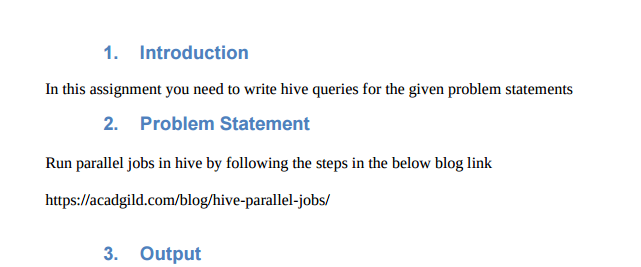
**Assignment 27.4**



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

* **Hive can converts a query into one or more stages and to save time executes multiple jobs parallely.**
* To understand parallel job on hive better first showing non parallel job so that its key concept can be well understood.

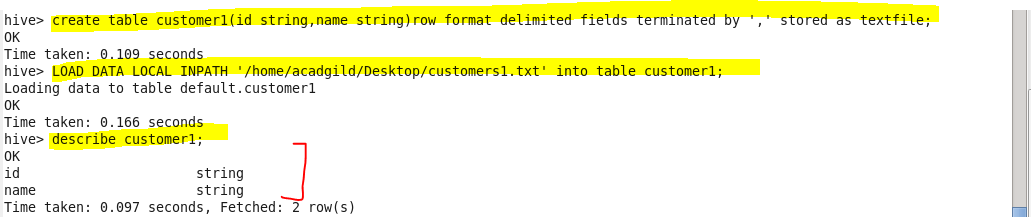
**Non-parallel job**:

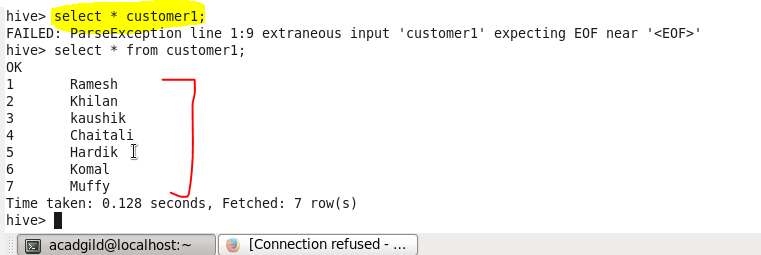
Firstly, creating 4 tables and then loading the data into the created tables.

Then, describing tables to see whether table is created correctly or not and displaying their columns and corresponding dataype.

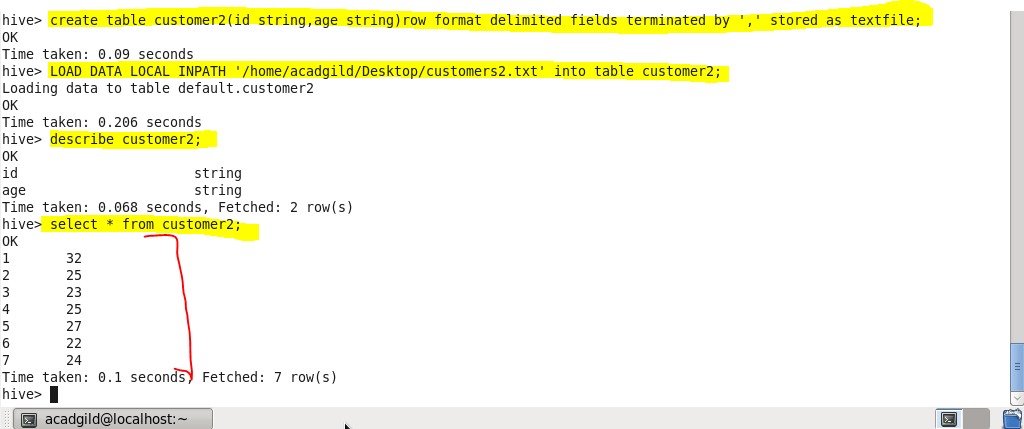
Then, displaying the data of the tables using select keyword.

**Creating table-> customer1**

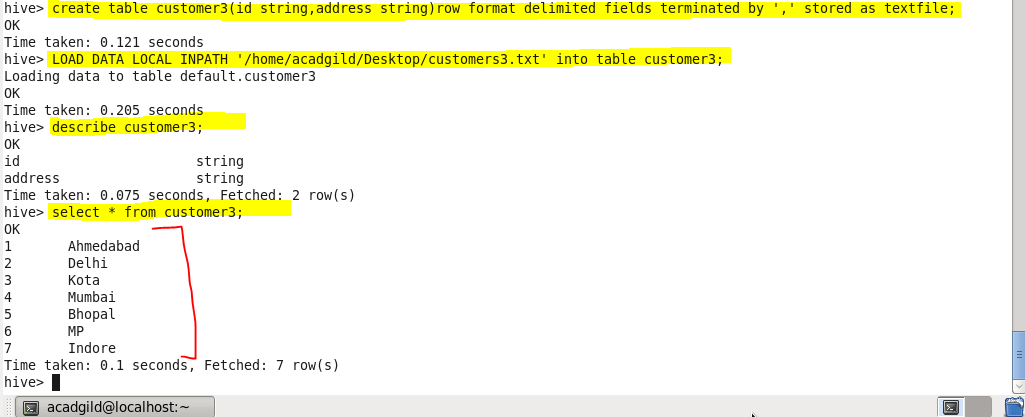
****



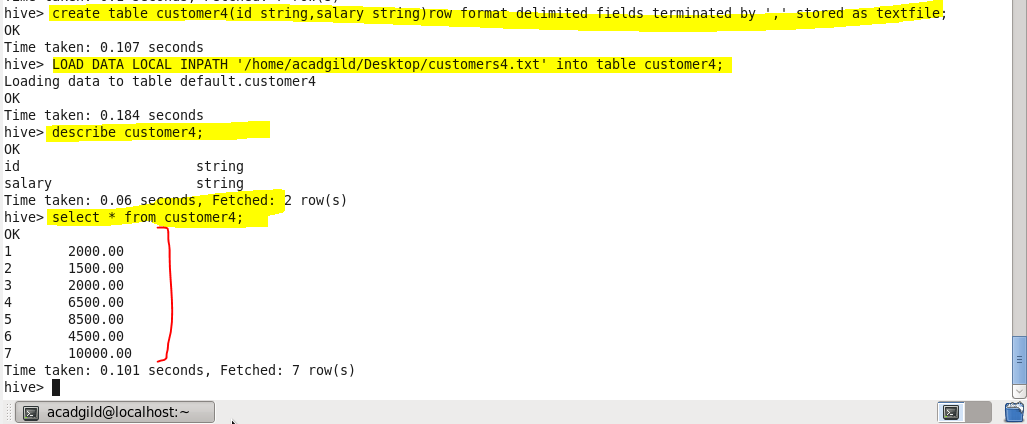
**Creating table –> customer2**



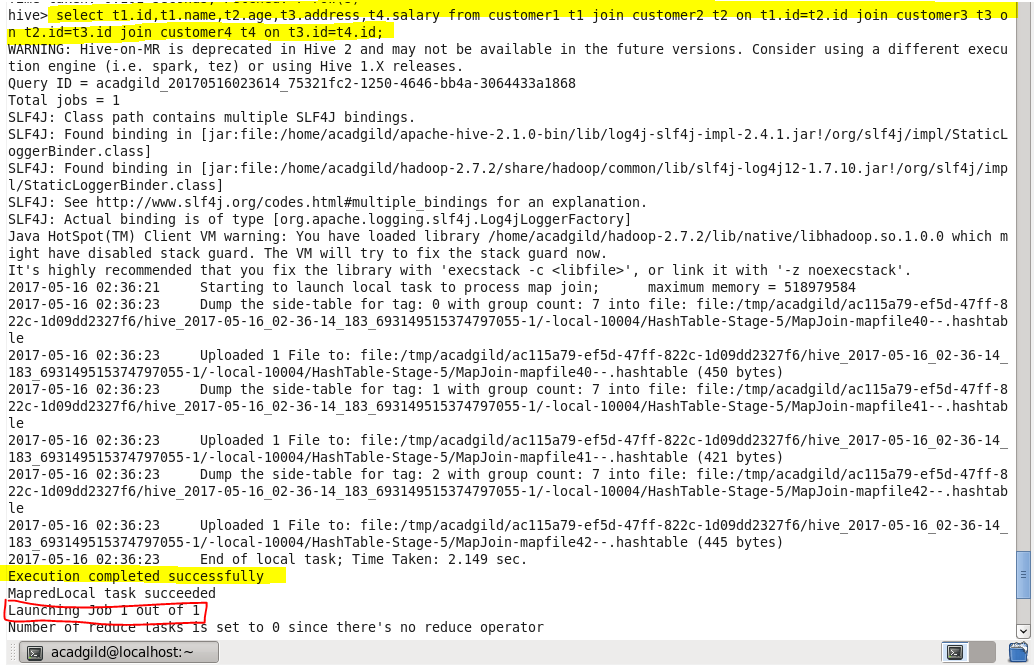
**Created table – customer3**



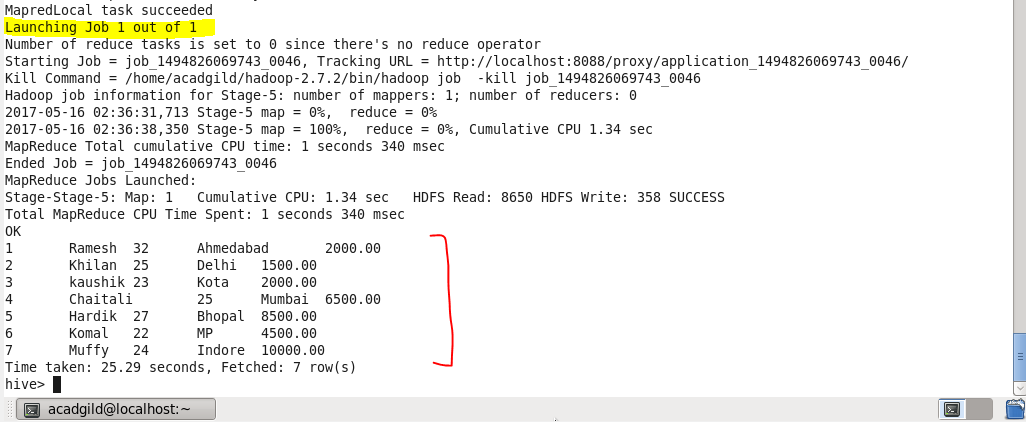
**Created table – customer4**



* **Performing join non parallel**



**Output :**

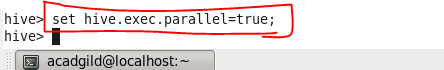


**Performing parallel job :**

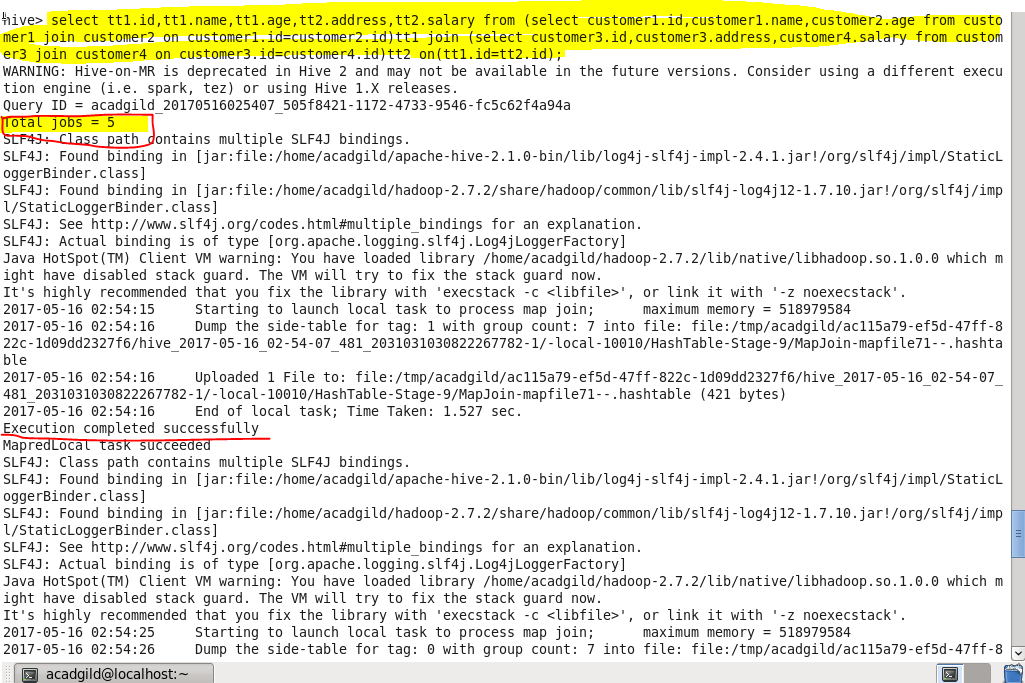
If the query is optimized and more stages are run simultaneously, the job may complete much faster.   
However,If a job is running more stages in parallel, it will increase its cluster utilization.

**Before running it we need to set configuration:**

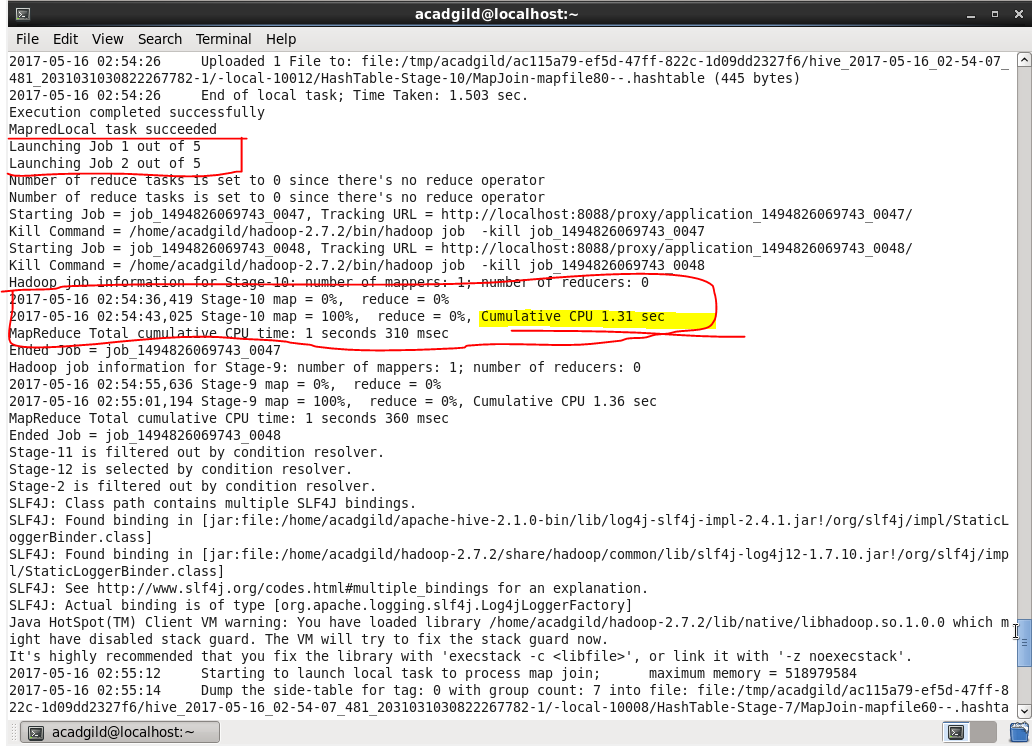
* **Set hive.exec.parallel = true;**



**Now,running parallel job -**

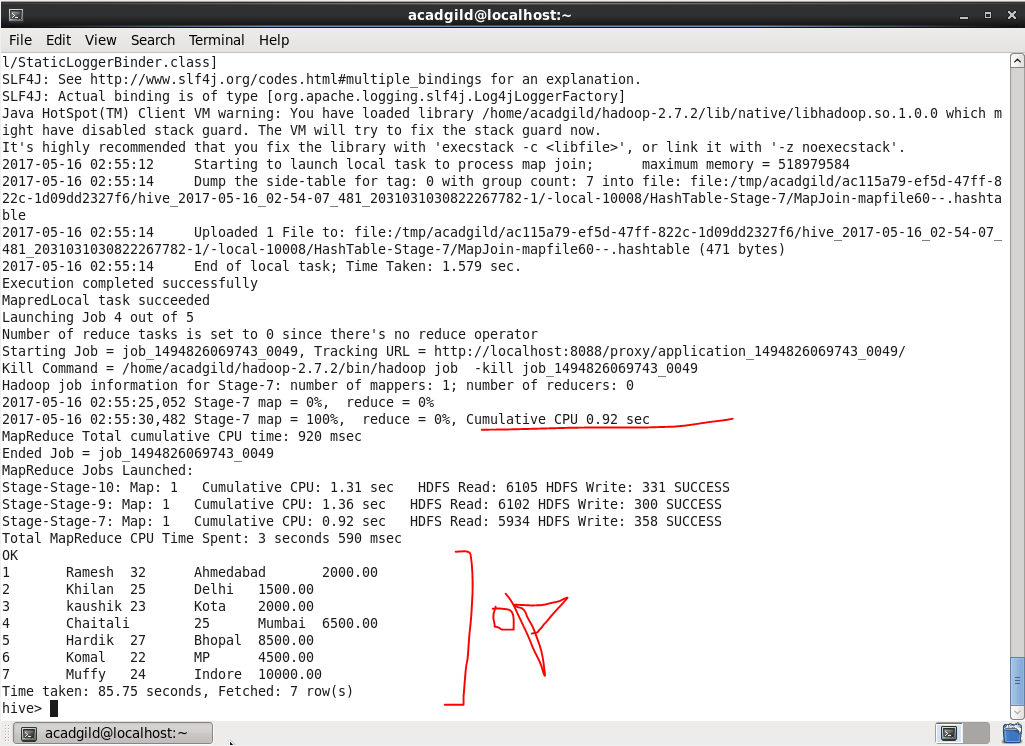


**We can see that job are executed in parallel and staging is also done in parallel as shown which will decrease the execution time**



**after job 1 is completed Job2 runs and so on**

**Output :**



* **We got the same result from parallel and non parallel jobs.**