Advanced Database System Lab
Assignment 6
Parallel Databases

Introduction to Parallel Databases

Companies need to handle huge amounts of data with high data transfer rates. The client server and centralized system is not very efficient. The need to improve the efficiency gave birth to the concept of Parallel Databases.

Parallel database system improves performance of data processing using multiple resources in parallel, like multiple CPU and disks are used parallely.

It also performs many parallelization operations like, data loading and query processing.

Goals of Parallel Databases

The concept of Parallel Database was built with a goal to:

Improve performance:
The performance of the system can be improved by connecting multiple CPU and disks in parallel. Many small processors can also be connected in parallel.

Improve availability of data:
Data can be copied to multiple locations to improve the availability of data.
For example: if a module contains a relation (table in database) which is unavailable then it is important to make it available from another module.

Improve reliability:
Reliability of the system is improved with completeness, accuracy and availability of data.

Provide distributed access of data:
Companies having many branches in multiple cities can access data with the help of a parallel database system.

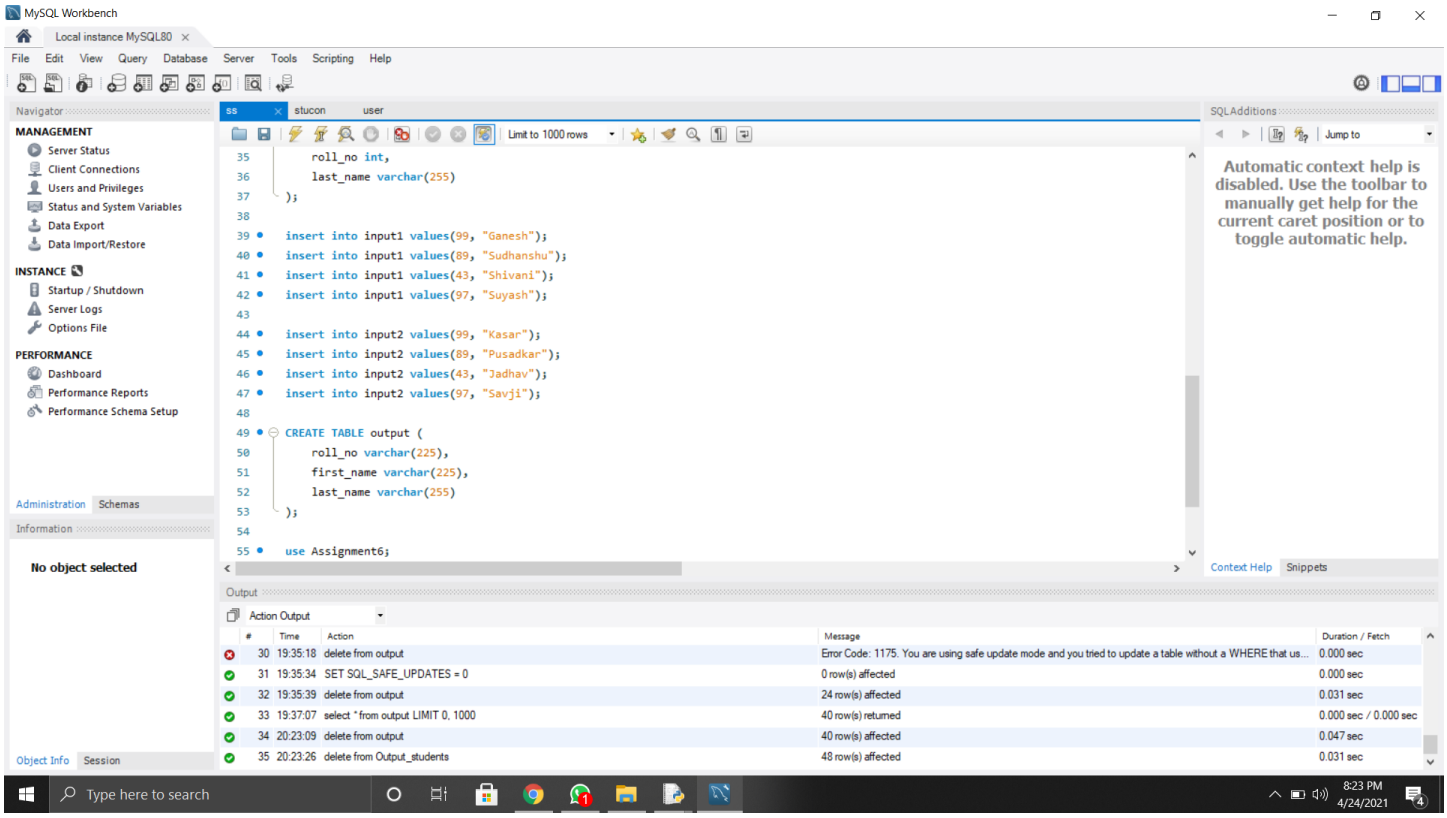
Steps: 1. Created the Tables

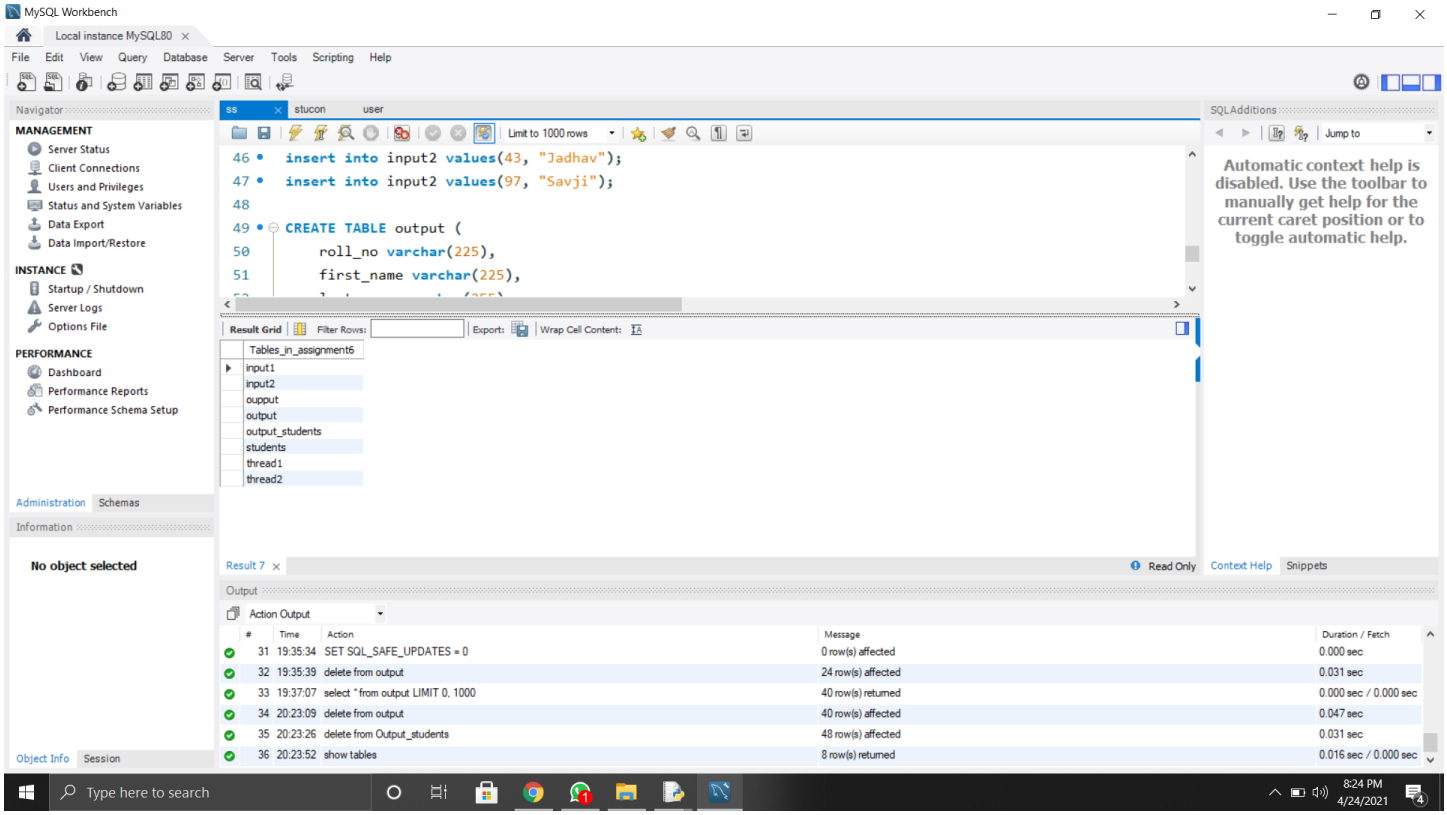
ParallelSort():

1. Students
2. Output_students

ParallelJoin():

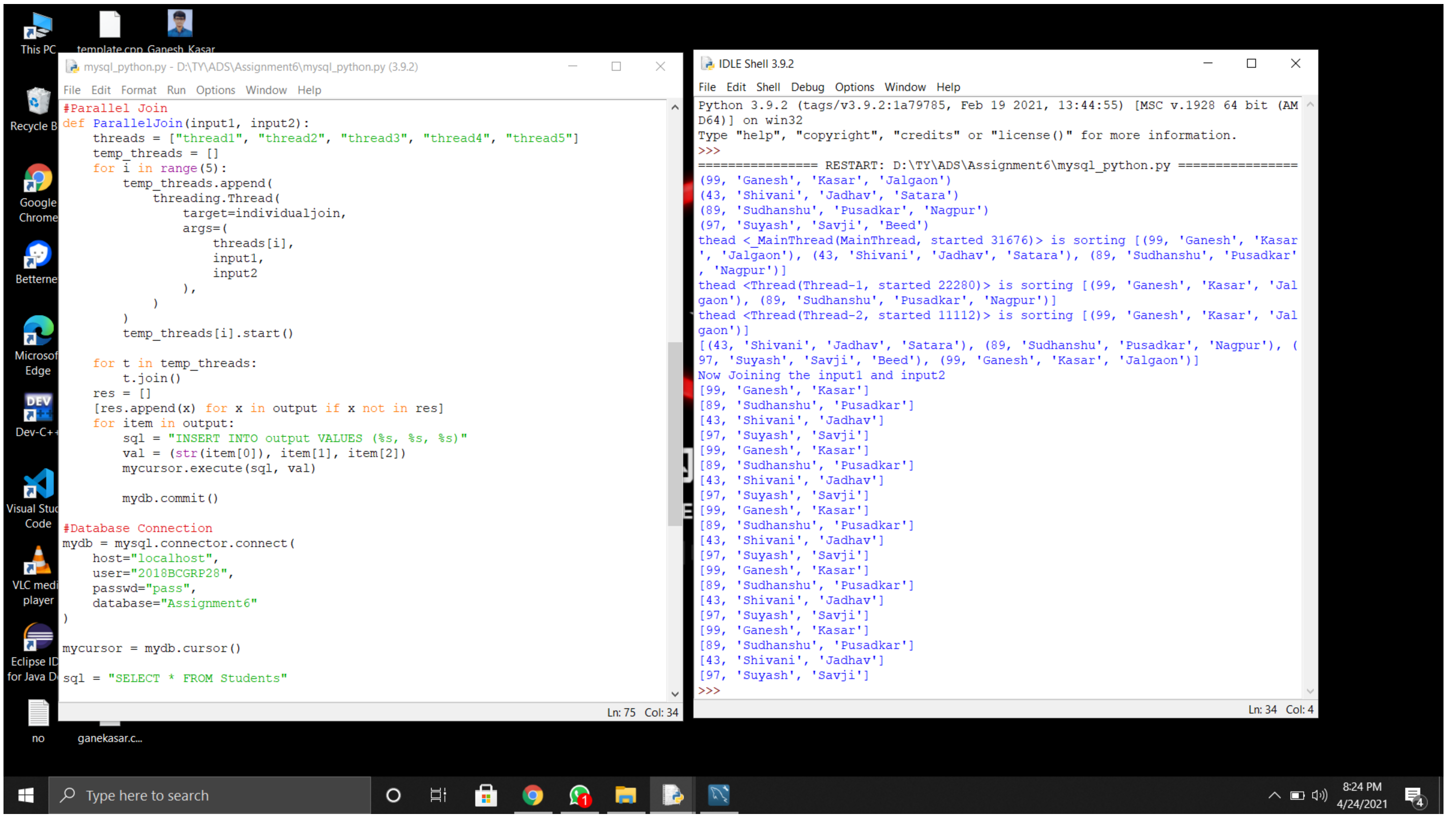
1. Input1
2. Input2
3. Output



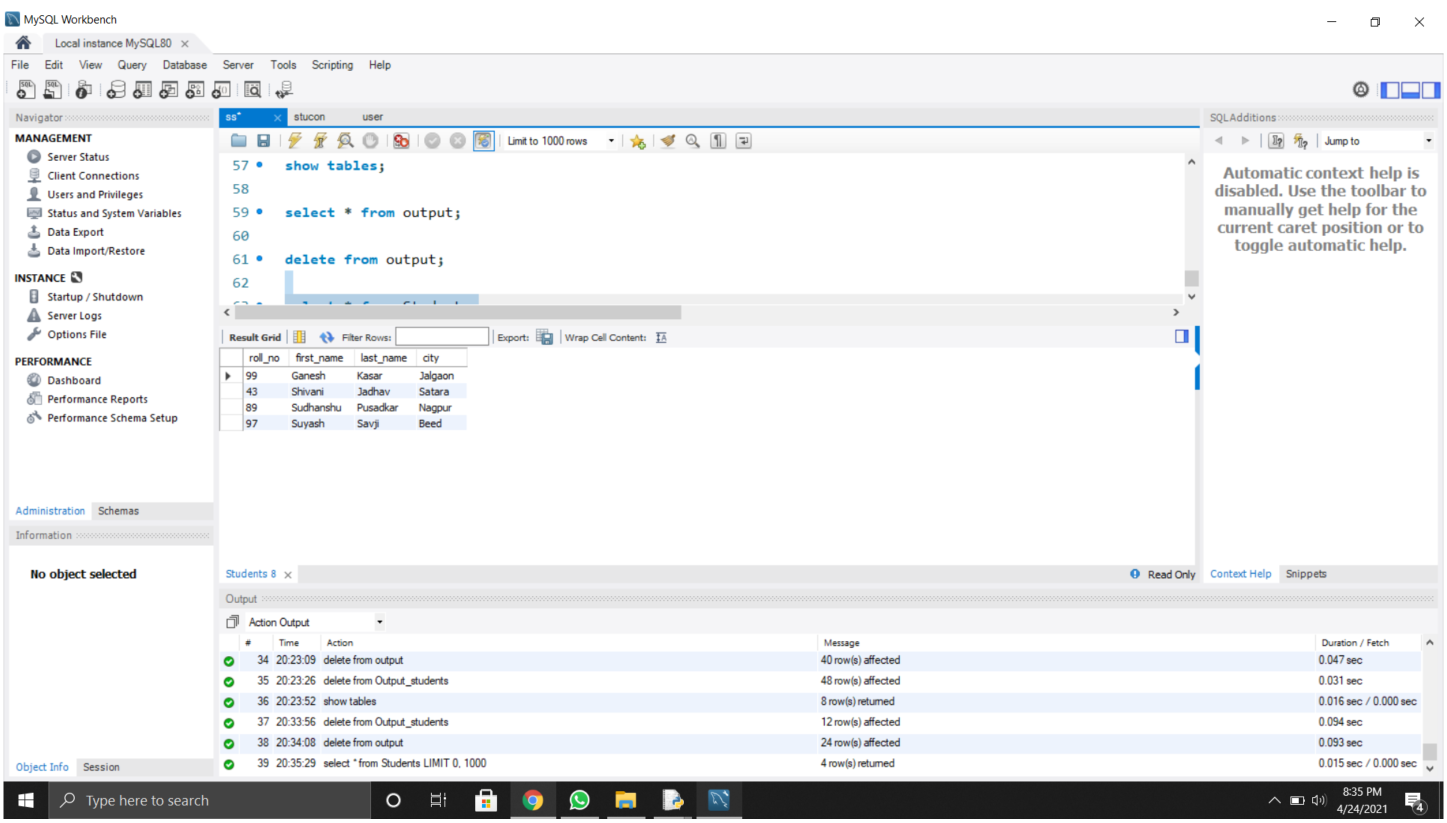


Step 2: Created Python Program

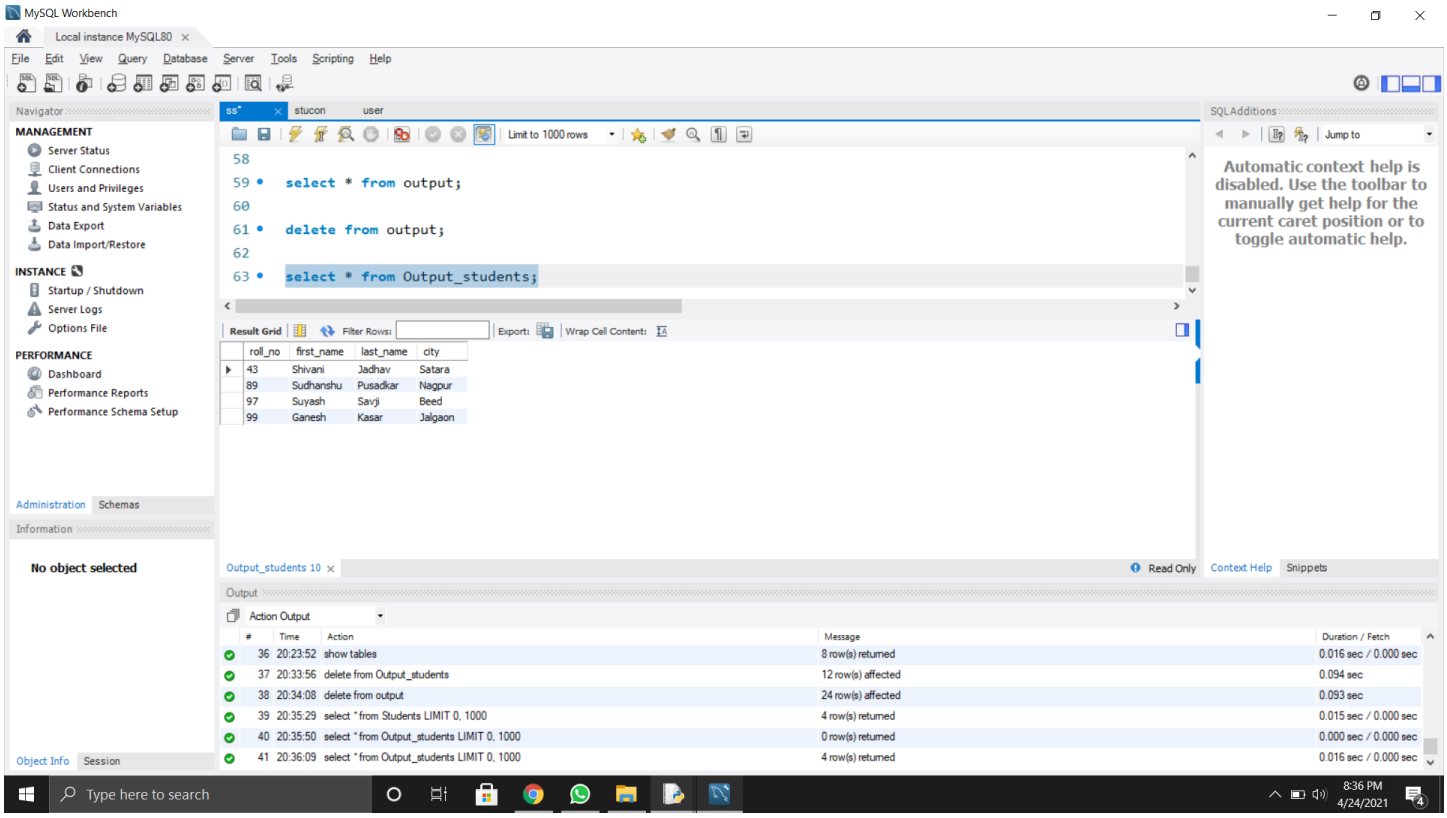
1. Created Connection to MySQL using MySQL connector
2. Sorted the Students table using threads
3. Joined the input1 & input2 stored in new table output



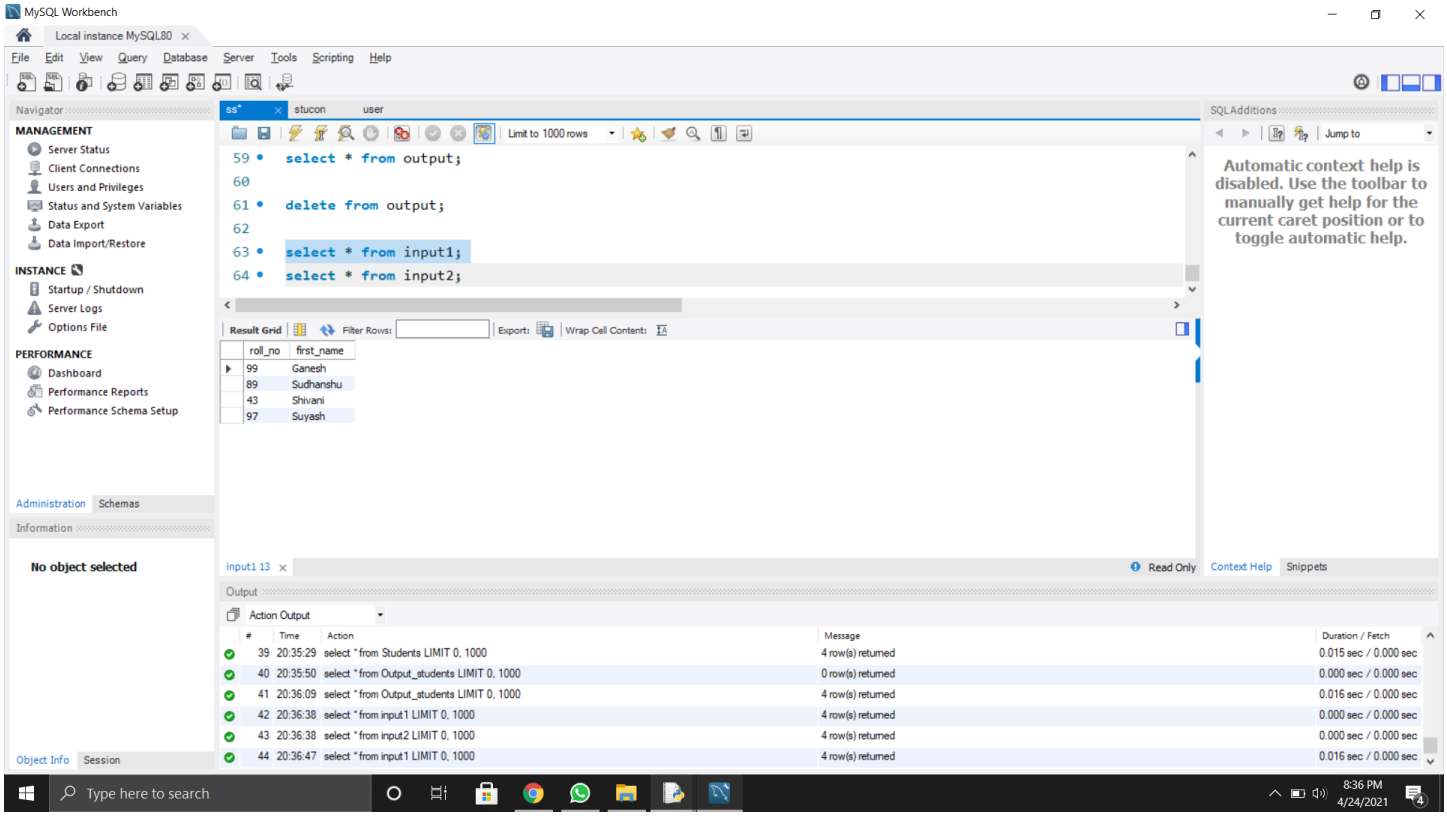
1. Before Sorting



2. After Sorting



1. Input Tables



2. Output tables