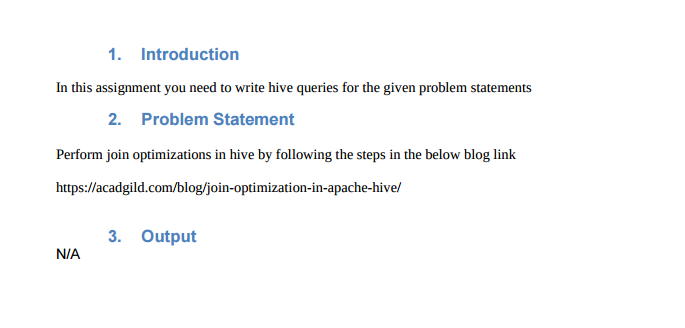
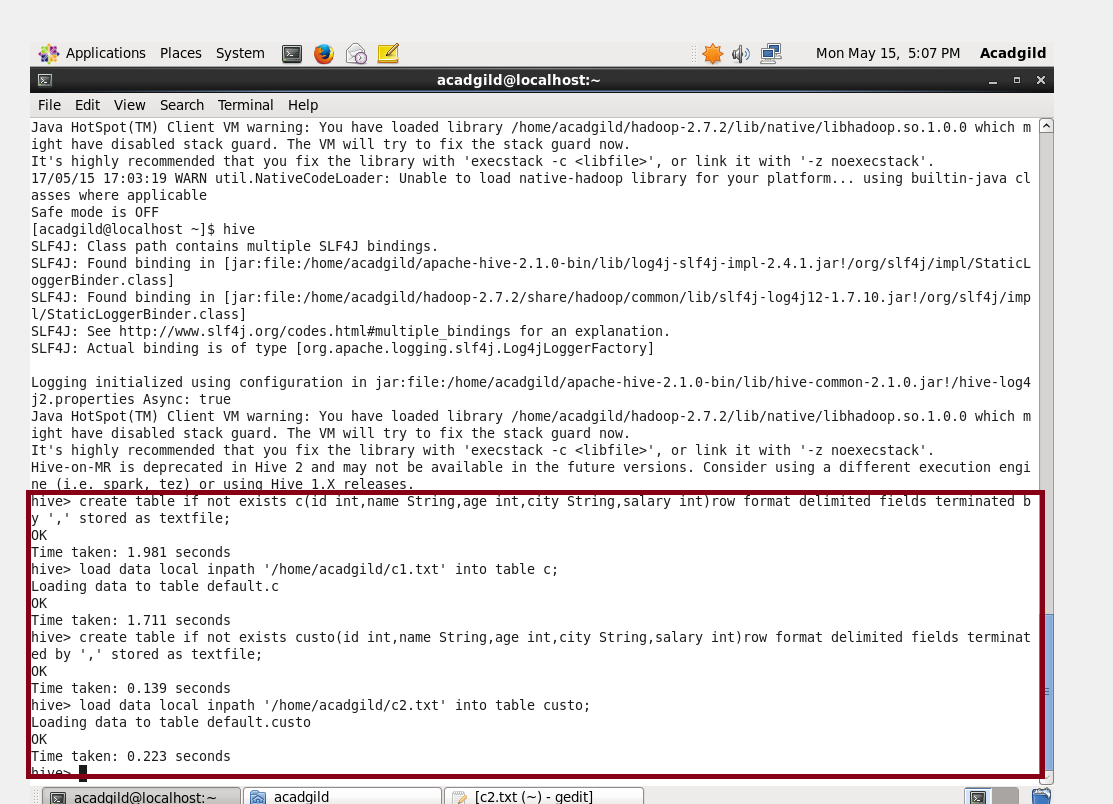
**ASSIGNMENT 27.5**

****

*Creating tables and loading data into the tables*



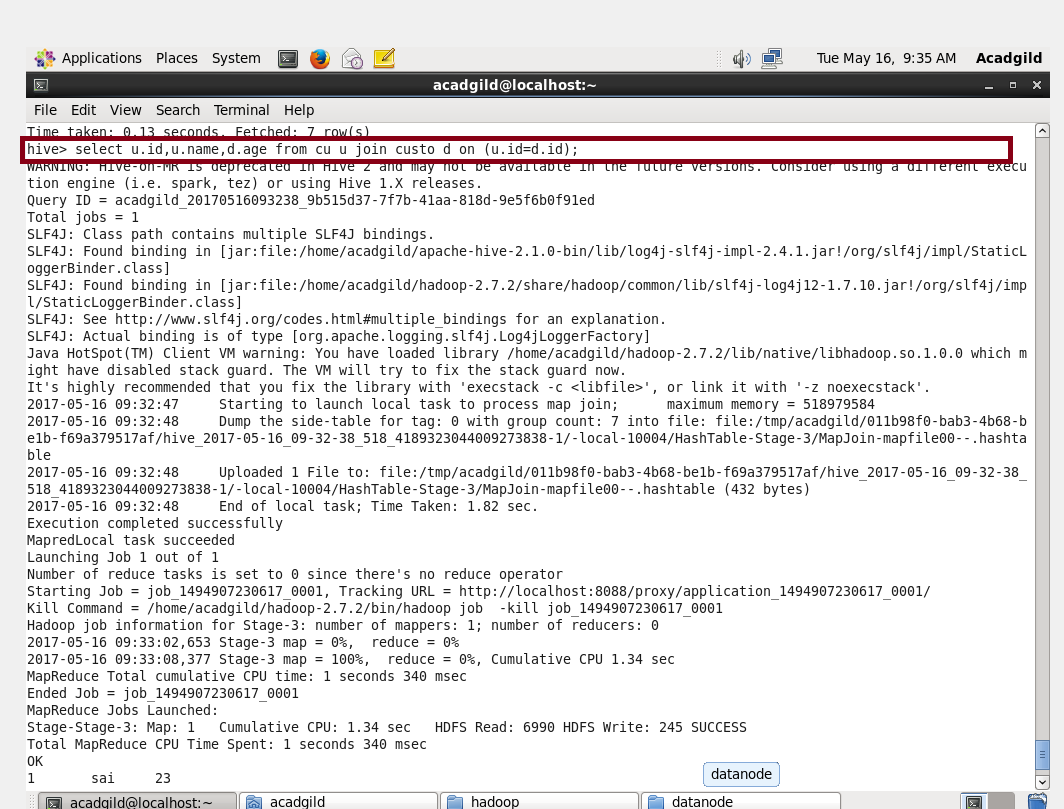
**DATA LOADED:**

* *When Hive executes a join, it has to select a table which is streamed.*
* *Hive will only take the last table in JOIN statement for streaming,*
* *And so the streaming table is largest among the two.*
* *Consider there are 2 tables.*

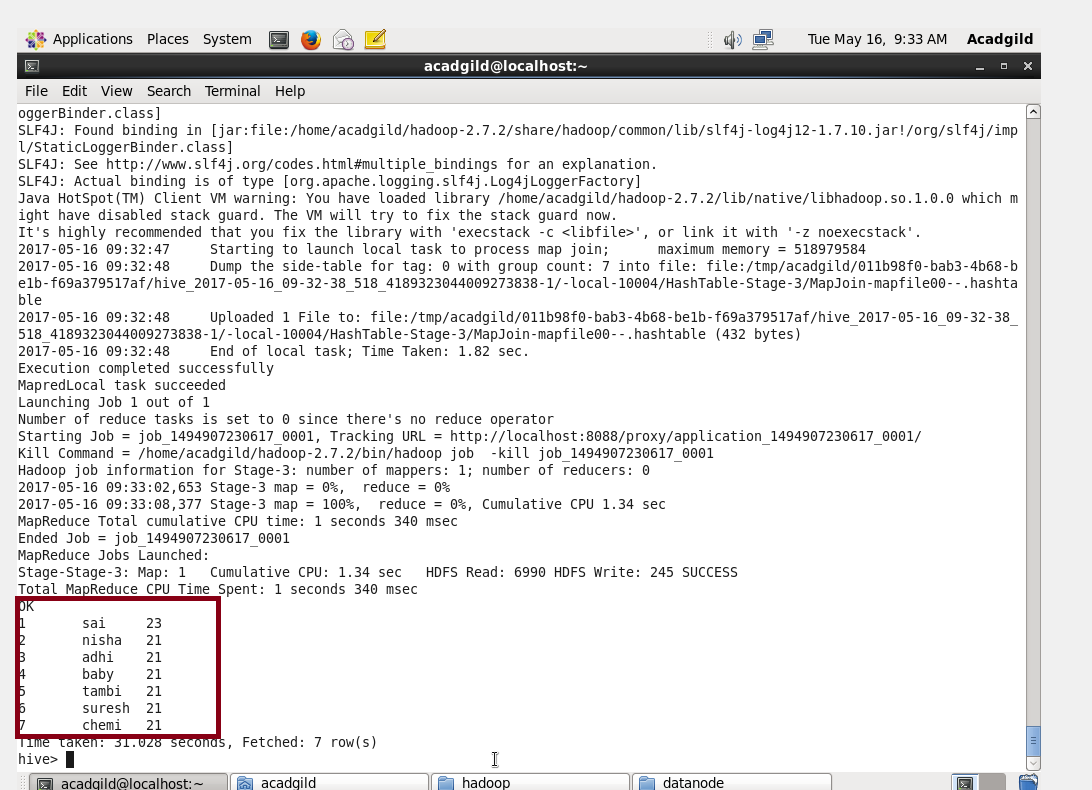
*1.) cu*

*2.)custo.*

* *Employee table large than department table.*
* *One need to set large table at the last.so here since both tavle are of same size we can consider anything.*
* *This is a method to optimize the joins in hive.*

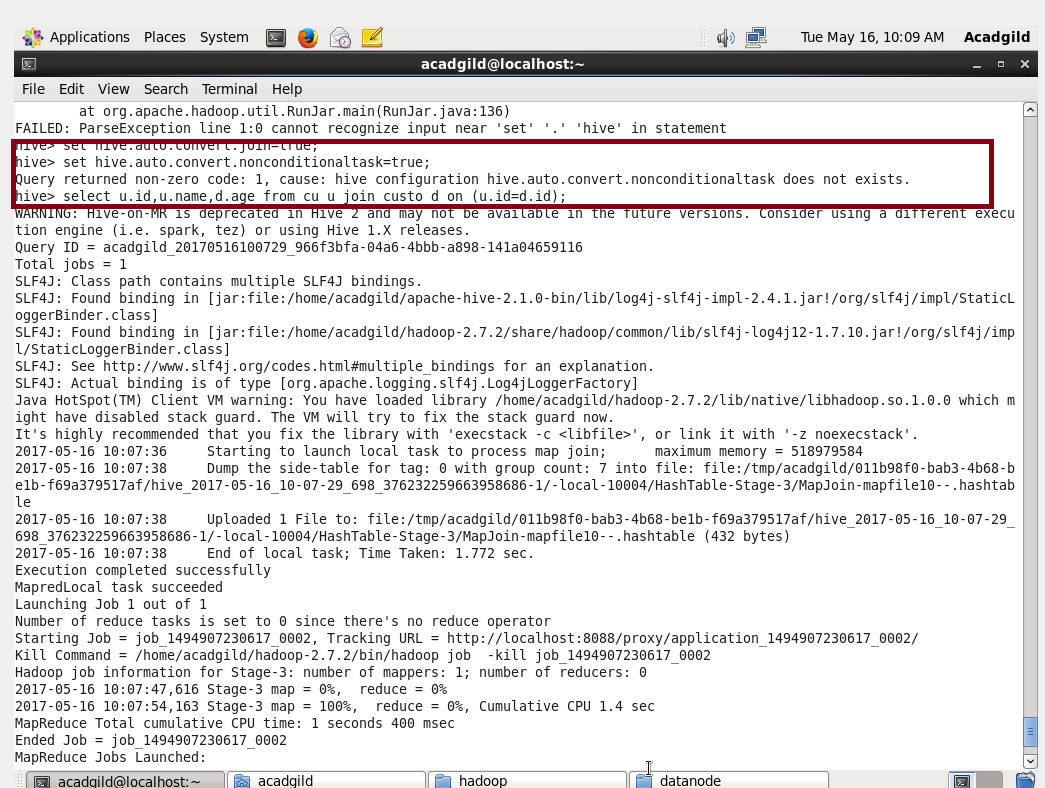


**OUTPUT:**

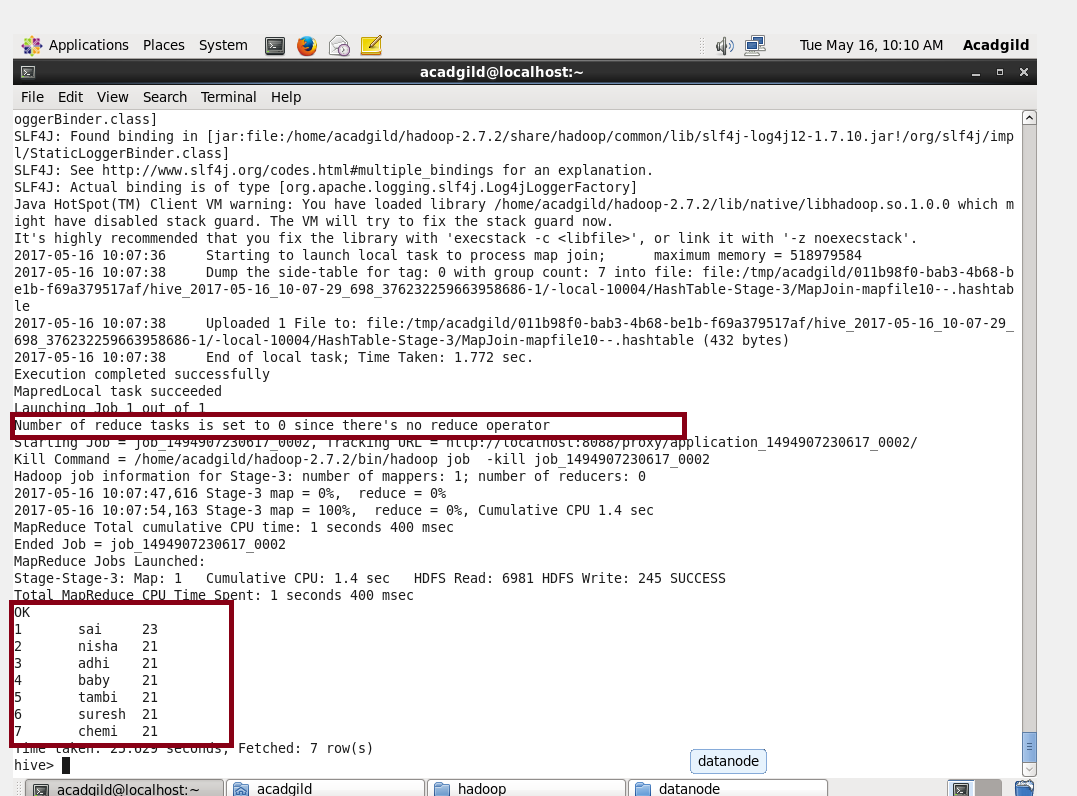


**Map Side Join:**

* ***In map side join the shell is set as***
* ***Set hive.auto.convert.join=true;***
* ***Once this set, map job will automatically will get start and run.***
* ***In case there is no reducer, then the time will get reduced.***



**From the below pic you can see that there is no reduce operation.**

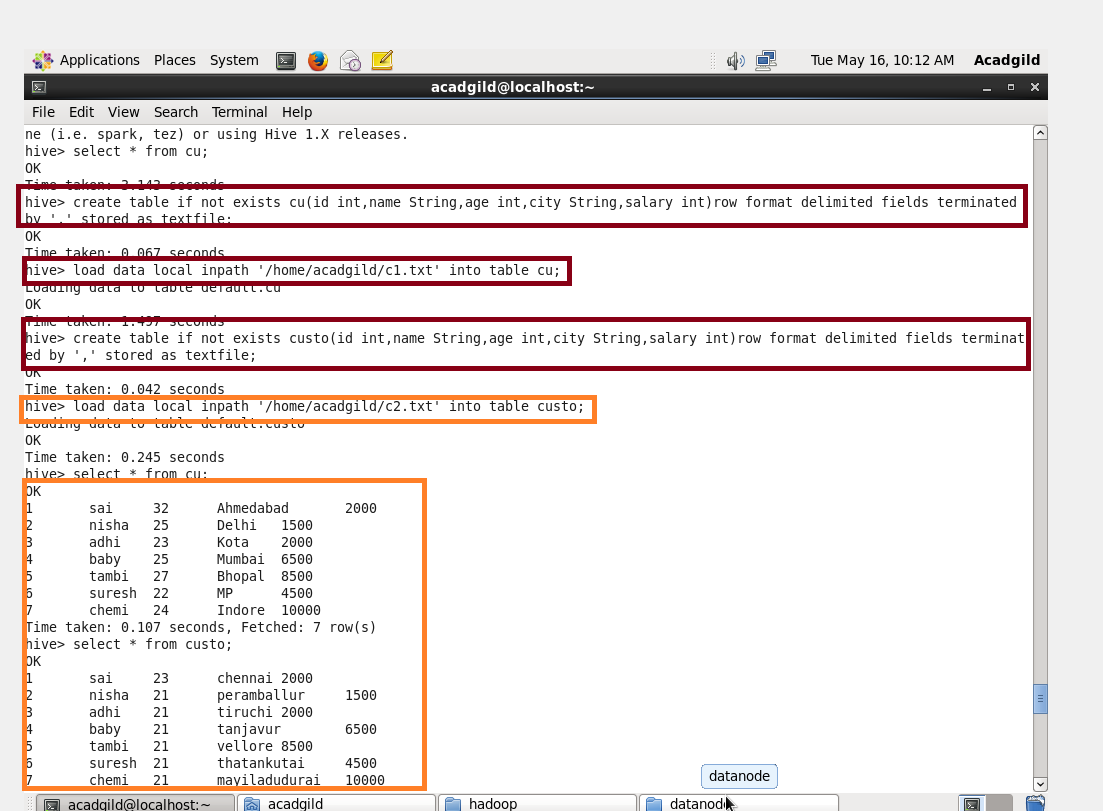


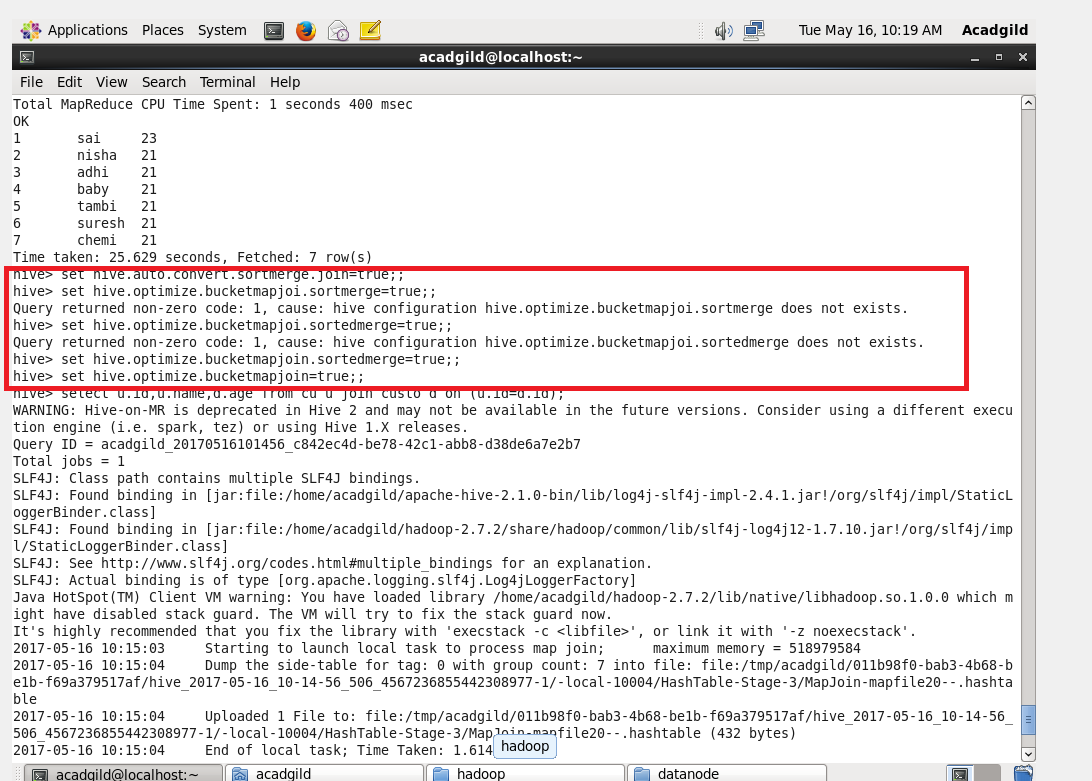
**Sort-Merge-Bucket (SMB) Map Join:**

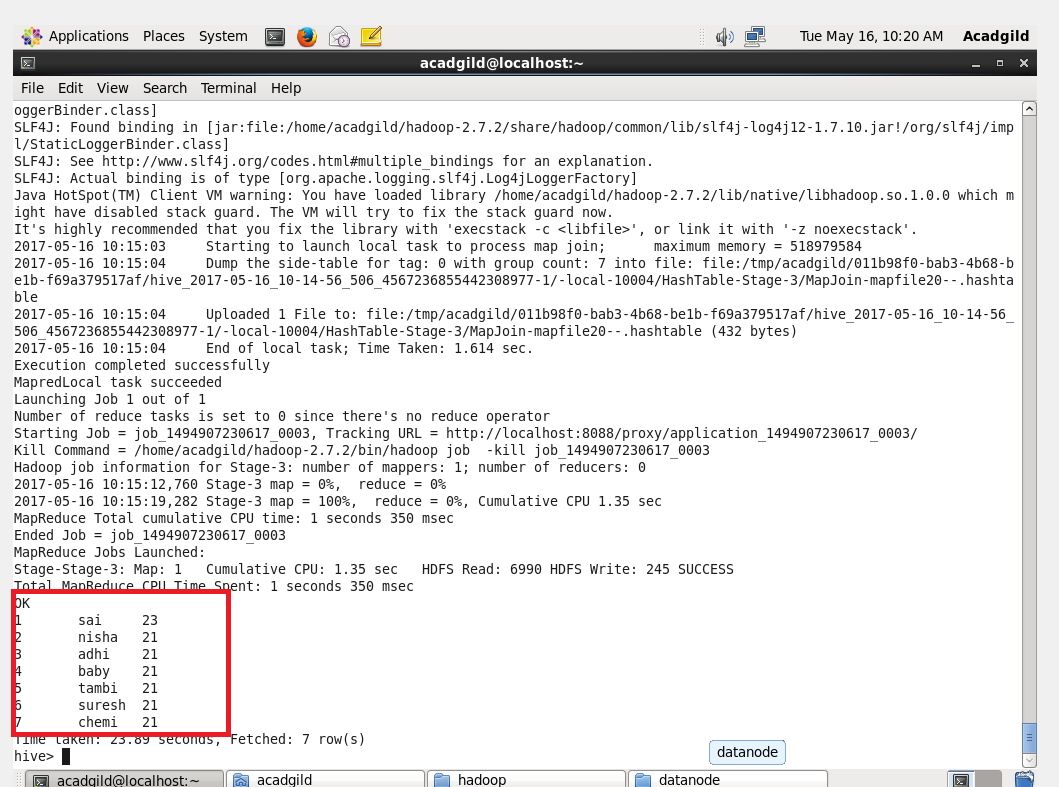
***We cannot use normal tables for this join.***

***We need to create a bucketed tables , sort the contents and then only we should join those tables .***

***Creating a bucketed table with 4 buckets, sorted it by deptid and loading data into it using insert command as well another bucketed table with 4 buckets , sorted it by deptid and loading data into it using insert command***

****



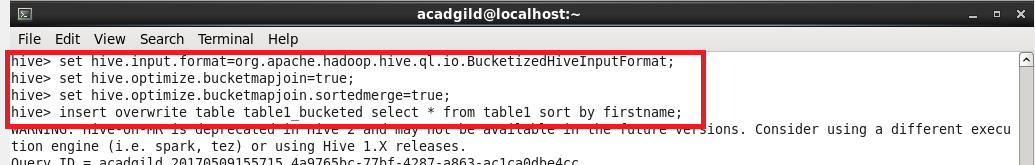


***SORT-MERGE BUCKET JOIN:***

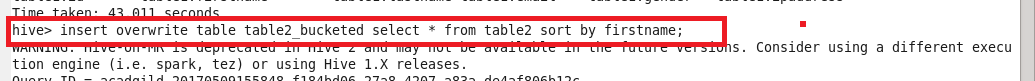
* *Sort-Merge bucket join will be like reduce side join.*
* *Before joining we should sort the data and then only we must merge that .*
* *Moreover in this join we should set*

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

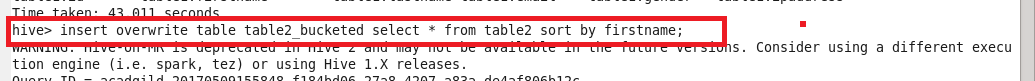
* *because we want run the reducer.*



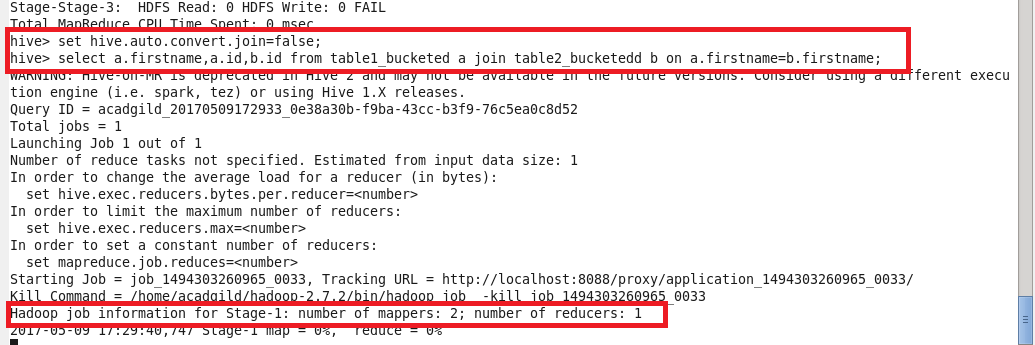
***Sorting the contents of bucketed table 2***



***Sorting the contents of bucketed table 1***



***Writing query for extracting data***



**OUTPUT:**