

1. **Map Side Join-**

Many of the times we face a situation where the hive table is very small but it takes too long to fetch data from that table. In such problems map side join is used. Map side joins takes place at mapper. It allows the table to be stored in memory, so that the join operation becomes faster. In map side join, a local job runs to create hash-table from content of HDFS file and sends it to every node.

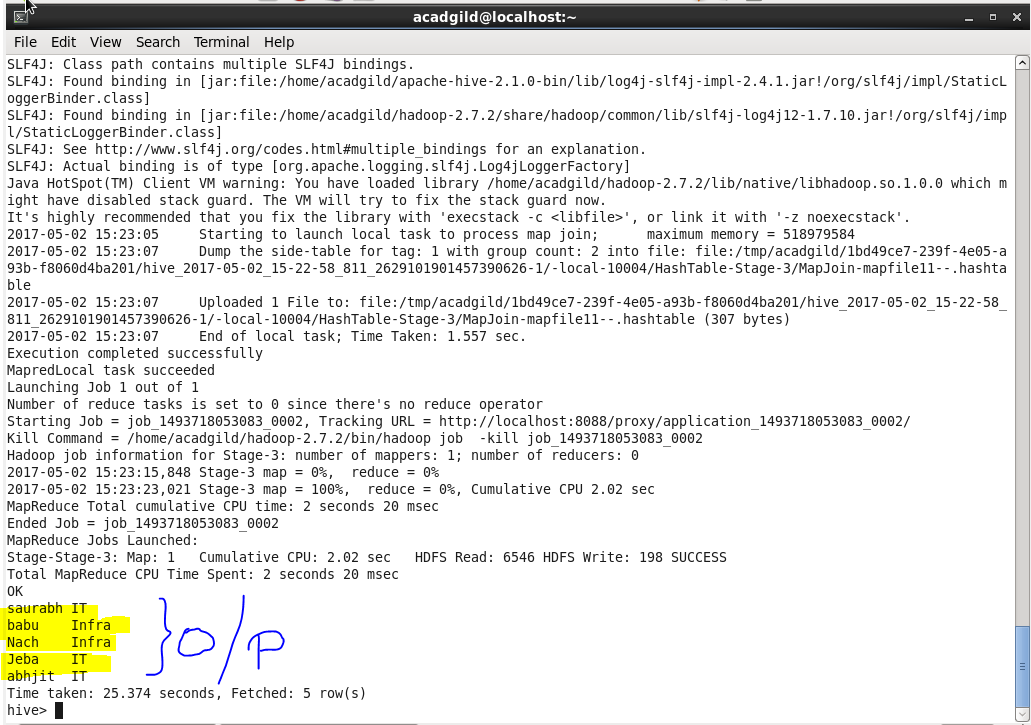
Map Side join can be enabled by setting the following flag-

Set hive.auto.convert.join=true;

**Example-**

****

**Output-**

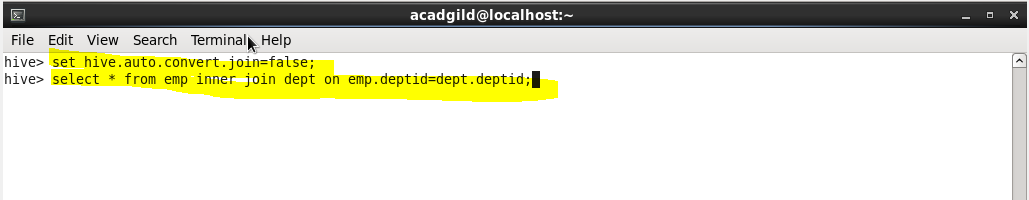
****

1. **Reduce Side Join-**

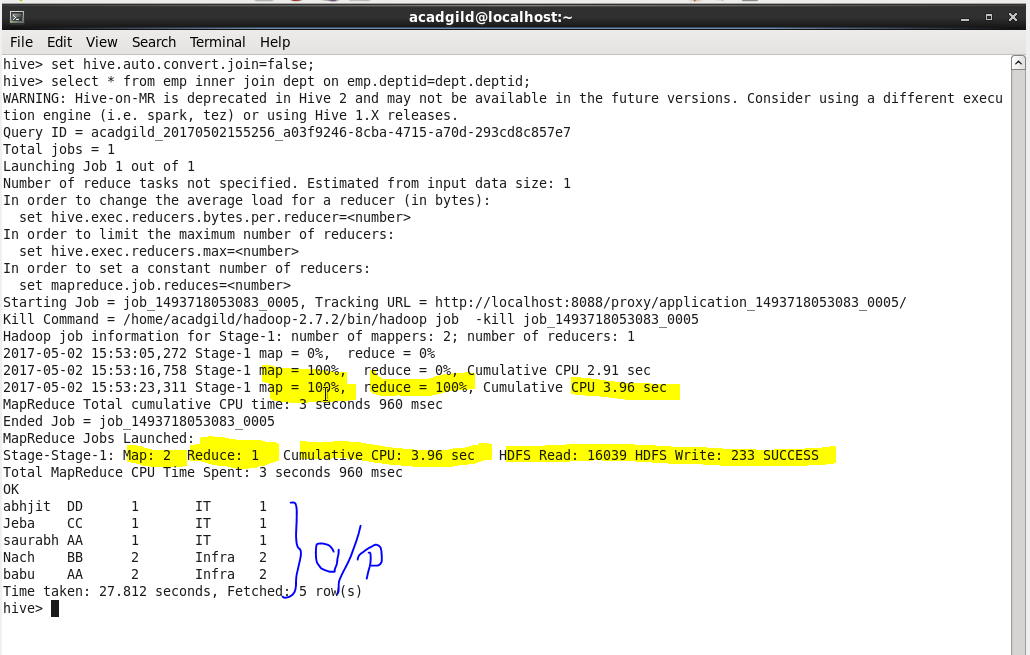
Reduce side joins takes place at reducer. Reduce side joins are much simpler that map side join, as dataset in reduce side join need not to be structured. But it is less efficient as compared to map side join. Reduce side joins take place when dataset is large. For performing Reduce set join following properties we have to set-

Set hive.auto.convert.join=false;

**Example-**

****

**Output-**

****

1. **Bucket Map Join-**

The data storage level is distributed in buckets in this process. Each bucket holds certain rows based on bucketing keys. 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.

Following properties must be set for bucket map join-

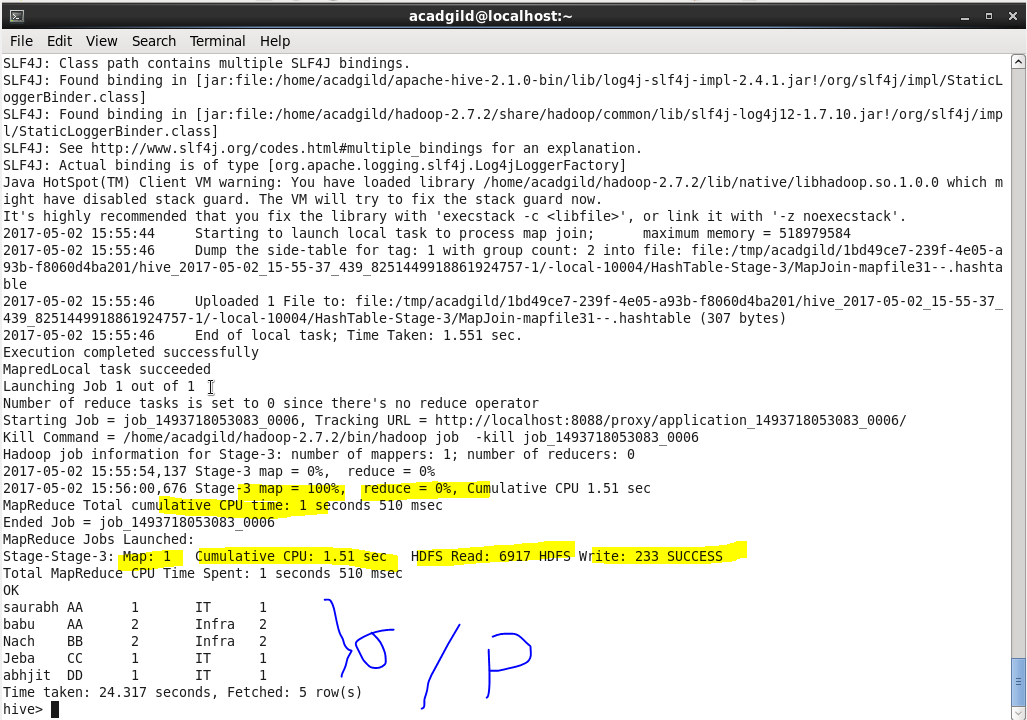
set hive.optimize.bucketmapjoin=true;

SET hive.auto.convert.join=true;

**Example-**

****

**Output-**

****

1. **SMBP Join-**

Sort Merge Join is a join technique, where each mapper reads a bucket from the first table and the corresponding bucket from the second table and then a merge sort join is performed. 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.

Following properties must be set for SMB join-

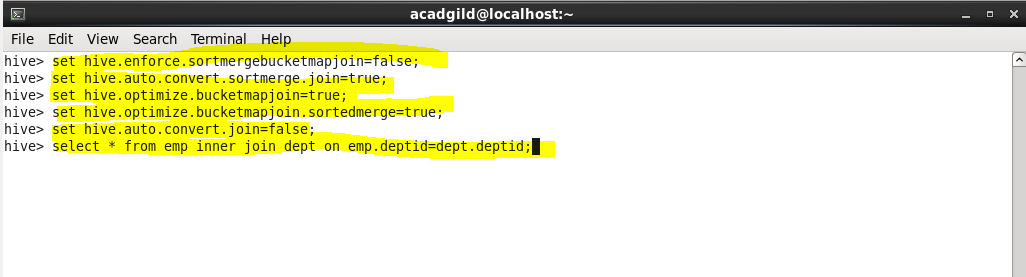
set hive.enforce.sortmergebucketmapjoin=false;

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

set hive.optimize.bucketmapjoin = true;

set hive.optimize.bucketmapjoin.sortedmerge = true;

**Example-**



**Output-**

