# DBMS LAB (CS355)

Lab Assignment 10
Indian Institute of Technology, Patna
October 26, 2022
Submission Deadline: **6:00 p.m. 26/10/2022**

--------------------------------------------------------------------------------------------------- -

Instructions:

1. Marks will be deducted in the case of plagiarism.

2. Appropriate comments (if necessary) are mandatory.

3. Submission File Name: ***roll_no_Assign10.txt, e.g.:1501cs11_Assign10.txt.***

4. Upload your assignment ( the text file ) at the following link:

https://www.dropbox.com/request/Qfo7RO0JfypTxDe3T7N7

The basic objective of this assignment is to get familiar with the index utility of MySQL. In database, an index is a data structure that improves the speed of operations in a table. Indices can be created on selected column(s) to facilitate fast search. Without index, a SELECT * FROM products WHERE productID = x needs to match with the productID column of all the records in the products table. If productID column is indexed, then the matching performance is expected to improve, especially when a large number of records to be processed. You should index columns which are frequently used in the WHERE clause; and as JOIN columns. In this assignment, you will try to monitor the performance improvement using indexing in MySQL.

1) **Task 1**: For this assignment, download One sample **transaction.csv** is available in Gmail. Along with this you can also experiment with a larger file and mention the downloadable link of the file in the submission file. (You can choose from the internet)

**1.1 SubTask A: Define the table and load data into it from the CSV file:** Go through the CSV file, check the properties of the various columns, and define a table with appropriate attributes. Don't define any constraints initially. Just define appropriate

data types so that you can load the values (from the CSV file) into the table. Then load all data from the CSV file into the table.

**1.2 SubTask B: Check various queries and monitor the timings:** Now, write various queries using **SELECT, INSERT, UPDATE** statements and monitor the timing. You execute the same query multiple times (in difierent sessions) and take the average. While writing the SELECT queries use relevant **where** clause to check both range and equality type of queries. Also, create some related tables and check the join operations as well.

**1.3 SubTask C: Create index on the copied table**: Now, create a new table and copy all the data from the earlier table. For this purpose, you can use the followings-

**CREATE TABLE new_table LIKE old_table;**

**INSERT new_table SELECT * FROM old_table;**

Choose an appropriate column(s) as an index and create index on it.

**1.4 SubTask D: Check various queries and monitor the timings:** Now, write the same queries that you have used in SubTask B on the indexed table and note down the average timing.

**1.5 SubTask E: Perform Comparative Analysis**: Perform a comparative analysis on the timings that you have obtained using SubTask B and SubTask D. State your observation.

**Note:** You need to submit a detailed report describing the steps (including all the MYSQL commands) you have used for performing Task A, B, C and D. Also, provide the details of the CSV file that you are using for the experiment purpose. Provide the description of the attributes and number of rows the le has. Also, provide justications

for index attribute selection, query selection, etc.

**Some Helpful tutorials**

1. https://dev.mysql.com/doc/refman/8.0/en/mysql-indexes.html#:
~:text=Indexes%20are%20used%20to%20find,table%2C%20the%20more%
20this%20costs.

2. https://www.mysqltutorial.org/mysql-index/mysql-use-index/