CSE 441: DATABASE SYSTEMS ASSIGNMENT 1

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

Programming Languages Allowed : Python, Java and C/C++

Dataset:

1. csv files for tables.
    i. If a file is named *File1.csv,* the table name would be *File1*.
    ii. There will be no tab-separation or space-separation, so you are not required to 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 **integers** only.
3. A file named *metadata.txt* (note the extension) will be given to you, which will have the following structure for each table:

    <begin_table>

    <table_name>

    <attribute1>

    ....

    <attributeN>

    <end_table>

Type of Queries:

Your code will be expected to run the following set of queries.

1. **Select all records** : `select * from table_name;`
2. **Aggregate functions**: Simple aggregate functions on a single column. Sum, average, max and min. They will be very trivial given that the data is only numbers: `select max(col1) from table1;`
3. **Project Columns**(could be any number of columns) from one or more tables : `select col1, col2 from table_name;`
4. **Select/project with distinct from one table** : `select distinct col1, col2 from table_name;`
5. **Select with where from one or more tables**: `select col1,col2 from table1,table2 where col1 = 10 AND col2 = 20;`
    a. In the where queries, there would be a maximum of one AND/OR operator with no NOT operators. (Bonus marks will be awarded if you handle cases with multiple AND/OR operators)
6. **Projection of one or more(including all the columns) from two tables with one join condition** :
    a. `select * from table1, table2 where table1.col1=table2.col2;`
    b. `select col1,col2 from table1,table2 where table1.col1 = table2.col2;`
    c. The joining column should be printed only once*. No repetitions.*(Bonus marks will be awarded if you handle cases with multiple tables join)

7. **Important:**
    a. ERROR HANDLING: 10% marks will be for error handling.

b. For the above queries, please note all the permutations and combinations of SQL that MySQL permits, especially when it comes to multiple tables. What is mentioned above are examples of what the queries could be.

8. **Parser**: You can use prebuilt SQL query parsers.
9. **Format of Input**
    a. Command to execute a command would be of the form: {compiled file} “SQL Query”. Here, the SQL Query would be a command line argument. Example :
        i. C++ – ./a.out “select * from table_name where condition”
        ii. Java – java classfile.class “select * from table_name where condition”
        iii. Python – python engine.py “select * from table_name where condition”
10. **Output Format:**

    <Table1.column1, Table1.column2....TableN.columnM>

    Row1

    Row2

    .....

    RowN

Deliverables:

a. Python/Java/C++ Source Code files.
b. Compiled Java/C++ files, if done in Java/C++.
c. Bash script with your roll number as the filename. (2015xxxx.sh, for example). The bash script should take

the SQL query as a command-line argument and run the program on it. For example: bash 2015xxxx.sh "select * from A" .

d. Keep all of this in a folder Roll-Number and *zip* it. Only zips will be considered, not tar/gz files.
e. Upload *only the zip file* on moodle.
f. **Copying in the Assignments can lead to anything from a Zero in the current assignment to an F in the course.**

DEADLINE: Jan 30, 2018 9:00PM