Run parallel jobs in hive by following the steps in the below blog link

The jobs may have many different stages to get completed.By default, Hive executes these stages one at a time.

Different stages may include a Map stage,Reduce stage, a sampling stage, a merge stage, a limit stage, or other possible tasks Hive needs to do.

A particular job may consist of some stages that are not dependent on each other and could be executed in parallel, possibly allowing the overall job to complete more quickly.

Hive can converts a query into one or more stages and to save time executes multiple jobs parallely.

We can do enable parallel execution of job stages by setting hive.exec.parallel to true .

<property>

<name>hive.exec.parallel</name>

<value>true</value>

<description>Whether to execute jobs in parallel</description>

</property>

Also numbers of mappers assigned to execute parallel processing can also be controlled by following tag.

<property>

<name>hive.exec.parallel.thread.number</name>

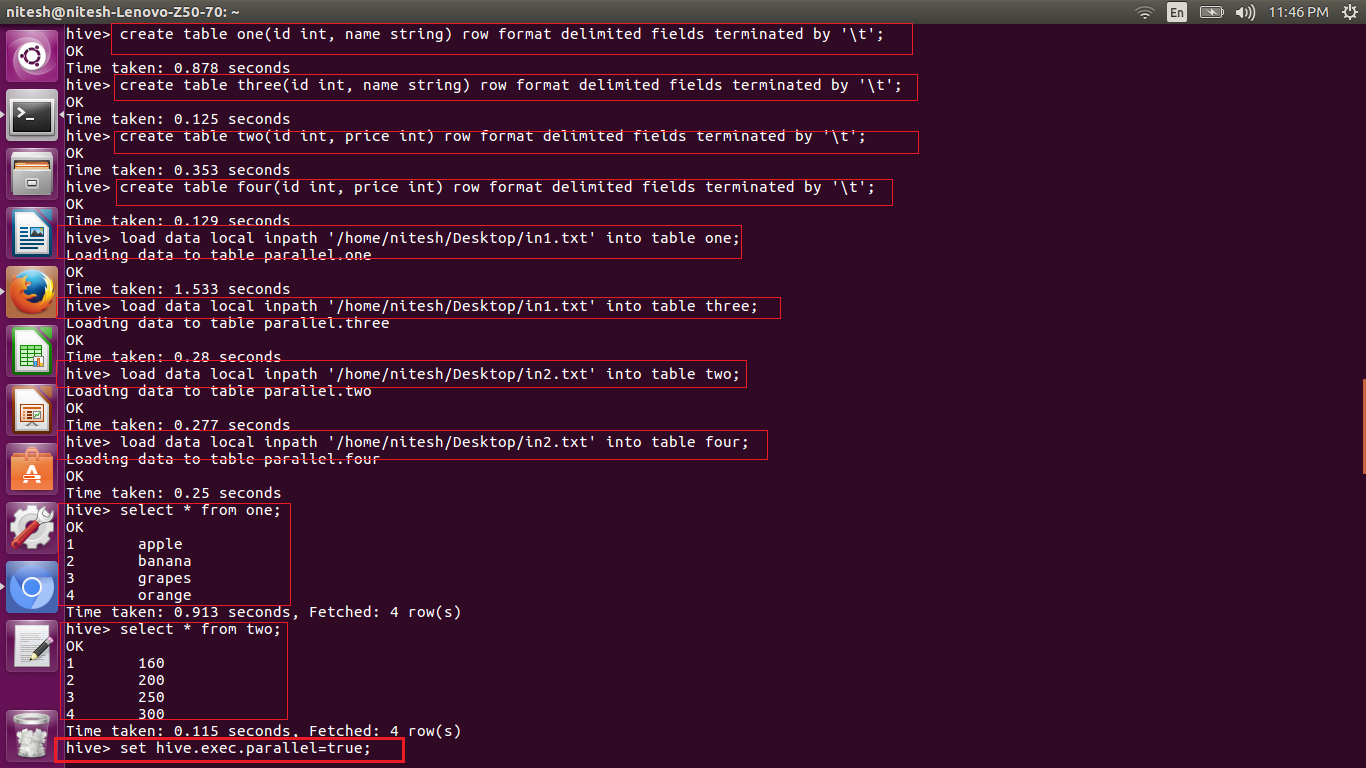
<value>8</value>

<description>How many jobs at most can be executed in parallel</description>

</property>

Or we can use the command set hive.exec.parallel=true;

Here we have created the four tables and data is loaded. As shown below



The property has also been set for the parallel execution.

The query for joins is written and the parallel execution can be seen as blow as the jobs are getting launched. We can see this in the below screen shots.

