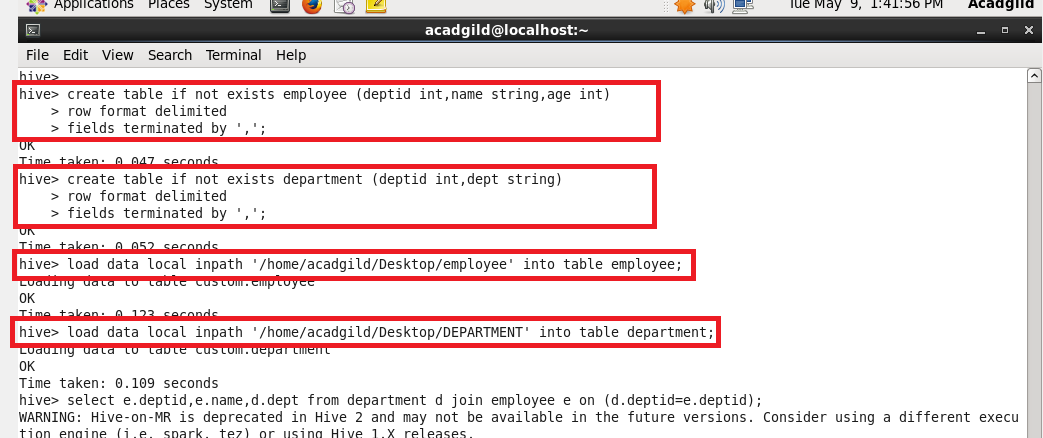
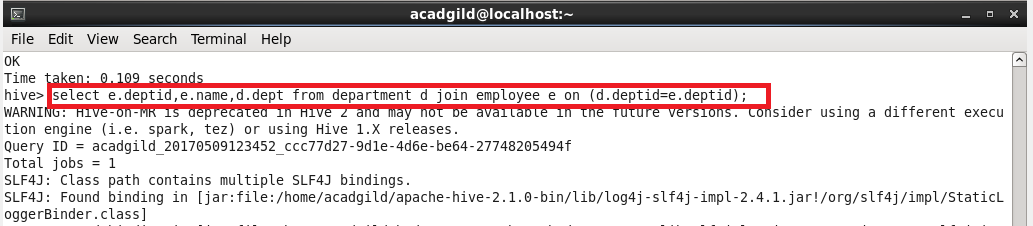
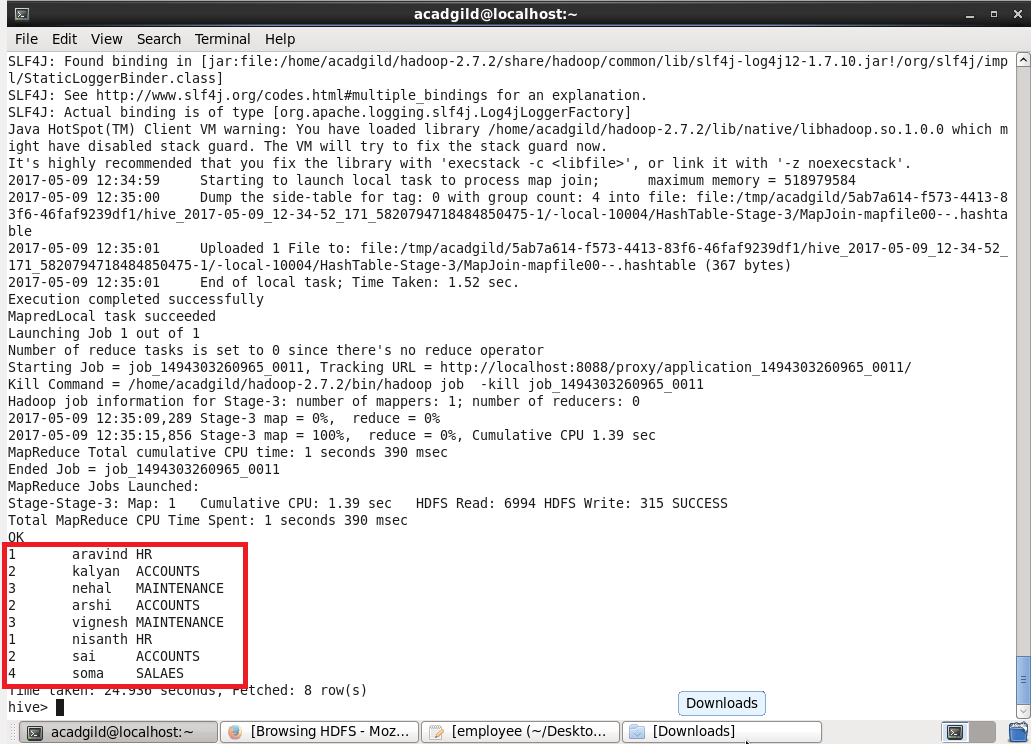
**Join table ordering (Largest table last):**

Creating tables and loading data into the tables





Output:



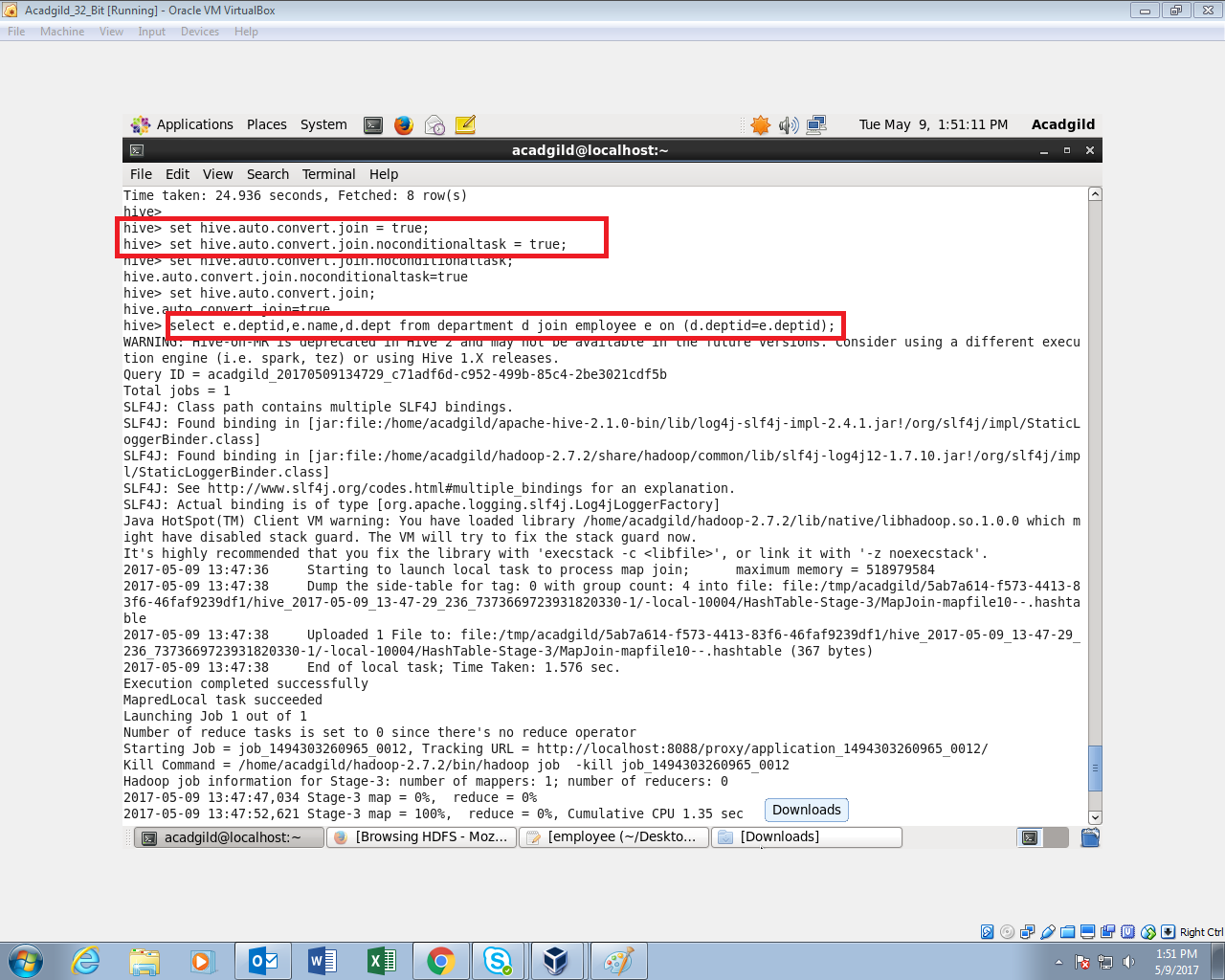
**Map Side Join:**

**In map side join we need to set this in shell**

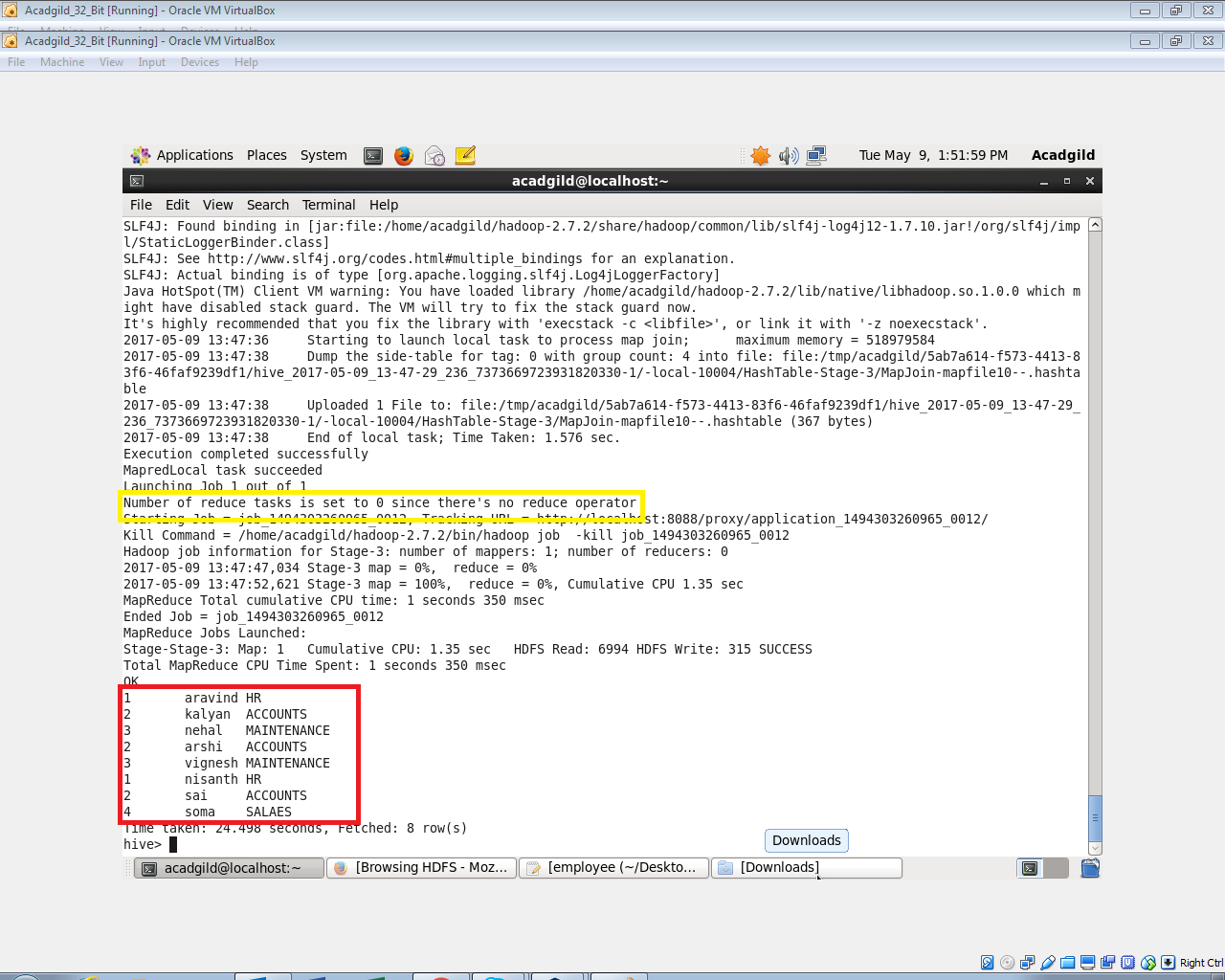
**Set hive.auto.convert.join=true;**

**If we set this, automatically map only job gets started and runs.**

**If there is no reducer, obviously the time gets reduced.**



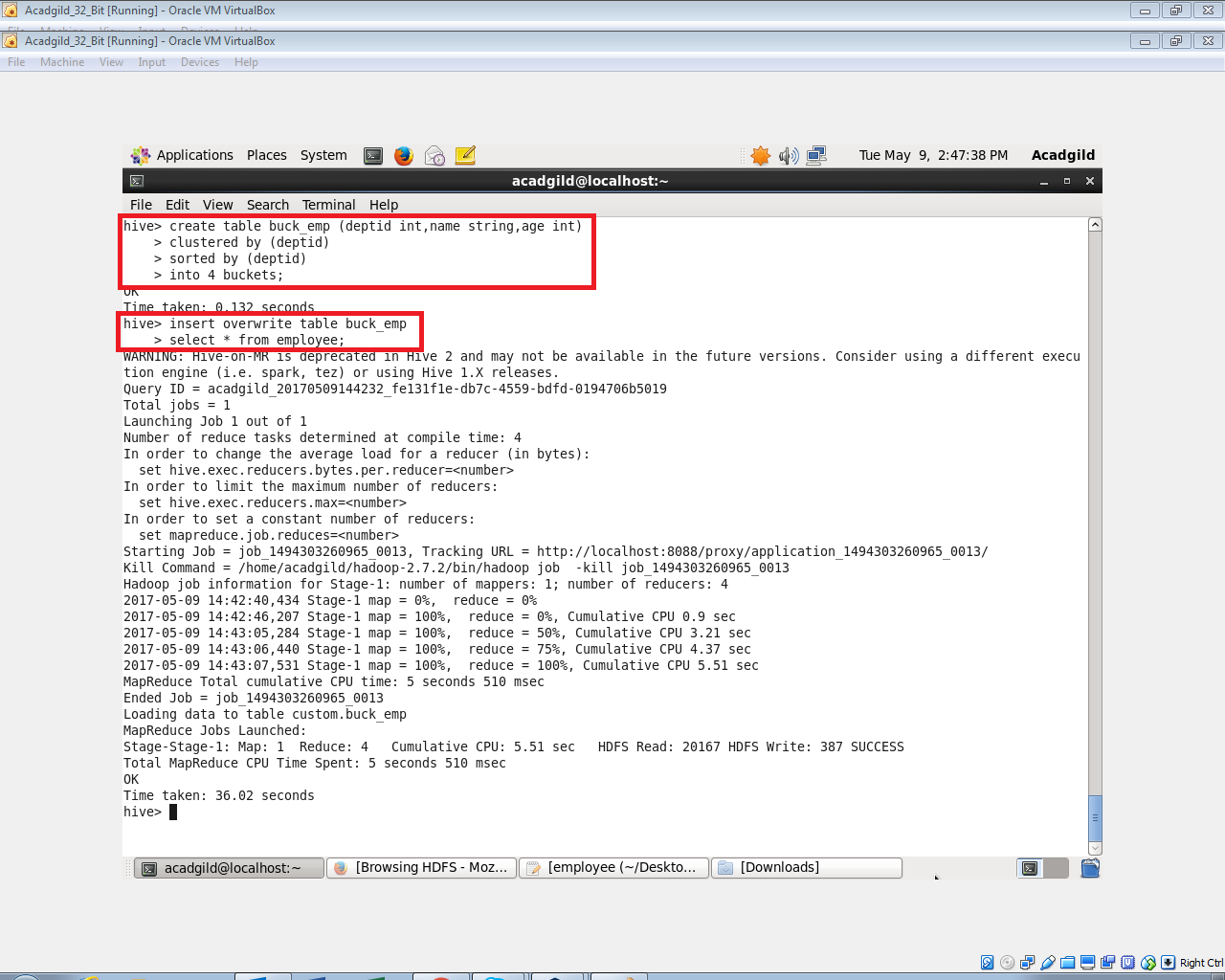
we can see that there is no reduce operation.



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

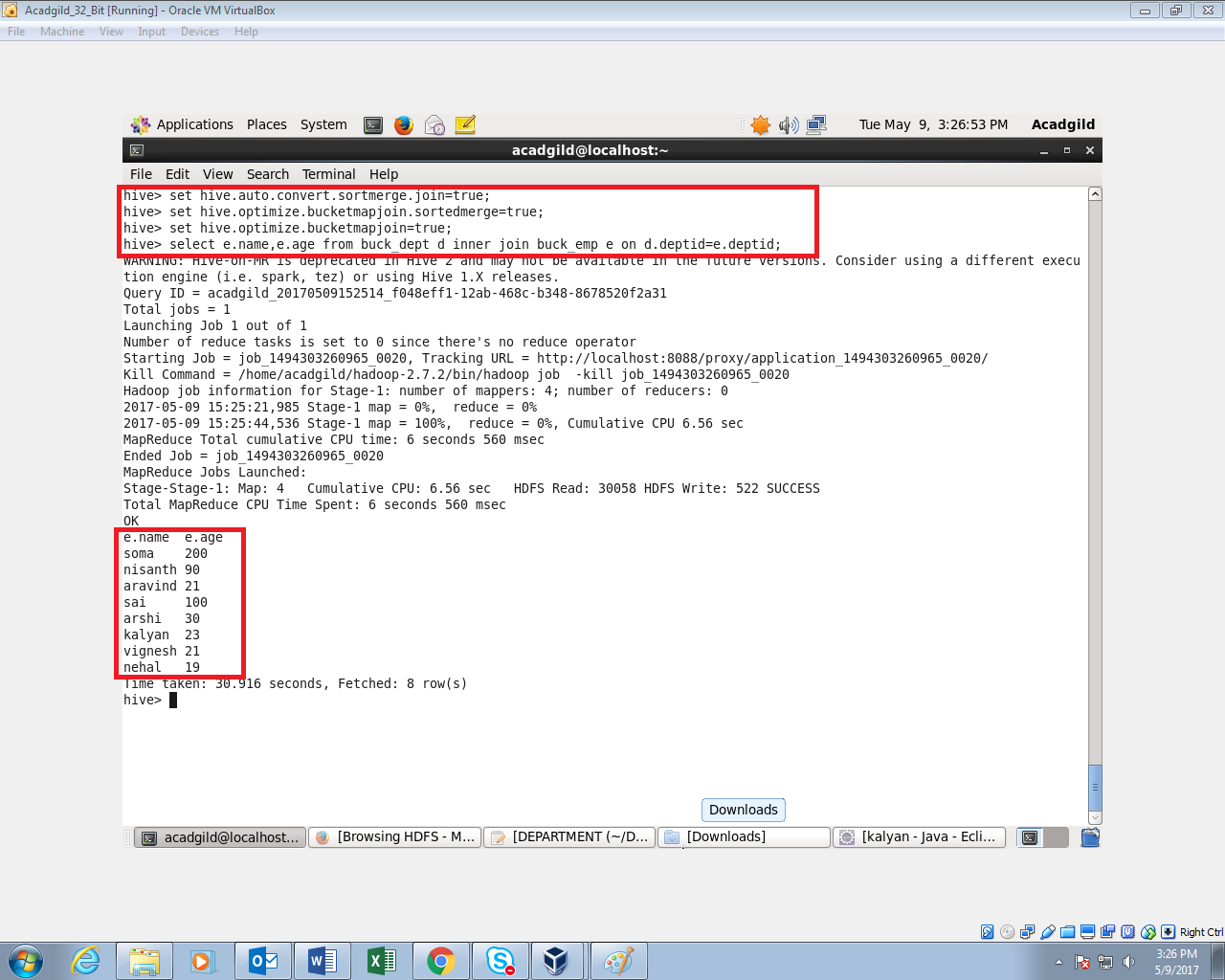
**We cannot use normal tables for this join. We need to create a bucketed tables, sort the contents and then only we should join those tables .**

**I’ve Created a bucketed table with 4 buckets, sorted it by deptid and load data into it**



**Creating another bucketed table with 4 buckets , and sorted it by deptid and loading data**



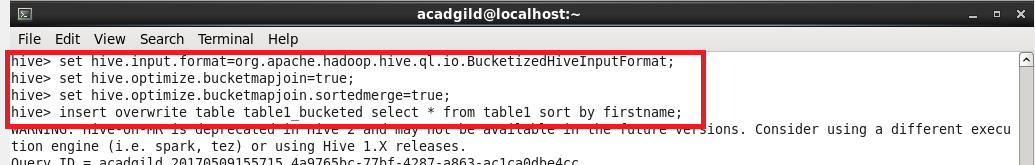


SORT-MERGE BUCKET JOIN:

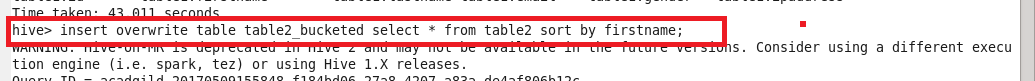
Sort-Merge bucket join is like reduce side join. Before joining we’ll have to sort the data and then only merge.

We have to set following

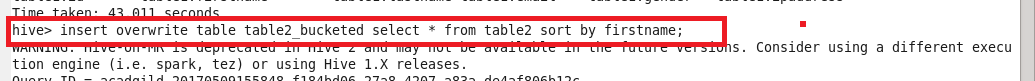
set hive.auto.convert.join= false;



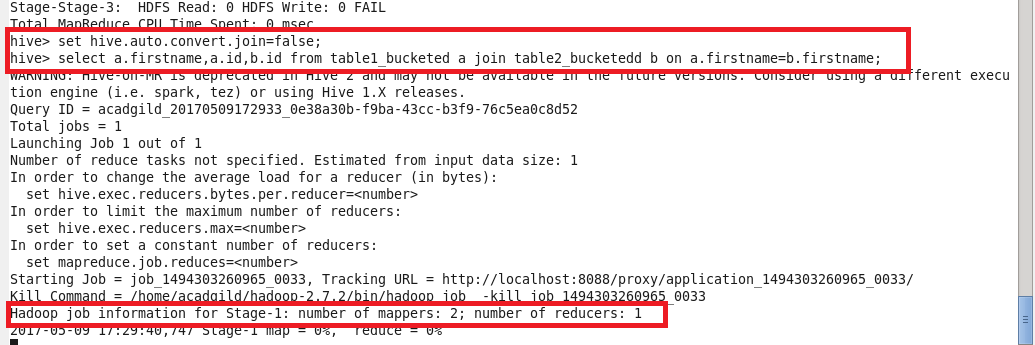
Sorting the contents of bucketed table 2



Sorting the contents of bucketed table 1



Extracting data from table:



Output:

