CHAPTER 12
DISCUSSIONS 2

Discussion 12-6

Let relations $r_1(A, B, C)$ and $r_2(C, D, E)$ have the following properties.

- r_1 : 20,000 tuples; 25 tuples/block
- r_2 : 45,000 tuples; 20 tuples/block
- Buffer size: 64 blocks

Estimate the number of block transfers required for $r_1 \bowtie r_2$ using *block nested-loop join*.

Discussion 12-7

Let relations $r_1(A, B, C)$ and $r_2(C, D, E)$ have the following properties.

- r_1 : 20,000 tuples; 25 tuples/block
- r_2 : 45,000 tuples; 20 tuples/block
- Buffer size: 64 blocks

Estimate the number of block transfers required for $r_1 \bowtie r_2$ using *hash join*.

Discussion 12-8

Perform a *hash join* of *r* and *s* on the numeric attribute. Assume $M = 4$ and use modulo as your hash function.

relation *r*

3	A
10	B
8	C
6	D
1	E
9	F
8	G
2	H
7	I
2	J
4	K
5	L
3	M
6	N
9	O
1	P

6	b
1	a
3	c
2	e
5	f
4	g
9	k
7	n

relation *s*

Discussion 12-9

What would be the most efficient way to evaluate the following query? Represent your (heuristic) answer in a query evaluation tree.

SELECT *i_id, c_id, title*

FROM *teaches*, *course*

WHERE *teaches.c_id = course.c_id*

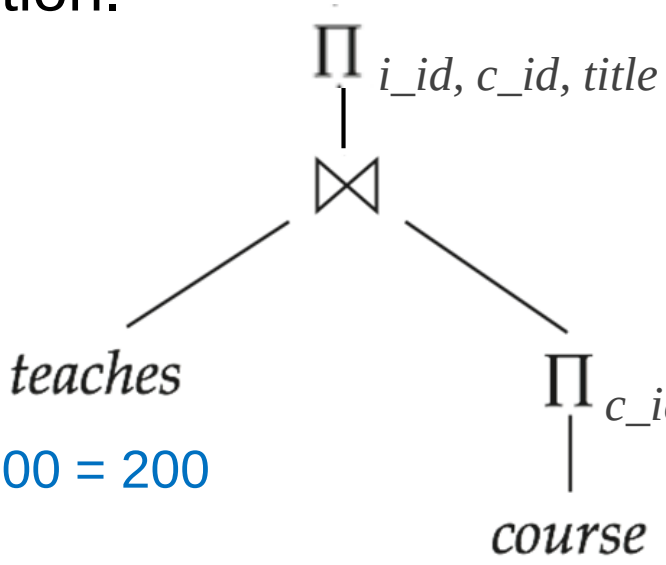
- ***teaches(i_id, c_id, year, semester)***

  $n_{teaches}$ = 20,000; $b_{teaches}$ = 200; *size(i_id, c_id)* = 50%
- ***course(c_id, title, dept, credit, level, description)***

  n_{course} = 10,000; b_{course} = 400; *size(c_id, title)*=20%
- ***M***=22 (# of buffer blocks)

Discussion 12-10

Suppose we chose the following evaluation plan for the previous query. Estimate the cost when we use materialization for all of the operations except for the topmost projection.



(20%)

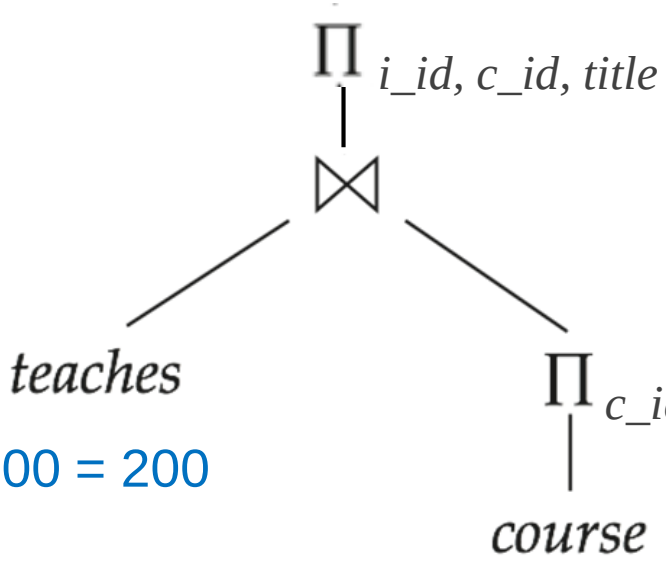
$b_{teaches}$ = 20,000/100 = 200

b_{course} = 10,000/25 = 400

M=22 (# of buffer blocks)

Discussion 12-11

Can you improve performance by *pipelining*? Which pairs of operations would you pipeline, and how much would you gain?



(20%)

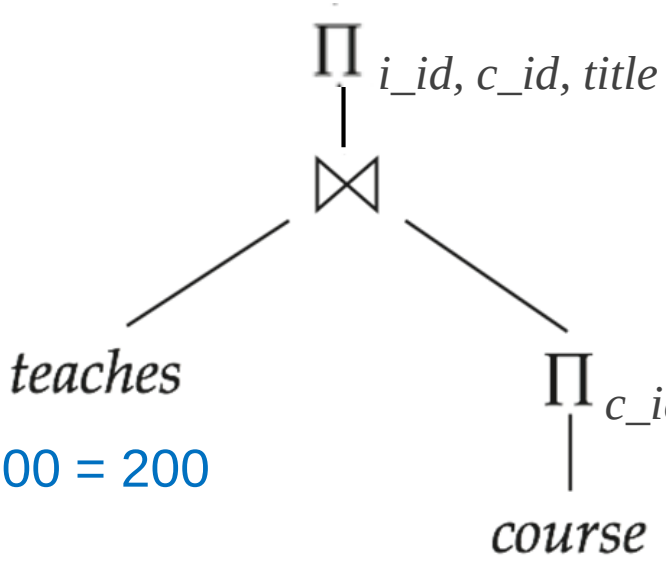
$b_{teaches}$ = 20,000/100 = 200

b_{course} = 10,000/25 = 400

M=22 (# of buffer blocks)

Discussion 12-12

Suppose relations *teaches* and *course* are each sorted on the join attribute *c_id*. Would you use a different join algorithm? Justify.



(20%)

$b_{teaches}$ = 20,000/100 = 200

b_{course} = 10,000/25 = 400

M=22 (# of buffer blocks)