**ASS-27.1**

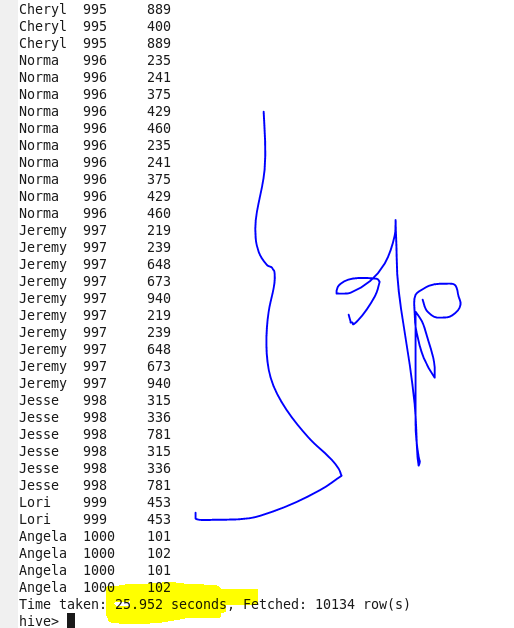
**Explain the following in brief with an example.**

**● Map side Join**

Map side join is used when one of the table is small .It is a join which is performed only by mapper. When we use map side join ,a local task is created which performs a local mapreduce function that creates a hash table and fill it with sorted small table. Now this hash table is sent to the distributed cache where each file is given a separate mapper files which is then copied to the original map reduce task when it is started.

This is suitable only in case one of the table is small and it finishes the task at very short time and less number of reducers.





**● Reduce side Join**

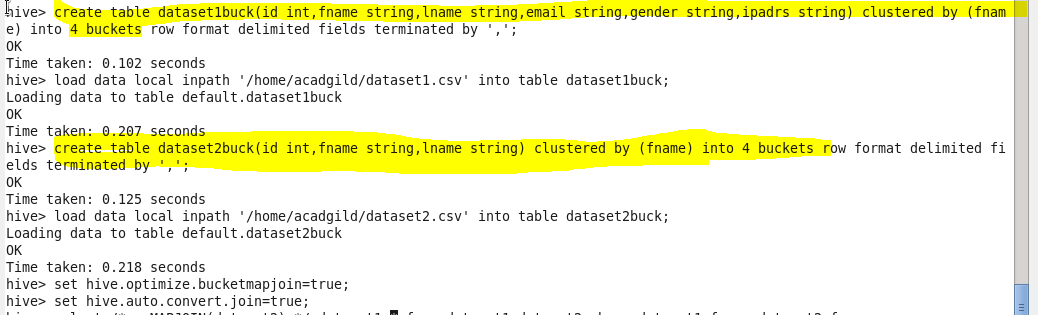
if both the datsets are large then we use reduce side join.This join is performed in reducer side. Reduce-Side joins are more simple than Map-Side joins since the input datasets need not to be structured. But it is less efficient as both datasets have to go through the MapReduce shuffle phase. the records with the same key are brought together in the reducer. We can also use the Secondary Sort technique to control the order of the records.

**● Bucket Map Join**

The data must be bucketed on the keys used in the ON clause and the number of buckets for one table must be a multiple of the number of buckets for the other table.

• When these conditions are met, Hive can join individual buckets between tables in the map phase, because it does not have to fetch the entire content of one table to match against each bucket in the other table.

• set hive.optimize.bucketmapjoin=true; • SET hive.auto.convert.join=true;





**● SMBM Join**

• Sort-Merge-Bucket (SMB) joins can be converted to SMB map joins as well.

• SMB joins are used wherever the tables are sorted and bucketed.

• The join boils down to just merging the already sorted tables, allowing this operation to be faster than an ordinary map-join.

• set hive.enforce.sortmergebucketmapjoin=false;

• set hive.auto.convert.sortmerge.join=true;

• set hive.optimize.bucketmapjoin = true;

• set hive.optimize.bucketmapjoin.sortedmerge = true;

