**Map-side join:**

For performing Map-side joins, there should be two files, one is of larger size and the other is of smaller size. You can set the small file size by using the following property:

Hive.mapjoin.smalltable.filesize=(default it will be 25MB)

Consider there are two dataset one is of larger in size and other is of smaller in size having id as common column.

Now let us perform map side join by joining the datasets based on their ID’s.

Select /\*+MAPJOIN(dataset2)\*/ dataset1.first\_name,dataset1.eid,dataset2.eid from dataset1 join dataset2 on datset1.first\_name=datset2.first\_name;

As it is amap-side join the number of reducer will be set o 0 automatically.

**Reduce-Side Joins:**  
  
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:**

If tables being joined are bucketed on the join columns, and the number of buckets in one table is a multiple of the number of buckets in the other table, the buckets can be joined with each other.

Consider there are two tables which are bucketed.we can perform bucket-map join between these two datasets.

For performing bucket-map join we need to set this property,

Set hive.optimize.bucketmapjoin=true;

**Sort merge bucket map join:**

you must define the tables to be CLUSTERED BY same column and SORTED BY same column in the same order INTO same amount of buckets.

In SMB join in Hive, 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) join in hive is mainly used as there is no limit on file or partition or table join.

SMB join can best be used when the tables are large. In SMB join the columns are bucketed and sorted using the join columns. All tables should have the same number of buckets in SMB join.

We have to set following options:

Set hive.input.format=org.apache.hadoop.hive.ql.io.BucketizedHiveInputFormat;

Set hive.optimize.bucketmapjoin=true;

Set hive.optimize.bucketmapjoin.sortedmerge=true;