

Project 2

CS 5300

Due date: November 30 (Firm Deadline)

You are extending the scope of project1 by adding an optimizer to your SQL compiler. The compiler will read (scan) a set of SQL queries as input and perform the following steps for each query:

- 1) The input query is scanned to identify different entities, operations, and order of execution.
- 2) The input query is then translated into its equivalent relational algebra.
- 3) The generated relational expression is then mapped into a query tree.
- 4) Your optimizer is called and applies various optimization rules to develop an efficient execution plan.
- 5) You will write a report about your project. Your report must be self- explanatory, clearly explaining the structure of your optimizer, the operation of the scanner, ..., and a justification of its correctness. As part of your report, for each input query (test data), you are expected to print out the relational algebra expression, the initial query tree, and the query tree after each optimization step (note, you need to identify the optimization rules that have been applied).

Please make note of the following:

- 1) The test data will be made available on the course web site on November 16.
- 2) The due date of the project is firm and will not change under any circumstances.
- 3) Each group is expected to submit in class a hard copy of the report along, with a print-out of the compiler. The compiler is expected to be well- documented.
- 4) SQL queries could be nested, consisting of set and relational operators.
- 5) Please note that randomly-selected groups may be asked to demo their project and execute it on input data.