Lecture 29: Final Review

Wednesday, December 11, 2002

1

Outline

- From the homework: Mr. Frumble's blues
- An exercise: counting the number of joins

- Redo logging – 17.3
- Redo/undo logging – 17.4

- Course evaluation forms

2

Counting the Number of Join Orders (Exercise)

$R_0(A_0,A_1)$ $R_1(A_1,A_2)$ . . . $R_n(A_n,A_{n+1})$

- The number of left linear join trees is:
  $n!$
- The number of left linear join trees without cartesian products is:
  2^n (why ?)
- The number of bushy join trees is:
  $n!/(n+1)*C^{2n}_n = (2n)!/((n+1)*(n!))$
- The number of bushy join trees without cartesian product is:
  $2^{n-1}/(n+1)*C^{2n}_n$ (why ?)

3

Number of Subplans Inspected by Dynamic Programming

$R_0(A_0,A_1)$ $R_1(A_1,A_2)$ . . . $R_n(A_n,A_{n+1})$

- The number of left linear subplans inspected is:
  $\Sigma_{k=1,n}C^n_k*k = n2^{n-1}$
- The number of left linear subplans without cartesian products inspected is:
  $\Sigma_{k=1,n}(n-k+1)*2 = n(n+1)$ why ?
- The number of bushy join subplans inspected is:
  $\Sigma_{k=1,n}C^n_k*2^k = 3^k$ why ?
- The number of bushy join subplans without cartesian product:
  $\Sigma_{k=1,n}(n-k+1)*(k-1) = n*n*(n-1)/2 – n(n-1)(2n-1)/6 = n(n-1)(n+1)/6$

4

The Final

- Date: Friday, December 13
- Time: 8:30 – 10:20
- Place: this room
- Open book exam !

5

What to Prepare for the Final

Everything !
- Data modeling
- Relational model
- XML
- Relational algebra
- SQL
- Storage
- Indexes
- Physical operators
- Optimization
- Recovery

6

Data Modeling

- E/R diagrams, ODL
- Keys
- Relationships
- Inheritance
- Mapping to relations

7

Relational Algebra

- Relations
- Keys
- Functional dependencies
- Decomposition
- Normal forms

8

XML

- XML syntax
- DTD
- From relations to XML
- From XML to relations

9

SQL

- Select-from-where
- Subqueries
- Aggregation
- Nulls
- Outer joins

10

SQL (continued)

- Database modification
- Defining and modifying relation schemas
- Constraints
  - On attribute values
  - Keys
  - Foreign keys
- Embedded SQL

11

Data Storage

- The I/O model of computation
- Representing data elements:
  - Grouping records into blocks
  - Variable length records
  - Overflow blocks

12

Index Structures

- Terminology:
  - Dense/sparse index
  - Primary/secondary index
- B^+-trees
- Hash tables

13

Physical Operators

- One-pass algorithms
- Nested-loop joins
- Two-pass algorithms based on sorting
- Two-pass algorithms based on hash tables
- Index-based algorithms

14

Optimizations

- Extended logical operators
- Algebraic identities
- Heuristic based optimization:
- Cost based optimization:
  - Size estimation
  - Dynamic programming for join order

15

Recovery

- Undo logging
- Redo logging
- Undo/redo logging

16

General Advice

- Some problems will require thinking
  - Use judgment
- Problem difficulty may be uneven:
  - do the easy ones first

17

Grading

- –Homework 25%
  - best 4 out of 5
- –Project: 25%
- –Midterm: 20%
- –Final: 25%
- –Intangibles: 5%

18

The End

19