CS127 Homework 6

Due: November 19, 2014 3:00PM

Warmup 1 (Textbook Problem 12.3)

Let relations $r_1(A,B,C)$ and $r_2(C,D,E)$ have the following properties: r_1 has 20,000 tuples, r_2 has 45,000 tuples, 25 tuples of r_1 fit on one block, and 30 tuples of r_2 fit on one block. Estimate the number of block transfers required, using each of the following join strategies for $r_1 \bowtie r_2$:

1. Nested-loop join
2. Block nested-loop join
3. Merge join
4. Hash join

Warmup 2 (Textbook Problem 12.6)

Consider the following bank database, where the primary keys are underlined:
branch(branch_name, branch_city, assets)
customer(customer_name, customer_street, customer_city)
loan(loan_number, branch_name, amount)
borrower(customer_name, loan_number)
account(account_number, branch_name, balance)
depositor(customer_name, account_number)

Suppose that a B+Tree index on *branch_city* is available on relation *branch*, and that no other index is available. List different ways to handle the following selections that involve negation:

1. $\sigma_{\neg(branch\_city < \text{"}Brooklyn\text{"})}(branch)$
2. $\sigma_{\neg(branch\_city = \text{"}Brooklyn\text{"})}(branch)$
3. $\sigma_{\neg(branch\_city < \text{"}Brooklyn\text{"}) \vee assets < 5000}(branch)$

Warmup 3 (Textbook Problem 13.4)

Consider the relations $r_1(A,B,C)$, $r_2(C,D,E)$, and $r_3(E,F)$, with primary keys A, C, and E, respectively. Assume that r_1 has 1000 tuples, r_2 has 1500 tuples, and r_3 has 750 tuples. Estimate the size of $r_1 \bowtie r_2 \bowtie r_3$ and give an efficient strategy for computing the join.

Warmup 4 (Textbook Problem 14.12)

List the ACID properties. Explain the usefulness of each.

Problem 5 (To Be Graded)

Consider again the simplified university registrar database from the previous homeworks:

Student		
name	**gradyear**	**gpa**
Amy	2016	3.95
Ben	2015	3.87
Carl	2016	3.29
Dan	2017	3.43
Eliza	2015	4.0

Course		
title	**semester**	**instructor**
CS33	2014F	Doeppner
CS127	2014F	Zdonik
CS195	2013F	Kraska
CS127	2012F	Zdonik
CS136	2012S	Fonseca

Enrollment			
name	**title**	**semester**	**grade**
Eliza	CS33	2014F	A
Eliza	CS127	2014F	A
Ben	CS127	2012F	A
Carl	CS195	2013F	C
Carl	CS127	2014F	B

The database has the following statistics:

- *Student(name,gradyear,gpa):* $n_{Student} = 10,000,\ f_{Student} = 100$
- *Course(title,semester,instructor):* $n_{Course} = 200,\ f_{Course} = 50$
- *Enrollment(name,title,semester,grade):* $n_{Enrollment} = 250,000,\ f_{Enrollment} = 25$

Answer the questions about the following queries:

1. $select\ name, gpa\ from\ student\ where\ gpa > 3.75$

   a. Draw a plan for this query.

   b. Assume *Student* is sorted. Does this affect how the query should be executed? Explain.

   c. Instead, we add an index to speed up the query. Should we use a B-tree or hash index? Explain.

2. $select\ *\ from\ student\ s, course\ c, enrollment\ e\ where\ s.name = e.name\ and\ c.title = e.title\ and\ c.semester = e.semester$

   a. Draw all possible plans for this query.

   b. Choose the best plan and give its cost with a block nested-loop join and a buffer size of 3 blocks.

   c. Choose the best plan and give its cost with a block nested-loop join and a buffer size of 6 blocks.