Name

CWID

Quiz
2

Due Friday, December 2nd by 11:59pm (Chicago local time)

CS525 - Advanced Database Organization
Fall 2022

Please leave this empty! 1 2 Sum

Instructions

- **You have to upload the assignment to the course blackboard**.
- This is an individual and not a group assignment. Fraud will result in **0** points
- For your convenience the number of points for each part and questions are shown in parenthesis.
- There are **2** parts in this quiz
  1. Result Size Estimations
  2. I/O Cost Estimation

Part 1 Result Size Estimations (Total: 20 Points)

Consider a table `student` with attributes `CWID`, `name`, `major`, `credits`, a table `course` with `title`, `instructor`, `credits`, and a table `registered` with attributes `student` and `course`. `registered.student` is a foreign key to `CWID`. Attribute `course` of relation `registered` is a foreign key to attribute `title` of relation `course`. Given are the following statistics:

T(student) = 30,000	T(course) = 80	T(registered) = 10,000
V(student,CWID) = 30,000	V(course,title) = 80	V(registered, student) = 3,000
V(student,name) = 29,500	V(course,instructor) = 50	V(registered, course) = 30
V(student,major) = 20	V(course,credits) = 6	
V(student,credits) = 32		

Question 1.1 Estimate Result Size (4 Points)

Estimate the number of result tuples for the query $q = \sigma_{instructor=Bob \wedge credits \leq 3}(course)$ using the first assumption presented in class (values used in queries are uniformly distributed within the active domain). Assume that the minimal and maximal values in the `credits` attribute are `1` and `6`, respectively.

Question 1.2 Estimate Result Size (5 Points)

Estimate the number of result tuples for the query $q = \sigma_{major=CS \vee credits<10}(students)$ using the first assumption presented in class. Assume that the minimal and maximal values in the `credits` attribute are `1` and `32`, respectively.

Question 1.3 Estimate Result Size (5 Points)

Estimate the number of result tuples for the query $q = \sigma_{credits>3 \wedge credits \leq 5}(course)$ using the first assumption presented in class. Assume that the minimal and maximal values in the `credits` attribute are `1` and `6`.

Question 1.4 Estimate Result Size (6 Points)

Estimate the number of result tuples for the query q below using the first assumption presented in class. Assume that the minimal and maximal values in the `credits` attribute are `1` and `32`,

$$q = \sigma_{credits \geq 20}(student) \bowtie_{CWID=student} registered \bowtie_{course=title} course$$

Part 2 I/O Cost Estimations (Total: 20 Points)

Question 2.1 I/O Cost Estimation (20 Points)

Consider two unordered, non-clustered relations `R` and `S` with `B(R) = 300,000` and `B(S) = 4,000` blocks, and `S(R) =` $\frac{1}{30}$, and `S(S) =` $\frac{1}{20}$. You have `M = 101` memory pages available. Let `R` has an index on the joining attribute `C`. Compute the minimum number of I/O operations needed to join these two relations using **tuple-based-nested-loop join** (relations are clustered), **block-nested-loop join** (relations are unclustered), **merge-join** (the inputs are not sorted and non-clustered), and **non-clustering index-join** (relations are clustered, read of `S` with uniform distribution assumption and expected `120` matching tuples on `R` for `S`). Justify you result by showing the I/O cost estimation for each join method.

Scratch Page - This page is intentionally left blank