CSE441: DATABASE SYSTEMS

ASSIGNMENT 1 (Mini-SQL Engine)

DEADLINE: 9:00 pm, 22nd January 2020

In this assignment, you are supposed to write a minisql engine which will run a subset of SQL Queries using command line interface

Programming Languages Allowed: Python and C++

Dataset:

1. CSV files for tables.
    a. If a file is : File1.csv , the table name would be File1.
    b. There will be no tab separation or space separation, so you are not required t e handle it but you have to make sure to take care of both csv file type cases: the one where values are in double quotes and the one where values are without quotes.
2. All the elements in files would be only **INTEGERS**.
3. A file named: metadata.txt (note the extension) would be given to you which will have the following structure for each table:
   <begin_table>
   <table_name>
   <attribute1>
   ....
   <attributeN>
   <end_table>

Types of Queries:

1. Select all records:
   Example Query: Select * from table_name;

2. Aggregate functions: Simple aggregate functions on a single column namely - **Sum, average, max and min**.
   Example Query: Select max(col1) from table1;

3. Project Columns (could be any number of columns) from one or more tables:
   Example Query: Select col1, col2 from table_name;

4. Select/project with distinct (for only one attribute) from one table.
   Example Query: Select distinct(col1) from table_name;

5. Select with where from one or more tables :
    a. In the where queries, there would be a maximum of one AND/OR operator with no NOT operators.
    b. Relational operators that are to be handled in the assignment, the operators include "<, >, <=, >=, =".

   Example Query: Select col1,col2 from table1,table2 where col1=10 AND col2=20;

6. Projection of one or more(including all the columns) from two tables with one join condition :
    a. NO REPETITION OF COLUMNS – THE JOINING COLUMN SHOULD BE PRINTED ONLY ONCE.

   Example Query1: Select * from table1, table2 where table1.col1=table2.col2;
   Example Query2: Select col1,col2 from table1,table2 where table1.col1=table2.col2 AND table2.col2>=10.

Input Format:

1. For C++ it will be – ./a.out “select * from table_name where condition”
2. For Python, it will be – python RollNumber.py “select * from table_name where condition”

Output format:

<Table1.column1, Table1.column2....TableN.columnM>
Row1
Row2
.....
RowN

Parser: You can use prebuilt parsers for SQL queries (Example: *sqlparse* in python)

Note: ERROR HANDLING: 10% of marks will be for error handling.

Deliverables

1. Python/C++ Source Code files.
2. Keep all the files in a folder RollNumber_Assignment1, zip it and upload.

Strict action will be taken for copying in the Assignments.