

CSE 444 TA Section 4/26

Lab4 has released

Debug using Eclipse

New version

- Parser.java
- JoinOptimizer.java
- LogicalJoinNode.java

HashEquiJoin

- A hash join implementation, only able to do equi-join
- How it works.
- Estimate cost.

Selinger algorithm

- 1. j = set of join nodes
- 2. for (i in $1 \dots |j|$): // First find best plan for single join, then for two joins, etc.
- 3. for s in {all length i subsets of j } // Looking at a concrete subset of joins
- 4. bestPlan = {} // We want to find the best plan for this concrete subset
- 5. for s' in {all length $i-1$ subsets of s }
- 6. subplan = optjoin(s') // Look-up in the cache the best query plan for s but with one relation missing
- 7. plan = best way to join ($s-s'$) to subplan // Now find the best plan to extend s' by one join to get s
- 8. if ($\text{cost}(\text{plan}) < \text{cost}(\text{bestPlan})$)
- 9. bestPlan = plan // Update the best plan for computing s
- 10. optjoin(s) = bestPlan
- 11. return optjoin(j)

Statistics

- TableStats
 - Each table, a TableStats
- StringHistogram
 - Already provided
- IntHistogram
 - Each attribute, a IntHistogram/StringHistogram
 - How it works?

Cardinalities and Costs

- Filter and Join
- Read code