Introduction to Database Systems
CSE 414

Lecture 17:
Basics of Query Optimization and
Query Cost Estimation

Announcements

• Midterm will be released by end of day today
• Need to start one HW6 step NOW:
  – Need to make an AWS account, can use existing
    Amazon account
  – Click on application button under Students and fill out
    form with your @uw.edu email
  – Will then be sent email for verification, must click to
    verify your email address

Two typical kinds of queries

- Point queries
  - What data structure should be used for index?
- Range queries
  - What data structure should be used for index?

Choosing Index is Not Enough

- To estimate the cost of a query plan, we still need to consider other factors:
  - How each operator is implemented
  - The cost of each operator
  - Let’s start with the basics

Cost of Reading Data From Disk

Cost Parameters

- Cost = I/O + CPU + Network BW
  - We will focus on I/O in this class
- Parameters (a.k.a. statistics):
  - \( B(R) \) = # of blocks (i.e., pages) for relation \( R \)
  - \( T(R) \) = # of tuples in relation \( R \)
  - \( V(R, a) \) = # of distinct values of attribute \( a \)
Cost Parameters

- Cost = I/O + CPU + Network BW
  - We will focus on I/O in this class
- Parameters (a.k.a. statistics):
  - \( B(R) \) = # of blocks (i.e., pages) for relation \( R \)
  - \( T(R) \) = # of tuples in relation \( R \)
  - \( V(R, a) \) = # of distinct values of attribute \( a \)

  When \( a \) is a key, \( V(R, a) = T(R) \)
  When \( a \) is not a key, \( V(R, a) \) can be anything \( \leq T(R) \)

DBMS collects statistics about base tables must infer them for intermediate results

---

Selectivity Factors for Conditions

How many tuples would this select:

SELECT *
FROM One_year
WHERE did = 32

1 tuple (out of 365)

---

Selectivity Factors for Conditions

How many tuples would this select:

SELECT *
FROM One_year
WHERE month = Jan

31 tuples (out of 365)
This is roughly 1/12 of the tuples, because 12 distinct values equally distributed.

---
Cost of Reading Data From Disk

- Sequential scan for relation R costs $B(R)$
- Index-based selection
  - Estimate selectivity factor $f$ (see previous slide)
  - Clustered index: $f B(R)$
  - Unclustered index $f T(R)$

Note: we ignore I/O cost for index pages

Index Based Selection

- Example: $B(R) = 2000$
  $T(R) = 100,000$
  $V(R,a) = 20$

- Table scan:
- Index based selection:

Index Based Selection

- Example: $B(R) = 2000$
  $T(R) = 100,000$
  $V(R,a) = 20$

- Table scan: $B(R) = 2,000$ I/Os
- Index based selection:
  - If index is clustered: $B(R) \cdot \frac{1}{V(R,a)} = 100$ I/Os
    Why: we know we can scan a full block to get the desired range
  - If index is unclustered: $T(R) \cdot \frac{1}{V(R,a)} = 5,000$ I/Os
Index Based Selection

- Example:
  \[
  B(R) = 2000 \\
  T(R) = 100,000 \\
  V(R, a) = 20
  \]

- Table scan: \( B(R) = 2,000 \) I/Os
- Index based selection:
  - If index is clustered: \( B(R) \times 1/V(R, a) = 100 \) I/Os
  - Why: we know we can scan a full block to get the desired range
  - If index is unclustered: \( T(R) \times 1/V(R, a) = 5,000 \) I/Os

Lesson: Don’t build unclustered indexes when \( V(R, a) \) is small!

Cost of Executing Operators
(Focus on Joins)

Outline

- Join operator algorithms
  - One-pass algorithms (Sec. 15.2 and 15.3)
  - Index-based algorithms (Sec 15.6)

- Note about readings:
  - In class, we discuss only algorithms for joins
  - Other operators are easier: read the book

Join Algorithms

- Hash join
- Nested loop join

Hash Join

Hash join: \( R \bowtie S \)
- Scan \( R \), build buckets in main memory
- Then scan \( S \) and join
- Cost: \( B(R) + B(S) \)
- Which relation to build the hash table on?
Hash Join

Hash join:  \( R \bowtie S \)
- Scan \( R \), build buckets in main memory
- Then scan \( S \) and join
- Cost: \( B(R) + B(S) \)
- Which relation to build the hash table on?
- One-pass algorithm when \( B(R) \leq M \)
  - \( M = \) number of memory pages available

Hash Join Example

<table>
<thead>
<tr>
<th>Patient</th>
<th>Insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>'Bob'</td>
</tr>
<tr>
<td>2</td>
<td>'Ela'</td>
</tr>
<tr>
<td>3</td>
<td>'Jill'</td>
</tr>
<tr>
<td>4</td>
<td>'Joe'</td>
</tr>
<tr>
<td>pid</td>
<td></td>
</tr>
<tr>
<td>name</td>
<td></td>
</tr>
<tr>
<td>address</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient</th>
<th>Insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>'Blue'</td>
</tr>
<tr>
<td>4</td>
<td>'Prem'</td>
</tr>
<tr>
<td>pid</td>
<td></td>
</tr>
<tr>
<td>provider</td>
<td></td>
</tr>
<tr>
<td>policy_nb</td>
<td></td>
</tr>
</tbody>
</table>

Two tuples per page

Memory \( M = 21 \) pages

Hash h: \( \text{pid} \% 5 \)

Disk

Input buffer

Output buffer

Write to disk or pass to next operator
Hash Join Example

Step 2: Scan Insurance and probe into hash table
Done during calls to next()

Disk
Patient Insurance
1 2 4 6 6
3 4 4 3 1 3
9 6 2 8
8 5 8 9

Memory M = 21 pages
Hash h: pid % 5

3 1 0 2 1 3 4 5

Input buffer
Output buffer

Keep going until read all of Insurance

Cost: B(R) + B(S)

Nested Loop Joins

• Tuple-based nested loop \( R \bowtie S \)
• \( R \) is the outer relation, \( S \) is the inner relation

\[
\begin{align*}
\text{for each tuple } t_1 \text{ in } R & \text{ do } \\
\text{for each tuple } t_2 \text{ in } S & \text{ do } \\
\text{if } t_1 \text{ and } t_2 \text{ join} & \text{ then output } (t_1, t_2)
\end{align*}
\]

What is the Cost?

Page-at-a-time Refinement

• Cost: \( B(R) + T(R) B(S) \)
• Multiple-pass since \( S \) is read many times

\[
\begin{align*}
\text{for each page of tuples } r \text{ in } R & \text{ do } \\
\text{for each page of tuples } s \text{ in } S & \text{ do } \\
\text{for all pairs of tuples } t_1 \text{ in } r, t_2 \text{ in } s & \text{ do } \\
\text{if } t_1 \text{ and } t_2 \text{ join} & \text{ then output } (t_1, t_2)
\end{align*}
\]

What is the Cost?

Page-at-a-time Refinement

Disk
Patient Insurance
1 2 4 6 6
3 4 4 3 1 3
9 6 2 8
8 5 8 9

Input buffer for Patient
Input buffer for Insurance

Output buffer

Disk
Patient Insurance
1 2 4 3
3 4 4 3 1 3
9 6 2 8
8 5 8 9

Input buffer for Patient
Input buffer for Insurance

Output buffer
### Page-at-a-time Refinement

- Disk
  - Patient
    - Input buffer for Patient
    - Output buffer
    - 12
    - 2 4 6 6
    - 3 4 4 3 1 3
    - 9 6 2 8
    - 8 9
  - Insurance
    - Input buffer for Insurance
    - Keep going until read all of Insurance
    - Then repeat for next page of Patient... until end of Patient

Cost: \( B(R) + B(R)B(S) \)

### Index Nested Loop Join

\( R \bowtie S \)

- Assume \( S \) has an index on the join attribute
- Iterate over \( R \), for each tuple fetch corresponding tuple(s) from \( S \)

- Cost:
  - If index on \( S \) is clustered:
    \[ B(R) + T(R) \cdot (B(S) \cdot 1/V(S,a)) \]
  - If index on \( S \) is unclustered:
    \[ B(R) + T(R) \cdot (T(S) \cdot 1/V(S,a)) \]

### Index Nested Loop Join

If index on \( S \) is clustered:

\[ B(R) + T(R) \cdot (B(S) \cdot 1/V(S,a)) \]

- Why is the multiplier term \( T(R) \)?
- What does \( 1/V(S,a) \) represent?

\( T(R) \) must be used because we cannot assume that a whole block of \( B(R) \) will have the same attribute to join on, and thus use the same index access on \( S \) for.

\( 1/V(S,a) \) represents the nature of the \( B^+ \) Tree index. We are only scanning as much as we need. Note that the performance of the index join will decrease as \( V \) decreases.

### Index Nested Loop Join

If index on \( S \) is unclustered:

\[ B(R) + T(R) \cdot (T(S) \cdot 1/V(S,a)) \]

Why did this change from \( B(R) \) to \( T(R) \)?

Remember that tuples are stored on contiguous blocks. In a clustered index from before we knew we can scan a single chunk of the disk to get the entire desired range. In an unclustered index we no longer can assume contiguous access. Thus we estimate that every tuple needs its own I/O operation.
Review: Logical vs Physical Plans

• Logical plans:
  – Created by the parser from the input SQL text
  – Expressed as a relational algebra tree
  – Each SQL query has many possible logical plans

• Physical plans:
  – Goal is to choose an efficient implementation for each operator in the RA tree
  – Each logical plan has many possible physical plans

Review: Relational Algebra

Relational algebra expression is also called the “logical query plan”

SELECT 
FROM Supplier x, Supply y
WHERE x.sid = y.sid 
and y.pno = 2 
and x.scity = 'Seattle'
and x.sstate = 'WA'

Supplier(sid, name, scity, sstate)
Supply(sid, pno, quantity)

Review: Physical Query Plan 1

(On the fly)  σscity='Seattle' and sstate='WA' and pno=2
(On the fly)  πsname

A physical query plan is a logical query plan annotated with physical implementation details

(Hash join)  SELECT name 
FROM Supplier x, Supply y
WHERE x.sid = y.sid 
and y.pno = 2 
and x.scity = 'Seattle'
and x.sstate = 'WA'

(On the fly)  sname = sid

Supplier (File scan)  Supply (File scan)

Review: Physical Query Plan 2

(On the fly)  σscity='Seattle' and sstate='WA' and pno=2
(On the fly)  πsname

Same logical query plan
Different physical plan

(Hash join)  SELECT name 
FROM Supplier x, Supply y
WHERE x.sid = y.sid 
and y.pno = 2 
and x.scity = 'Seattle'
and x.sstate = 'WA'

(On the fly)  sname = sid

Supplier (File scan)  Supply (File scan)

Query Optimization: Overview

• Compute cost of each operator
  – This depends on:
    • Table statistics (# of tuples etc)
    • Algorithm used

• Cost of a physical plan = sum(each operator cost)

• Cost each plan and choose the one with lowest cost

Cost of Query Plans
Logical Query Plan 2

SELECT name
FROM Supplier x, Supply y
WHERE x.sid = y.sid
and x.scity = 'Seattle'
and y.pno = 2
and x.scity = 'WA'

Physical Plan 1

Cost of Supplier(scity) = 4
Cost of Supply(pno) = 100
Total cost: 100*4 = 400

Logical Query Plan 2

SELECT name
FROM Supplier x, Supply y
WHERE x.sid = y.sid
and x.scity = 'Seattle'
and y.pno = 2
and x.scity = 'WA'

Physical Plan 2

Cost of Supplier(scity) = 2500
Cost of Supply(pno) = 100
Total cost: 100*2500 = 250000
Query Optimizer Summary

- Input: A logical query plan
- Output: A good physical query plan
- Basic query optimization algorithm
  - Enumerate alternative plans (logical and physical)
  - Compute estimated cost of each plan
  - Choose plan with lowest cost

- This is called cost-based optimization