

# Fall 2007 CS186 Discussion Section:

## Weeks 10, 10/29 - 11/02

Your friendly TAs

October 31, 2007

## 1 Relational Operators

1. In this question, you will consider joining two relations from an ecommerce database. The **CARTS** relation represents shopping cars of items. the **CONTENTS** relations has the contents of all the carts, with a foreign key called **cartID** to **CARTS**:

```
CARTS(cartID, customerID, date, comment)
CONTENTS(cartID, productID, quantity, price)
```

Assume the following:

- fixed size tuples in both relations
  - 10 **CARTS** tuples per page
  - 100 **CONTENTS** tuples per page
  - 1000 pages in the **CARTS** relation
  - 5000 pages in the **CONTENTS** relation
  - 1002 frames in the buffer pool
  - join on **CARTS.cartID=CONTENTS.cartID**
- (a) How many I/Os are required for a page-oriented nested loops join, with **CARTS** as the outer relation and **CONTENTS** as the inner relation? Assume that the buffer manager is not used (i.e. every page reference generates an I/O).
- (b) Consider the previous question.  
Assuming that the buffer manager is fully used and:
- The buffer manager starts empty
  - One output buffer frame is pinned by the join, used to hold output tuples until they are ready to be flushed to disk. This frame is not unpinned by the join until the end of the query.
  - One input buffer frame is pinned by the join and holds the current page of the outer relation at all times. All I/Os to the outer relation are placed explicitly into this frame, which is not unpinned until the end of the query
  - The buffer manager is running MRU replacement policy
  - This is a single query system

Symbol	Meaning
$p(R)$	number of blocks of relation R
$t(R)$	number of tuples of relation R
$l(R)$	number of leaf pages (data entry pages) of B index of relation R
$sel$	selectivity factor

Table 1: Symbolisms used in the solution of the cost estimation exercises.

Given these assumptions, how many I/Os are required for a page-oriented nested loops join?

- (c) How many I/Os are required for a block nested loops join, with `CONTENTS` as the outer relation, `CARTS` as inner relation, and 1000 pages per block?
- (d) Assume there are only 52 frames in the buffer pool total. Give the I/Os for the following join algorithms.
- Hash Join, using `CARTS` to build the hash table in the second phase
  - Sort-Merge Join
  - Block Nested Loops, `CARTS` as outer, block-size=50
2. Consider the same database in Question 1, now running a selection range query on `CONTENTS.quantity` (e.g. `CONTENTS.quantity > c`). Let the selectivity of the query be  $sel = 0.4$ .
- (a) How many I/Os do we need if the whole table is scanned and `CONTENTS` is unsorted?
- $$p(CONTENTS) = 5,000$$
- (b) What if it is sorted on `CONTENTS.quantity`?
- (c) How many I/Os are required given the following:
- unclustered 3 level B+ tree index on `CONTENTS.quantity`.
  - data distributed evenly among data record pages
  - 500 data entries/page
  - query only requires rids of the `CONTENTS` records.
- (d) What if the previous query requires entire data records?
- (e) Let there be an index similar to the previous two questions. Assume that each data records page has a 10% chance of having an overflow page. How many I/Os are required using the index if the query requires entire data records?
- (f) Let there be an unclustered 3 level B+ tree index on `(CONTENTS.price, CONTENTS.quantity)`. How many I/Os are required using the index if the query requires the entire data record?