**ASS-27.6**

**Using the below given two datasets you need to give a demo on the below joins in hive.**

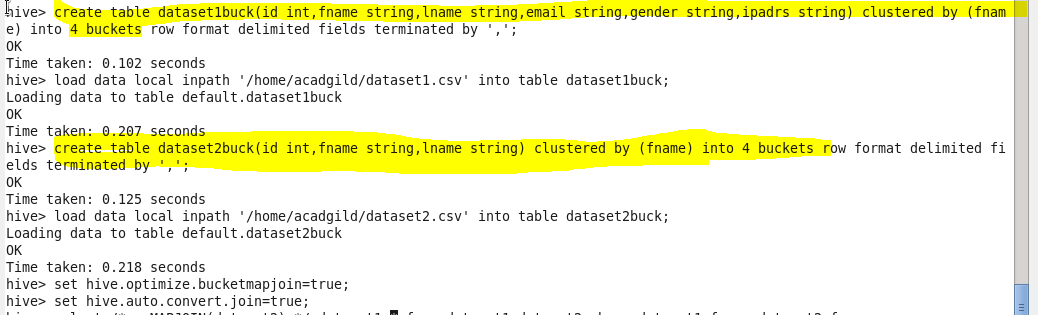
**Dataset description of dataset1.csv is as follows:** id,first\_name,last\_name,email,gender,ip\_address

**Dataset description of tdataset2.csv is as follows:**

id,first\_name,last\_name

**Bucket Map join :**

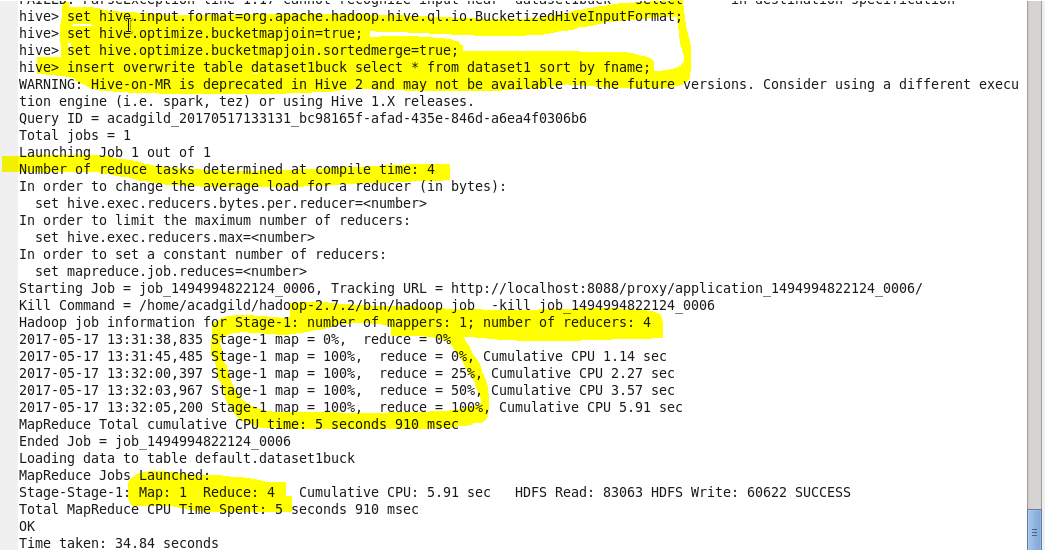
In this , the tables are bucketed and then the content in bucket is joined together .Here there is no need of reducer.

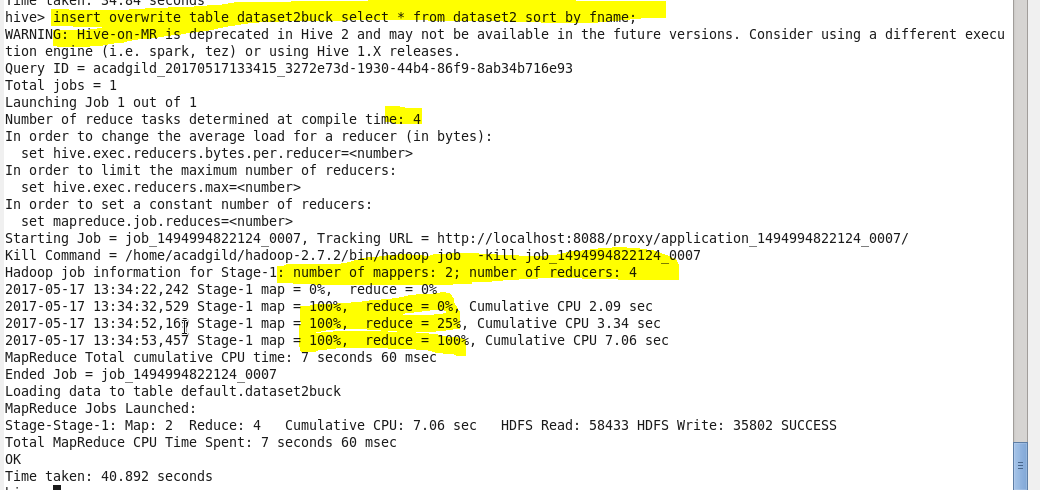
****

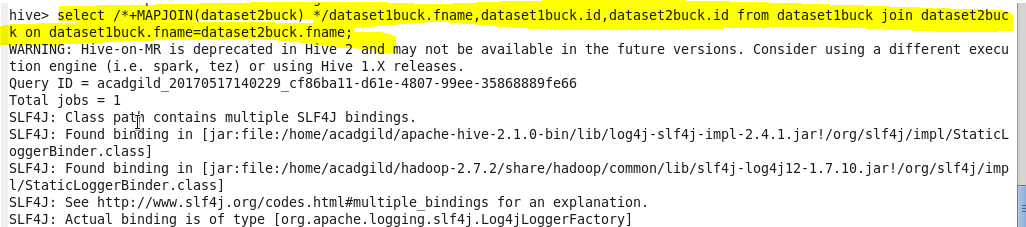


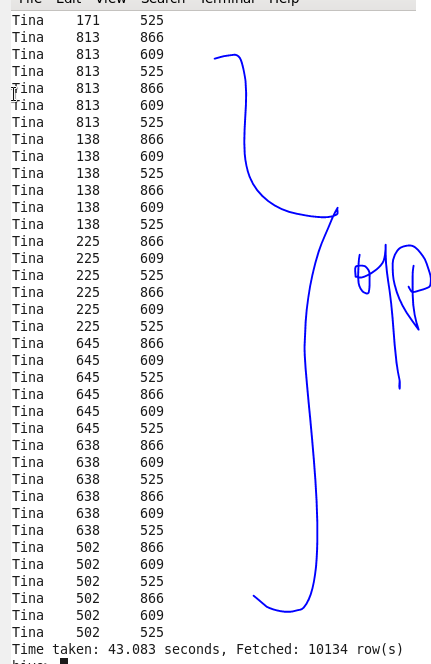
**Sort-Merge Bucket map join**

The main difference b/w smb map join and smb join is the number of buckets is equal in smb map join where as smb join has multiples of buckets of another table.Since equal buckets ,it takes less time than smb join

****

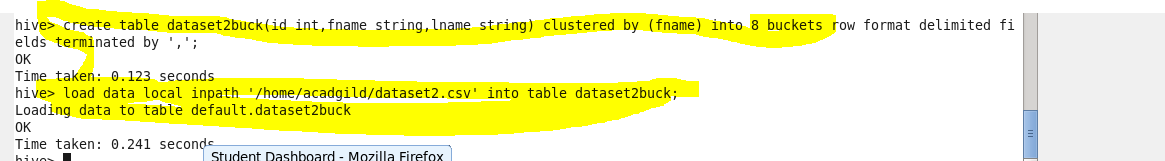
****

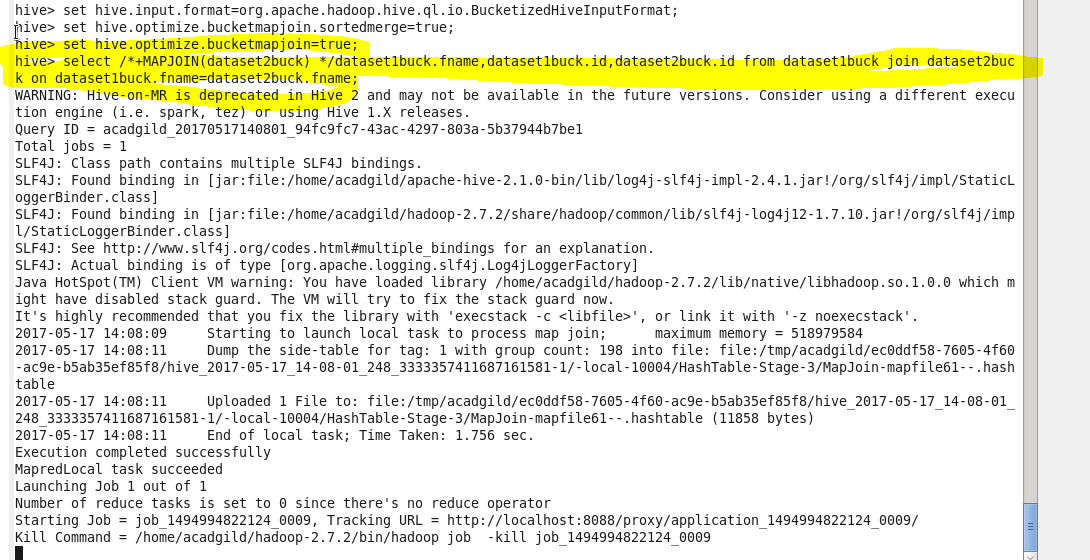


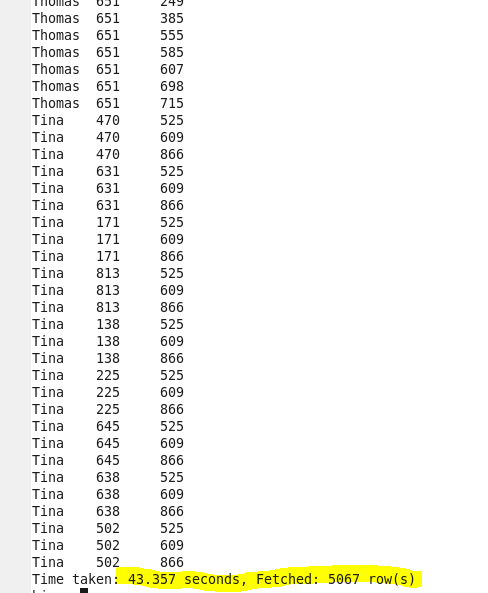


**Sort-Merge Bucket join**

In this first the table that is bucketed is sorted by the column which is on clause condition ,so it is sort merger bucket join.The second table can contain buckets which is a multiples of first table bucket number.







**Left semi join**

Here this join returns the left hand table content when it is present in the right table .Since hive does not perform ‘where’ this helps in joining on conditions .

