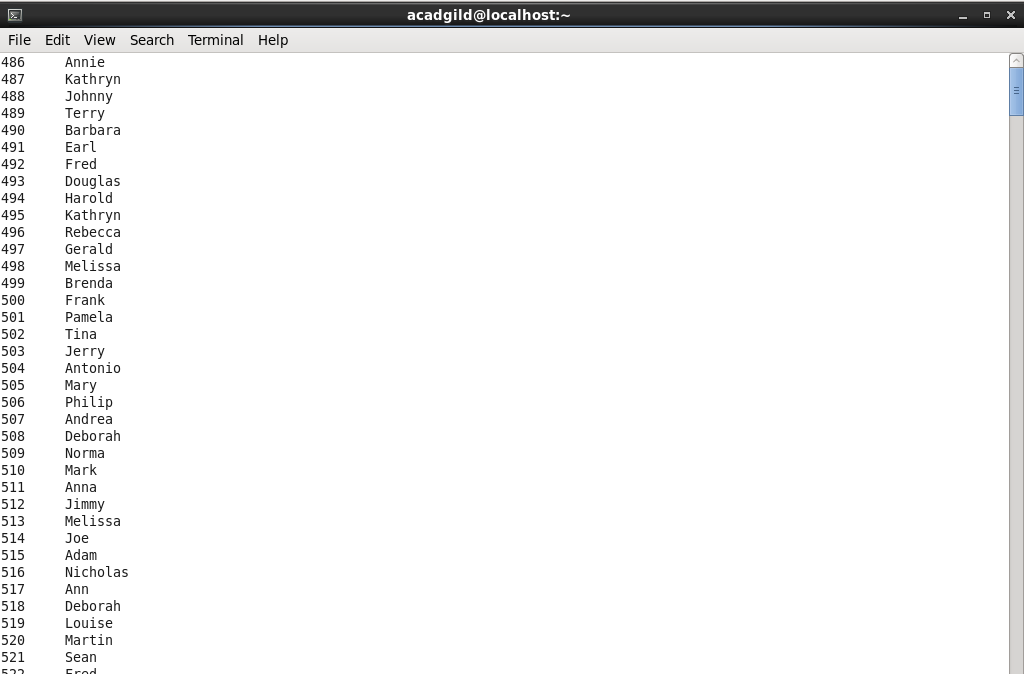
**LEFT SEMI JOIN:**

The left semi join is used in place of the IN/EXISTS sub-query in Hive. In a traditional RDBMS, the IN and EXISTS clauses are widely used whereas in Hive, the left semi join is used as a replacement of the same.

**QUERY:**



**OUTPUT:**

****