Lecture 24:
• Quiz 4
• Query processing (cont)

Announcements:
• R-3 due Mon
• Project out
• HW-4 out

Query Operators
FROM clause ▷ ◁ (join)

SELECT *
FROM R JOIN S ON (R.sid = S.sid)

Assumptions:
• 8 KB sized pages, B available buffer page slots
• 100,000 Page I/Os per second ... 1 Page I/O = 0.01 milliseconds
• use Page I/Os as cost metric ... but relate to time to help see difference

Data Assumptions:
• R has \( m \) rows in \( M \) pages ... \( M = 1000, m = 100,000 \)
• S has \( n \) rows in \( N \) pages ... \( N = 500, n = 40,000 \)
• In \( R ▷ ◁ S \), \( R \) is outer and \( S \) is inner relation
Query Operators

**Naive (Simple) Nested Loop Join**

for each row \( r \) in outer
for each row \( s \) in inner
if \( r \) and \( s \) match then add new row \( r, s \) to output buffer

*How does this actually work?*

- Read a page from \( R \) costs \( M \) total Page I/Os
- Compare a row \( r \) to each row \( s \) each \( r \) costs \( N \) Page I/Os

The total cost: \( M + m \times N \)

- In our scenario, \( 1000 + 100,000 \times 500 = 50,001,000 \) Page I/Os
- Or about 500 seconds!

What if we make \( S \) the outer? (Yes 40s) ... use smaller relation as outer

---

**Page-Oriented Nested Loop Join**

for each page in outer
for each page in inner
for each row \( r \) in outer page
for each row \( s \) in inner page
if \( r \) and \( s \) match then add new row \( r, s \) to output buffer

*How does this actually work?*

- For each page in \( R \), scan each page in \( S \)

The total cost: \( M + M \times N \)

- In our scenario, \( 1000 + 1000 \times 500 = 501,000 \) Page I/Os
- Or about 5 seconds ... better, but still slow

What if we make \( S \) the outer? ... 500,500 Page I/Os (so marginally)
Query Operators

Block Nested Loop Join

- Read $B - 2$ outer block pages save 1 for inner page, 1 for output
- Compare all $B - 2$ pages of rows with all inner page rows

The total cost: $M + \lceil M/(B - 2) \rceil \times N$

- Assume $B = 102$ ... very small! (MySQL default is 128 MB)
- In our scenario, $1000 + \lceil 1000/100 \rceil \times 500 = 6,000$ Page I/Os
- Or about 0.06 seconds

What if we make $S$ the outer? ... 5,500 Page I/Os (so marginally)