

Lab4 Report

Jiangbei Li

- Design Decision

- IntHistogram
 - In the construction function, I calculate the width of each bucket and create an array to save the histogram and a variable to save the number of tuples.
 - In the function addValue, I get the bucket that the variable belongs to and add it to corresponding position in the array.
 - In function estimateSelectivity, I calculate the selectivity according to the type of operation.
- TableStats
 - In the construction function, I create three arrays to save the max value, min value and the histogram for each field. Firstly, traverse the tuples in the table and save the max and min value for each field. Next, create IntHistogram or StringHistogram for each field using the max and min value. Finally, traverse the table again and add all values to the histograms.
 - In function estimateSelectivity, estimateScanCost and estimateTableCardinality, I just call the estimateSelectivity function in class histogram and do some simple calculations.
- Join cost estimation
 - In function estimateJoinCost, just use simple calculation.
 - In function estimateTableJoinCardinality, I need discussions on some cases. If only one of tables is the primary-key table, the cardinality of the other table will return. If both the tables is the primary-key table or not the primary-key table, the larger will return. Moreover, if the operation is not equal, $\text{card1} * \text{card2}$ - calculated card will return. For range scans, just return $0.3 * \text{card1} * \text{card2}$.
- Join Ordering
 - Firstly, save the number of joins and create a new PlanCache.
 - Using the method of dynamic programming, we just deal with the subset of the join set whose size from 1 to the number of joins.

Friday, May 31, 2019

- For each subset, I traverse all the joins in the subset and treat each join as the last join. I can get the best plan for the subset and the CostCard. Then add it to the PlanCache.
- Finally, I just need to return the best plan in the PlanCache for the whole join set.

- Difficulty and time

- I spend two days on this lab
- The most difficulty is how to use the class PlanCache, Scanseq and CostCard and how to use the private function enumerateSubsets, computeCostAndCardOfSubplan and so on.