

Minh Nguyen

Programming Assignment 4 Report

Simple DB

Professor George Kollios

I. Design decisions:

- IntHistogram.java

With the help of the instruction (2.2.3 Filter Selectivity), I was able to calculate the size of the current histogram. And with that size I was able to get the index for the addValue and estimateSelectivity function.

The addValue function was simple once I was able to get the index, increasing the histogram[index] by 1.

With the help of the instruction (2.2.3), I solved EstimateSelectivity by dividing the problems into 2 part:

- + Solving for equal (op is EQUALS or NOT EQUALS) (1)
- + Solving for Greater than, less than, greater than or equal to, less than or equal to. (2). This one was then divided one more time into 2 big cases, one for greater than and one for less than.
- TablesStats.java: This class was given fully implemented and passed all tests.
- JoinOptimizer.java:
 - + estimateJoinCost: Section (2.2.2) in the instruction gives the formula for this function.
 - + EstimateJoinCardinality: (2.2.4) in the instruction guided through this function. For no primary key I used the suggestion as in the instruction, returning the larger size of the two tables. For either that has the primary key, I simply return the smaller size of the two tables, so the nTuples can't get larger than either one.
 - + For range scan, I return the cross product of the two tables, which means the ratio I set is 1.
 - + OrderJoin.java: I strictly followed the pseudo code given in the instructions (2.3).

II. Changes to the API:

None

III. Incompleteness:

None

IV. Feedback:

I'm very grateful for the detailed instructions that are given. I can see myself sitting on this assignment for at least 20 hours without the instructions.

So, with the instructions, and the given TableStats.java, which should have taken me at least 5 to 8 hours in addition but luckily did not, I spent around 10 hours on this assignment, and most of the hours I spent on debugging EstimateSelectivity and OrderJoin.

Most of the bugs I had with IntHistogram were initializing values and when to increment them, and handling all the possible cases.

Bugs with JoinOptimizer were from the loops. I set the looping range incorrectly, increment the variables incorrectly, like values that should not be changed and should be changed in this loops but not outside the other loop,...etc

Overall, with the given TableStats and all the instructions, this assignment was shorter than the other ones, not easier by itself, but easier with the instructions. Thank you once again for it.