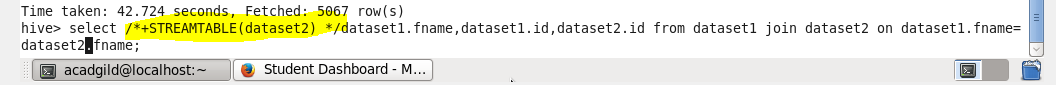
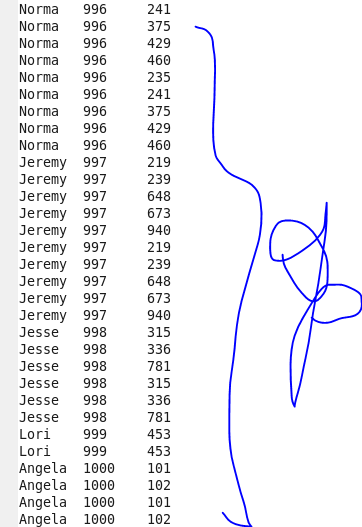
**ASS-27.5**

**Perform join optimizations in hive**

**>**While we are joining two tables, the hive takes the last table for streaming and the first one acts as cache. So we must make sure that the last table should be the large one, so it can process soon.

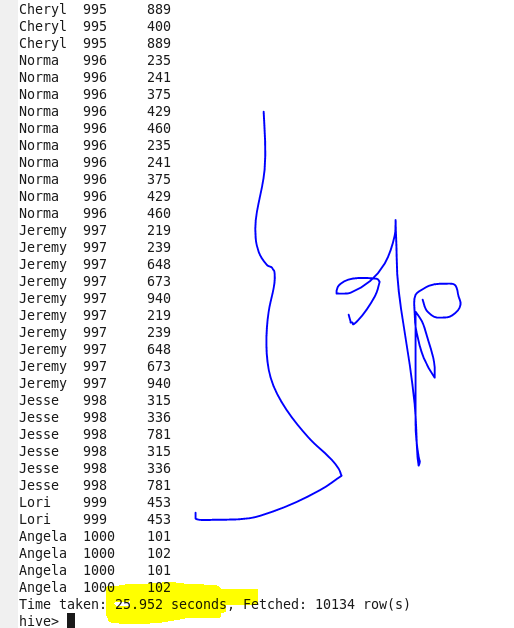
>This we can perform by introducing streamtable





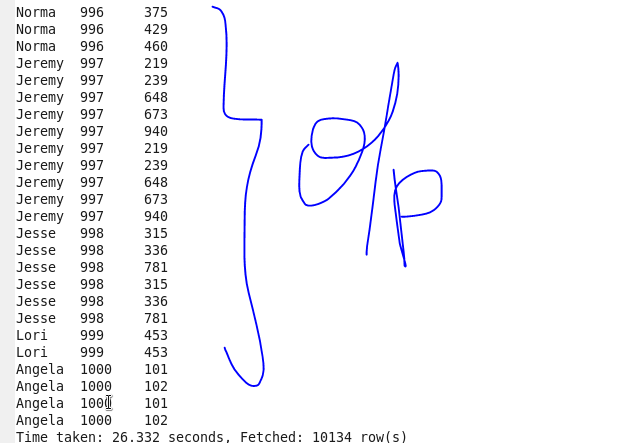
>Map Side Join: Without configuration of map side join if one of the table is very small, it takes a longer time. Map side join is used when one of the table is small, which moves to a temporary hash table the local mapreduce task sort it .Then it is introduced into the distributed cache.Hence map side join takes less time ,since there is no separate reduce part





With Map side 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.

