

## Evaluation of Relational Operations: Other Operators

### Chapter 14, Part B

Database Management Systems 3ed, R. Ramakrishnan and Johannes Gehrke
1

1

---

---

---


---

---

---

---

---



## Using an Index for Selections

- ❖ Cost depends on #qualifying tuples, and clustering.
  - Cost of finding qualifying data entries (typically small) plus cost of retrieving records (could be large w/o clustering).
  - In example, assuming uniform distribution of names, about 10% of tuples qualify (100 pages, 10000 tuples). With a clustered index, cost is little more than 100 I/Os; if unclustered, upto 10000 I/Os!
- ❖ Important refinement for unclustered indexes:
  1. Find qualifying data entries.
  2. Sort the rid's of the data records to be retrieved.
  3. Fetch rids in order. This ensures that each data page is looked at just once (though # of such pages likely to be higher than with clustering).

2


2 now not an I/O for each tuple qualified  
but an I/O for each page that has these qualified tuples!  
But this algorithm is blocking (need to get all tuple ids & sort them & return together), instead of send one after another to output on the go

The easiest way: read the table one page after another. For a page, read one tuple ~~one after another and evaluate if the tuple match predicate P~~. If so, return it, if not, iterate to next.

using an index for selections  
(assume select sid = 123 from reserves, average 2.5 tuples per side  
assume we use alternative 2, if there are n tuples with key = 123, then we need to do n I/Os if not clustered because we need to retrieve all pages where these tuples reside in. Or just 1 I/O if its clustered, we can just retrieve 1 page to get all tuples

but what if we select a non-key attribute  
that has tons of matches, i.e. name





## Two Approaches to General Selections

- ❖ First approach: Find the *most selective access path*, retrieve tuples using it, and apply any remaining terms that don't match the index:
  - Most selective access path: An index or file scan that we estimate will require the fewest page I/Os (eliminate most)
  - Terms that match this index reduce the number of tuples retrieved; other terms are used to discard some retrieved tuples, but do not affect number of tuples/pages fetched.
  - Consider *day<8/9/94 AND bid=5 AND sid=3*. A B+ tree index on *day* can be used; then, *bid=5* and *sid=3* must be checked for each retrieved tuple. Similarly, a hash index on *<bid, sid>* could be used; *day<8/9/94* must then be checked.

3

3

---

---

---

---

---

---

---

---

## Intersection of Rids

- ❖ **Second approach** (if we have 2 or more matching indexes that use Alternatives (2) or (3) for data entries):
  - Get sets of rids of data records using each matching index.
  - Then **intersect** these **sets of rids** (we'll discuss intersection soon!)
  - Retrieve the records and apply any remaining terms.
  - Consider **day<8/9/94 AND bid=5 AND sid=3**. If we have a B+ tree index on *day* and an index on *sid*, both using Alternative (2), we can retrieve rids of records satisfying *day<8/9/94* using the first, rids of recs satisfying *sid=3* using the second, intersect, retrieve records and check *bid=5*.

Database Management Systems 3ed, R. Ramakrishnan and Johannes Gehrke

4

what if all predicates are not selective?

second approach: get the set of rids using indexes that qualifies each predicate, intersect them to find the ones that qualifies all predicates. Then access the table (doing I/Os)

);  
sets of rids of data records using each mat  
intersect these sets of rids (we'll discuss i  
!)  
even the records and apply any remaining  
sider **day<8/9/94 AND bid=5 AND sid=3**. If  
ee index on *day* and an index on *sid*, both

4

## The Projection Operation

```
SELECT DISTINCT
  R.sid, R.bid
FROM Reserves R
```

- ❖ An approach based on **sorting**:
  - **Modify Pass 0 of external sort to eliminate unwanted fields.** Thus, runs of about 2B pages are produced, but tuples in runs are smaller than input tuples. (Size ratio depends on # and size of fields that are dropped.)
  - **Modify merging passes to eliminate duplicates.** Thus, number of result tuples smaller than input. (Difference depends on # of duplicates.)
  - **Cost:** In Pass 0, read original relation (size M), write out same number of smaller tuples. In merging passes, fewer tuples written out in each pass. Using Reserves example, 1000 input pages reduced to 250 in Pass 0 if size ratio is 0.25

Database Management Systems 3ed, R. Ramakrishnan and Johannes Gehrke

5

New term: query rewrite: rewrite to another query that does the same but cheaper

project only sid and bid, then sort the tuples  
external sort (sort-merge): get one page at a time from table, get the attributes that matter and remove others  
then sort the tuples in each page

get two pages, merge them, smaller tuples come first, throw the duplicates and keep only one if there is an equal

5

## Projection Based on Hashing

- ❖ **Partitioning phase:** Read R using one input buffer. For each tuple, discard unwanted fields, apply hash function *h1* to choose one of B-1 output buffers.
  - Result is B-1 partitions (of tuples with no unwanted fields). 2 tuples from different partitions guaranteed to be distinct.
- ❖ **Duplicate elimination phase:** For each partition, read it and build an in-memory hash table, using hash fn *h2* ( $\neq h1$ ) on all fields, while discarding duplicates.
  - If partition does not fit in memory, can apply hash-based projection algorithm recursively to this partition.
- ❖ **Cost:** For partitioning, read R, write out each tuple, but with fewer fields. This is read in next phase.

Database Management Systems 3ed, R. Ramakrishnan and Johannes Gehrke

6

remove unwanted attributes and also duplicates

6

## Discussion of Projection



- ❖ Sort-based approach is the standard; better handling of skew and result is sorted.
- ❖ If an index on the relation contains all wanted attributes in its search key, can do *index-only* scan.
  - Apply projection techniques to data entries (much smaller!)
- ❖ If an ordered (i.e., tree) index contains all wanted attributes as *prefix* of search key, can do even better:
  - Retrieve data entries in order (index-only scan), discard unwanted fields, compare adjacent tuples to check for duplicates.

Database Management Systems 3ed, R. Ramakrishnan and Johannes Gehrke

7

7

i.e. on prefix of search key - if we want to project on a,b  
if index is a,b,c, then it's good  
if index of c,a,b, then (a,b)s are all over the place, it's not good

## Set Operations



- ❖ Intersection and cross-product special cases of join.
- ❖ Union (Distinct) and Except similar; we'll do union.
- ❖ Sorting based approach to union:
  - Sort both relations (on combination of all attributes).
  - Scan sorted relations and merge them.
  - *Alternative:* Merge runs from Pass 0 for *both* relations.
- ❖ Hash based approach to union:
  - Partition R and S using hash function *h*.
  - For each S-partition, build in-memory hash table (using *h2*), scan corr. R-partition and add tuples to table while discarding duplicates.

Database Management Systems 3ed, R. Ramakrishnan and Johannes Gehrke

8

8

intersection: tuples in both table A and table B whose every each attributes are the same

schema compatible: same amount of attributes in both schema  
types of corresponding attributes have to match  
attribute names don't have to match

similar as sort-merge, but pick one and throw the other if we find a match (because union eliminates duplicates)

## Aggregate Operations (AVG, MIN, etc.)



- ❖ Without grouping:
  - In general, requires scanning the relation.
  - Given index whose search key includes all attributes in the SELECT or WHERE clauses, can do index-only scan.
- ❖ With grouping:
  - Sort on group-by attributes, then scan relation and compute aggregate for each group. (Can improve upon this by combining sorting and aggregate computation.)
  - Similar approach based on hashing on group-by attributes.
  - Given tree index whose search key includes all attributes in SELECT, WHERE and GROUP BY clauses, can do index-only scan; if group-by attributes form prefix of search key, can retrieve data entries/tuples in group-by order.

Database Management Systems 3ed, R. Ramakrishnan and Johannes Gehrke

9

9

i.e select avg gpa from students group by age  
the index should be (age, gpa)  
should not be (gpa, age)

## Impact of Buffering



- ❖ If several operations are executing concurrently, estimating the number of available buffer pages is guesswork.
- ❖ Repeated access patterns interact with buffer replacement policy.
  - e.g., Inner relation is scanned repeatedly in Simple Nested Loop Join. With enough buffer pages to hold inner, replacement policy does not matter. Otherwise, MRU is best, LRU is worst (*sequential flooding*).
  - Does replacement policy matter for Block Nested Loops?
  - What about Index Nested Loops? Sort-Merge Join?

10

---

---

---

---

---

---

---

## Summary



- ❖ A virtue of relational DBMSs: *queries are composed of a few basic operators*; the implementation of these operators can be carefully tuned (and it is important to do this!).
- ❖ Many alternative implementation techniques for each operator; no universally superior technique for most operators.
- ❖ Must consider available alternatives for each operation in a query and choose best one based on system statistics, etc. This is part of the broader task of optimizing a query composed of several ops.

11

---

---

---

---

---

---

---